World Journal of Microbiology and Biotechnology (2023) 39:246
https://doi.org/10.1007/s11274-023-03682-5

RESEARCH

Functional characterization of the *Komagataella phaffii 1033* gene promoter and transcriptional terminator

Yanelis Robainas-del-Pino[1] · José María Viader-Salvadó[1] · Ana Lucía Herrera-Estala[1] · Martha Guerrero-Olazarán[1]

Received: 22 October 2022 / Accepted: 16 June 2023 / Published online: 8 July 2023

Abstract

The methylotrophic yeast *Komagataella phaffii* (syn. *Pichia pastoris*) is a widely used host for extracellularly producing heterologous proteins via an expression cassette integrated into the yeast genome. A strong promoter in the expression cassette is not always the most favorable choice for heterologous protein production, especially if the correct folding of the protein and/or post-translational processing is the limiting step. The transcriptional terminator is another regulatory element in the expression cassette that can modify the expression levels of the heterologous gene. In this work, we identified and functionally characterized the promoter (P_{1033}) and transcriptional terminator (T_{1033}) of a constitutive gene (i.e., the *1033* gene) with a weak non-methanol-dependent transcriptional activity. We constructed two *K. phaffii* strains with two combinations of the regulatory DNA elements from the *1033* and *AOX1* genes (i.e., P_{1033}-T_{AOX1} and P_{1033}-T_{1033} pairs) and evaluated the impact of the regulatory element combinations on the transcript levels of the heterologous gene and endogenous *1033* and *GAPDH* genes in cells grown in glucose or glycerol, and on the extracellular product/biomass yield. The results indicate that the P_{1033} has a 2–3% transcriptional activity of the *GAP* promoter and it is tunable by cell growth and the carbon source. The combinations of the regulatory elements rendered different transcriptional activity of the heterologous and endogenous genes that were dependent on the carbon source. The promoter-terminator pair and the carbon source affected the heterologous gene translation and/or protein secretion pathway. Moreover, low heterologous gene-transcript levels along with glycerol cultures increased translation and/or protein secretion.

Keywords *1033* promoter · *1033*-transcriptional terminator · *AOX1*-transcriptional terminator · *GAPDH* gene · Gene transcript levels · *Pichia pastoris*

Introduction

The methylotrophic yeast *Komagataella phaffii* (formerly known as *Pichia pastoris*) is a widely used host for producing heterologous proteins, due to its efficient protein secretion system that simplifies protein recovery, its ability to perform post-translational modifications, and its ability to be grown to a high biomass concentration in low-cost defined media. Furthermore, it can yield high levels of extracellular heterologous proteins in controlled bioprocesses (Cereghino and Cregg 2000). Usually, an expression cassette harboring a promoter sequence, the heterologous coding sequence, and a transcriptional terminator sequence is integrated into the yeast genome to express the heterologous gene during cell growth or in the presence of an inductor.

The transcription of the heterologous gene is a crucial step in recombinant protein production. Therefore, strong and controllable promoters are usually preferred for the efficient production of recombinant proteins (Vogl and Glieder 2013). Since the promoter from the alcohol oxidase 1 gene (P_{AOX1}) is a strong and tightly regulated methanol-inducible promoter, it is the most commonly used promoter to produce foreign proteins in the *K. phaffii* expression system (Cereghino and Cregg 2000). Nevertheless, when methanol is used as a carbon source and inductor, some safety and

✉ José María Viader-Salvadó
jose.viadersl@uanl.edu.mx

✉ Martha Guerrero-Olazarán
martha.guerrerool@uanl.edu.mx

[1] Facultad de Ciencias Biológicas, Instituto de Biotecnología, Universidad Autónoma de Nuevo León UANL, Av. Universidad S/N Col. Ciudad Universitaria, 66455 San Nicolás de los Garza, Nuevo León, Mexico

operational constraints must be considered in large-scale bioreactors (Mattanovich et al. 2014). In contrast, the promoter from the glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene (P_{GAP}) is a strong and constitutive promoter that has emerged as an alternative to P_{AOX1} to avoid the use of methanol (García-Ortega et al. 2019). Nevertheless, a strong promoter is not always the most favorable choice for heterologous protein production, especially when the correct folding of the heterologous protein and/or post-translational processing in the secretory pathway is the limiting step for obtaining high amounts of the extracellular product (Cereghino and Cregg 2000; Mattanovich et al. 2004; Hohenblum et al. 2004). The high level of heterologous gene expression can overwhelm the post-translational machinery of the cell, causing a significant proportion of the foreign protein to be misfolded, unprocessed, mislocalized, and degraded. Weak promoters can be used for expressing genes that produce toxic proteins to the host cells (Ruth et al. 2010) and in synthetic biology applications that commonly require the co-expression of multiple genes (Vogl et al. 2016; Jin et al. 2019). To date, the lowest strength *K. phaffii* promoter is reported to be from the *YPT1* gene (P_{YPT1}), which encodes for a GTPase involved in secretion (Sears et al. 1998). This promoter provides constitutive gene expression levels in glucose cultures that are 100-fold lower than those obtained from P_{GAP}, though it is far less studied than other promoters.

On the other hand, protein production levels are determined mainly by the synergistic action of the promoter and the transcriptional terminator (Matsuyama 2019). The main function of terminators is the termination of the transcription process, but the 3′-untranslated region (3′UTR) contained in the terminator sequence also modulates the half-life of mRNA, and transcriptional and translational efficiencies, nuclear export, and the cellular localization of mRNA (Kuersten and Goodwin 2003; Mayr 2019). Therefore, besides the promoter, the transcriptional terminator is another regulatory element in the expression cassette that can modify the expression levels of the gene of interest.

In mining homemade RNA-seq data from a *K. phaffii* KM71 strain grown in methanol, glucose, or glycerol, we observed that the PAS_chr3_1033 gene (hereafter *1033* gene) showed almost constant and low transcript levels. Therefore, the promoter of the *1033* gene (P_{1033}) should be weak and constitutive. The *1033* gene (also referred to as PP7435_CHR3-0135 gene for the *K. phaffii* CBS 7435 genome) codes for a hypothetical protein with 38% and 29% sequence identity with D-amino-acid oxidases from the yeast *Scheffersomyces stipitis* CBS 6054 (Genbank accession no. XP_001384459.2) and *K. phaffii* CBS 7435 (Genbank accession no. CCA39107.2), respectively.

To fully characterize the P_{1033} and the transcriptional terminator of the *1033* gene (T_{1033}), we identified the P_{1033} and T_{1033} sequences, along with the 5′UTR sequence, the transcription start site (TSS), the 3'UTR sequence and A- and T-rich sequences downstream of the stop codon for the *1033* gene. The sequence of the TATA-like motif, a mammalian-type initiator (INR) element, and the potential *K. phaffii* transcription factors (TF) for the P_{1033} sequence were also predicted. Furthermore, we constructed two *K. phaffii* strains with two combinations of regulatory DNA elements (i.e., P_{1033}-T_{AOX1} and P_{1033}-T_{1033} pairs; T_{AOX1}: transcriptional terminator of the *AOX1* gene) in the expression cassette that harbors the *FTEII* gene coding for a beta-propeller phytase (Viader-Salvadó et al. 2010) as the reporter gene, which was functionally associated with the alpha-factor prepro-secretion signal coding sequence. We selected clones of each construct with a single copy of the expression cassette integrated into the yeast genome and evaluated the impact of the regulatory DNA element combinations on the transcript levels of the heterologous *FTEII* gene in cells grown in glucose or glycerol as the carbon source. Moreover, we analyzed the transcript levels of the endogenous *1033* and *GAPDH* genes as transcript-level references. The extracellular product/biomass yields ($Y_{p/x}$) were also determined to correlate it with the transcript levels.

Materials and methods

Strains, plasmids, enzymes, media composition, and chemicals

The *K. phaffii* KM71 (*his4*) strain was purchased from Thermo Fisher Scientific (Waltham, MA, USA). Plasmid pUCIDT-AMP used for cloning was from Integrated DNA Technologies, Inc. (Coralville, IA, USA). Plasmid pGAH-FTEII was previously constructed in our laboratory (Herrera-Estala et al. 2022). This vector harbors an expression cassette containing the P_{GAP} sequence followed by the *Saccharomyces cerevisiae* alpha-factor prepro-secretion signal coding sequence (flanked by *Aat*II and *Xho*I restriction sites), a nucleotide sequence encoding the beta-propeller phytase FTEII with *K. phaffii*-preferred codons (Viader-Salvadó et al. 2010), the transcriptional terminator from the *AOX1* gene (T_{AOX1}), and a functional copy of the histidinol dehydrogenase (*HIS4*) gene. Endo H glycosidase, and *Aat*II, *Xho*I, and *Bsu36*I endonucleases were from New England Biolabs (Beverly, MA, USA). *Sal*I endonuclease was from Clontech (Palo Alto, CA, USA). M-MLV reverse transcriptase, RQ1 RNase-free DNase, GoTaq DNA polymerase, and oligo$(dT)_{15}$ primer were from Promega (Madison, WI, USA). Other primers and probes were from Integrated DNA Technologies, Inc. (Coralville, IA, USA); sequences are described in Supplementary Table S1. Yeast extract-peptone-dextrose (YPD), regeneration dextrose base (RDB) agar, and buffer minimal glycerol (BMG) were prepared

according to the manual from the Pichia Expression Kit (Thermo Fisher Scientific, Waltham, MA, USA). BMGly and BMGlc media were similar to BMG medium but with 30 mM glycerol (0.28% [w/v]) or glucose (0.54% [w/v]) instead of 1% (w/v) glycerol, respectively. All chemicals used were from Sigma-Aldrich Co. (St. Louis, MO, USA) or Productos Químicos Monterrey (Monterrey, Nuevo León, Mexico).

Promoter and transcriptional terminator sequences of the *1033* gene

The sequence located between the coding sequences (inter-CDS region) of the *1033* gene and the adjacent upstream gene of the *K. phaffii* GS115 chromosome 3 sequence (i.e., *NOB1* gene) was considered the P_{1033} sequence. The sequence surrounding the TSS was identified by analyzing the data for five homemade RNA-seq of a *K. phaffii* KM71 strain grown in methanol or glycerol, available at NCBI Sequence read archive (SRA) under the BioProject accession number PRJNA930494. The HISAT2 program (Kim et al. 2015) (Galaxy version 2.1.0) was used to map the RNA reads to the inter-CDS region of the *1033* gene and the upstream gene at the *K. phaffii* GS115 chromosome 3 sequence. The mapping results were visualized with the Integrative Genomics Viewer (Thorvaldsdóttir et al. 2013) to detect a subregion of the inter-CDS region without aligned RNA reads. The adjacent downstream and upstream nucleotides of the 3′ end of the subregion with non-aligned RNA-reads were considered as the sequence surrounding the dominant TSS and the beginning of the 5′UTR sequence of the *1033* gene. The sequence of the TATA-like motif and a mammalian-type INR element downstream of the TATA-like box were predicted using the program YAPP Eukaryotic Core Promoter Predictor (http://www.bioinformatics.org/yapp/cgi-bin/yapp.cgi), which was also used to predict the specific dominant TSS nucleotide through a synergistic combination of the TATA-like box and the INR element. Moreover, an *in silico* analysis for predicting the putative transcription factor-binding sites (TFBS) in the P_{1033} sequence was performed using Transcription Factor Affinity Prediction Web tools (Thomas-Chollier et al. 2011) with the fungi [transfac_2010.1 + jaspar] matrix and the yeast promoters (only for fungi matrices) model, together with the Benjamini–Hochberg multiple correction test. Only results with p-values less than 0.05 were considered. The UniProt database (https://www.uniprot.org/) was used to retrieve the *S. cerevisiae* protein sequences involved with the TFBS. The *K. phaffii* orthologs of these sequences were identified by the Best Reciprocal Blast Hit technique (Ward and Moreno-Hagelsieb 2014) and were considered potential *K. phaffii* TF for the P_{1033} sequence.

The sequence from the 3′ end of the *1033* CDS until 50 nucleotides downstream of the 3′UTR of the *1033* gene was considered as the T_{1033} sequence. The 3′URT sequence was determined by analyzing the data for five homemade RNA-seq (BioProject accession number PRJNA930494 at NCBI SRA) as described above for 5′UTR determination mapping the RNA reads to the inter-CDS region of the *1033* gene and the downstream gene at the *K. phaffii* GS115 chromosome 3 sequence. The beginning of the subregion with non-aligned RNA-reads was considered as the 3′UTR end of the *1033* gene. Moreover, the T_{1033} sequence was analyzed *in silico* to find A- and T-rich sequences described in the literature with a high number of occurrences in the sequences located downstream of the stop codon in yeast genes (van Helden et al. 2000).

Construction of *K. phaffii* KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033} strains

A synthetic DNA sequence harboring an *Aat*II restriction site, the P_{1033} sequence, the coding sequence for the *S. cerevisiae* alpha-factor prepro-secretion signal, based on *K. phaffii*-preferred codons (De Schutter et al. 2009) until the *Xho*I site, a *Not*I restriction site, 11 spacer nucleotides, the T_{1033} sequence, and the sequence from the 3' end of the T_{AOX1} until the *Bsu36*I site of the vector pPIC9 was synthesized, cloned into the vector pUCIDT-AMP and sequenced by Integrated DNA Technologies, Inc. (Coralville, IA, USA) to generate the plasmid pUCIDT$P_{1033}T_{1033}$. The liberated fragment from *Aat*II and *Xho*I digestion of the plasmid pUCIDT$P_{1033}T_{1033}$ was ligated into the vector pGAH-FTEII, previously digested by the same endonucleases, to produce the expression vector pP_{1033}-FTEII-T_{AOX1}. Similarly, vector pP_{1033}-FTEII-T_{1033} was constructed by *Not*I and *Bsu36*I digestion of the plasmid pUCIDT$P_{1033}T_{1033}$ and ligation into the vector pP_{1033}-FTEII-T_{AOX1}. The correct constructions of the two vectors were confirmed by PCR using PP1033 and FTE2 primers, and FTE1 and 3TH primers, directed to the P_{1033} and the *FTEII* gene or to the *FTEII* gene and a region of the linearized vector integrated into the yeast genome downstream of the transcriptional terminator sequence, respectively. All DNA manipulations were performed according to standardized methods (Green and Sambrook 2012).

K. phaffii KM71 cells were transformed with *Sal*I-linearized pP_{1033}-FTEII-T_{AOX1} or pP_{1033}-FTEII-T_{1033} DNA by electroporation according to the manufacturer's instructions (Thermo Fisher Scientific, Waltham, MA, USA). Transformants were plated onto RDB agar plates at 30 °C for subsequent screening by histidine prototrophy. The integration of the expression cassette at the *HIS4* locus of the *K. phaffii* genome in randomly selected His^+ colonies from the two transformations was verified by PCR using the PP1033 and FTE2 primers, and FTE1 and 3TH primers.

Transformant selection from KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033} constructions

Thirty-two and 20 transformants of each strain (KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033}, respectively) were grown in 10 mL of YPD medium at 30 °C and 250 rpm for 12–14 h until reaching an optical density at 600 nm (OD_{600}) of 8–10. The grown cells harvested by centrifugation (3000 g, 10 min, 4 °C) were used to inoculate 25 mL of BMG medium supplemented with 0.1% (w/v) $CaCl_2$, to an initial OD_{600} of 1.4. Further incubation was performed for 24 h at 30 °C and 250 rpm, with the addition of 10% (w/v) glycerol to a final concentration of 1% (w/v) after 14 h of culture. The biomass concentration was estimated based on 1.0 OD_{600} unit corresponding to 0.23 g dry cell weight (DCW) per liter (Herrera-Estala et al. 2022). The cell-free culture medium from each culture was recuperated by centrifugation (3000 g, 10 min, 4 °C), and the protein concentration was determined by the Bradford protein assay, using bovine serum albumin as the standard. The extracellular protein/biomass yield was calculated as the ratio of extracellular protein concentration to biomass concentration.

The three clones (His^+ transformants) of each strain that yielded the lowest extracellular protein/biomass yield were analyzed by quantitative PCR (qPCR) to determine the copy number of the integrated expression cassette in the yeast genome as described previously (Herrera-Estala et al. 2022). One transformant of each constructed strain with a single copy of the *FTEII* expression cassette was selected for further experiments.

P_{1033} and T_{1033} functionality analysis

The P_{1033} and T_{1033} functionalities on the heterologous gene were verified by detecting the *FTEII* transcripts using reverse transcription polymerase chain reaction (RT-PCR), and FTEII protein using SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and phytase activity.

Total RNA was isolated as described previously (Caballero-Pérez et al. 2021), from KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033} cells stored in RNAlater solution (Ambion, Grand Island, NY, USA), previously grown in BMG medium supplemented with 0.1% (w/v) $CaCl_2$, for 10 h at 30 °C and 250 rpm. The isolated RNA was treated with RQ1 RNase-free DNase (Promega, Madison, WI, USA). First-strand cDNA was synthesized by reverse transcription of the RNA using oligo$(dT)_{15}$ primer and M-MLV reverse transcriptase, according to the manufacturer's recommendations. PCR amplifications were performed using a PCR Multigene Mini Thermal Cycler (Labnet International Inc., Edison, NJ, USA) in a 25 µL reaction volume containing 0.5 µM of FTE1 and FTE2 primers, 0.2 mM dNTP's each, 1X GoTaq reaction buffer, 1.0 U GoTaq DNA polymerase, and 2 µL of primary cDNA. Detection of β-actin transcripts was used as a positive RT-PCR control with 5ACT and 3ACT primers in the PCR step. A 30-cycle amplification program was used: 95 °C for 1 min, 50 or 56 °C (*FTEII* or β-actin transcript detection, respectively) for 1 min, and 72 °C for 1 min, with a first denaturation step at 95 °C for 1 min, and a final extension step at 72 °C for 5 min. The amplified products were visualized by agarose gel electrophoresis.

Proteins from the cell-free culture medium at 24 h culture of each strain were analyzed by SDS-PAGE to assess the migration shift of the Endo H-treated proteins in a Coomassie blue-stained 12% SDS-polyacrylamide gel. The cell-free culture medium was previously concentrated 100-fold and desalted by ultrafiltration at 4 °C using 10-kDa Amicon Ultra-4 filters (Millipore, MA, USA). Concentrated samples were then incubated with Endo H for 1 h at 37 °C in accordance with the manufacturer's instructions.

Volumetric extracellular phytase activity was determined by measuring the phosphate released from sodium phytate as described previously (Guerrero-Olazarán et al. 2010) for the cell-free medium of 24 h culture, previously desalted using a PD-10 column (GE Healthcare Bio-Sciences Corp, Piscataway, NJ, USA) as described previously (Viader-Salvadó et al. 2010). One unit of phytase activity was defined as the amount of enzyme required to liberate 1 µmol of phosphate per min from sodium phytate under the assay conditions (pH 7.5, 37 °C).

Growth kinetics and activities of the regulatory elements (P_{1033}-T_{AOX1} and P_{1033}-T_{1033})

The two single-copy strains (KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033}) were grown in BMGly and BMGlc media supplemented with 0.1% (w/v) $CaCl_2$ (initial OD_{600} of 1.0) for 24 h at 30 °C and 250 rpm. Samples were taken every 3 h up to 6 h of culture and later every 6 h to measure growth by the turbidimetry assay at OD_{600}. Specific growth rates (µ) were determined by the slope of the natural log-linear regression of biomass concentration vs. time of the exponential phase (3 to 12 h) and statistically compared using a Student *t*-test with a significance level of 0.05.

The P_{1033} activities for the heterologous *FTEII* and endogenous *1033* genes in each single-copy strain were determined by reverse transcription-quantitative polymerase chain reaction (RT-qPCR) using the thermocycler Mx3005P QPCR system (Agilent Technologies, Santa Clara, CA, USA) and *YPT1* as a normalizer gene. The *GAPDH*-transcript levels were also determined as transcript-level references. Cell samples were harvested by centrifugation (3000 g, 10 min, 4 °C) at 6, 12, and 18 h of culture and stored in RNAlater solution (Ambion, Grand Island, NY, USA) until used. Total RNA was obtained from the cell samples, the cDNA was synthesized and transcripts were amplified as described

previously (Caballero-Pérez et al. 2021). PrimeTime qPCR Probe assays (Integrated DNA Technologies, Inc.) were used for the *FTEII* and *GAPDH* analyses (Herrera-Estala et al. 2022), and the SYBR Green DNA intercalator agent with 5q1033 and 3q1033 or 5qYPT1 and 3qYPT1 primers for the endogenous *1033* and *YPT1* gene expression level determinations.

Extracellular product/biomass yield ($Y_{p/x}$) was estimated for each culture at each sampling time, as the ratio of volumetric extracellular phytase activity (U per liter) to biomass concentration (g DCW per liter), which were estimated as above. The cell-free culture medium, obtained by centrifugation (3000 g, 10 min, 4 °C), was previously concentrated 5- to 50-fold and diafiltrated by ultrafiltration using 10-kDa Amicon Ultra-4 filters (Millipore, MA, USA) and 100 mM Tris-HCl (pH 8.5) buffer with 50 mM NaCl, 2% glycerol, and 5 mM $CaCl_2$.

The transcript levels and $Y_{p/x}$ values were statistically compared using a Student's t-test for independent samples and paired samples with a significance level of 0.05.

Results

P_{1033} and T_{1033} sequences

The P_{1033} sequence that was considered contained 176 nucleotides located upstream of the *1033* CDS. The TSS was identified at − 29 nucleotides upstream of the start codon (Fig. S1), the TATA-like box (TAAAAAGA) and the mammalian-type INR element (CTAATCG) were located at − 10 to − 17 and + 2 to + 8 nucleotides from the TSS, respectively (Fig. S2). The bioinformatics pipeline analysis of the P_{1033} sequence showed six TFBS and potential TF (i.e., Ecm22, Cat8, Rox1, Asg1, Yjl103c, and Stp2; Fig. S2).

The transcriptomic analysis showed that RNA reads aligned up to 107 nucleotides downstream of the *1033* CDS (Fig. S1), ending with a CAA that is a characteristic RNA cleavage and polyadenylation site for yeasts (Graber et al. 2002). The RNA cleavage and polyadenylation site was flanked by two A + T-rich sequences (ATATGTAT and ATTTATT) that are over-represented in downstream CDS sequences for yeasts (van Helden et al. 2000). Therefore, the considered T_{1033} sequence contained 157 nucleotides located downstream of the *1033* CDS.

Construction and selection of the *K. phaffii* KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033} strains

The expression vectors pP_{1033}-FTEII-T_{AOX1} and pP_{1033}-FTEII-T_{1033} were constructed harboring a DNA coding sequence for a beta-propeller phytase FTEII (Viader-Salvadó et al. 2010) as the reporter gene, in-frame with the *S. cerevisiae* alpha-factor prepro-secretion signal and between P_{1033} and T_{AOX1} or P_{1033} and T_{1033}, respectively (Fig. 1). PCR analysis of plasmid DNAs (pP_{1033}-FTEII-T_{AOX1} and pP_{1033}-FTEII-T_{1033}), and the genomic DNAs isolated from KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033} strains (His^+ transformants), showed the expected bands of 1,357 and 870 bp, which confirmed the expression plasmid construction and the integration of the expression cassette into the *K. phaffii* genome.

Cell-free culture media from 24 h cultures of 32 and 20 recombinant clones of each strain (KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033}, respectively), showed biomass concentrations ranging from 7.33 to 8.83 and 8.06 to 8.99 g DCW/L, extracellular protein concentrations from 5.93 to 17.20 and 5.31 to 8.52 mg/L, and extracellular protein/biomass yields from 0.7 to 2.0 and 0.6 to 1.0 mg/g, respectively (Fig. S3). The selected strain from each construction rendered the lowest protein production yield. The qPCR analysis of DNA from the selected strain of each construction confirmed the presence of one copy of the heterologous gene in the yeast genome.

P_{1033} and T_{1033} functionality

The RT-PCR assays for KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033} cells in BMG medium with $CaCl_2$

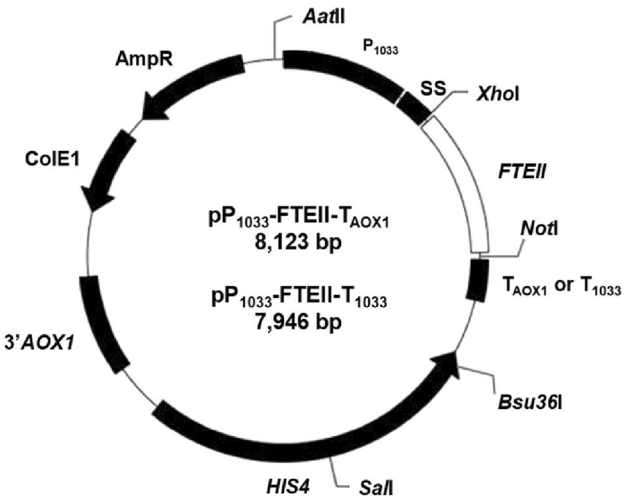

Fig. 1 pP_{1033}-FTEII-T_{AOX1} and pP_{1033}-FTEII-T_{1033} expression vectors. P_{1033}, promoter of the *1033* gene; SS, alpha-factor prepro-secretion signal coding sequence; *FTEII,* gene coding for the beta-propeller phytase FTEII; T_{AOX1}, *AOX1* transcriptional terminator; T_{1033}, transcriptional terminator of the *1033* gene; *HIS4*, *K. phaffii* wild-type gene coding for histidinol dehydrogenase; *3'AOX1*, *AOX1* downstream region; ColE1, *Escherichia coli* origin of replication; AmpR, ampicillin resistance gene; *Aat*II, *Xho*I, *Not*I and *Bsu36*I restriction sites for cloning P_{1033} and T_{1033} sequences; *Sal*I, restriction site for vector linearization before *K. phaffii* transformation by electroporation

showed the expected 536-bp band of an *FTEII*-transcript fragment (Fig. 2a and b).

The SDS-PAGE analysis for the cell-free culture medium of the KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033} cultures showed the characteristic smear for phytase FTEII ranging from 39 to over 66 kDa (Fig. 2c, lanes 2 and 5) that shifted to a defined band of 39 kDa (Fig. 2c, lanes 1 and 4) after N-deglycosylation by Endo H which corresponds to the theoretical molecular mass for the phytase FTEII based on its amino acid sequence (Viader-Salvadó et al. 2010). This result confirms that the phytase FTEII produced in the *K. phaffii* expression system is secreted as a highly N-glycosylated protein (Viader-Salvadó et al. 2010). The cell-free media from the KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033} cultures also showed phytase activities of 94.0 and 44.0 U/L, respectively.

These results (*FTEII*-transcript detection, phytase FTEII detection by SDS-PAGE, and enzyme activity) confirmed the correct functionality of the regulatory sequences and their combinations (i.e., P_{1033}-T_{AOX1} and P_{1033}-T_{1033}).

Growth kinetics

Figure 3a shows the growth kinetics for the two strains using glucose or glycerol as the carbon sources. Cell growth increased exponentially from 3 to 12 h of culture for the four kinetic assays with μ values ranging from 0.203 ± 0.029 to 0.257 ± 0.001 h^{-1}. No significant differences were seen in terms of μ between the same strain grown in glucose and glycerol or between different strains grown in the same carbon source. Although the four kinetic assays reached plateaus at 24 h of culture, the KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033} cultures in glucose yielded 1.29 and 1.35 times more biomass concentrations, respectively, compared to the cultures grown in glycerol.

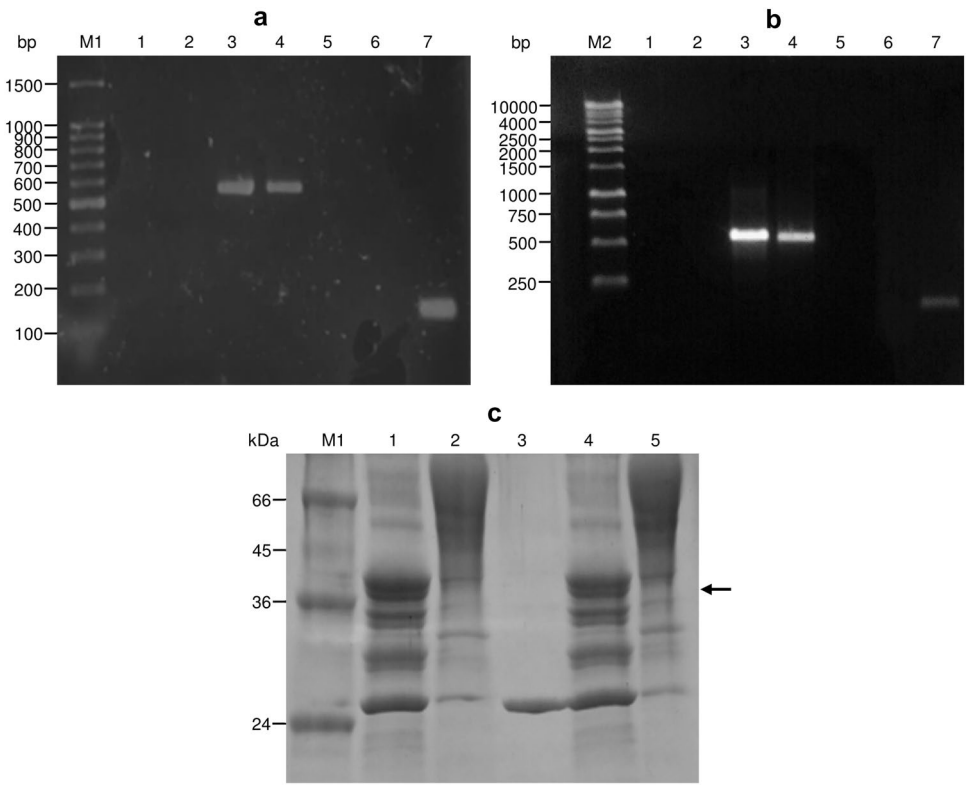

Fig. 2 Agarose gels for RT-PCR assays from RNA of KM71/P_{1033}-FTEII-T_{AOX1} cells (**a**) and KM71/P_{1033}-FTEII-T_{1033} cells (**b**). Lane M1, molecular size marker, 100 bp DNA ladder (Promega). Lane M2, molecular size marker, 1 kb DNA ladder (Promega). Lanes 1 to 4, assays with the FTE1 and FTE2 primers. Lanes 1, negative control of reverse transcriptase step (without reverse transcriptase); lanes 2, PCR negative control (without cDNA); lanes 3, PCR positive control (pP_{1033}-FTEII-T_{AOX1} plasmid); lanes 4, RT-PCR amplified product. Lanes 5 to 7, assays with the 5ACT and 3ACT primers. Lanes 5, negative control of reverse transcriptase step; lanes 6, PCR negative control; lanes 7, RT-PCR amplified product. **c** SDS-polyacrylamide gel of cell-free culture medium from BMG cultures, previously concentrated by ultrafiltration. Lane M, molecular mass marker. Lanes 1 and 2, proteins treated with and without Endo H glycosidase from the KM71/P_{1033}-FTEII-T_{AOX1} culture, respectively. Lane 3, Endo H glycosidase. Lanes 4 and 5, proteins treated with and without Endo H glycosidase from the KM71/P_{1033}-FTEII-T_{1033} culture, respectively. The arrow indicates N-deglycosylated recombinant FTEII

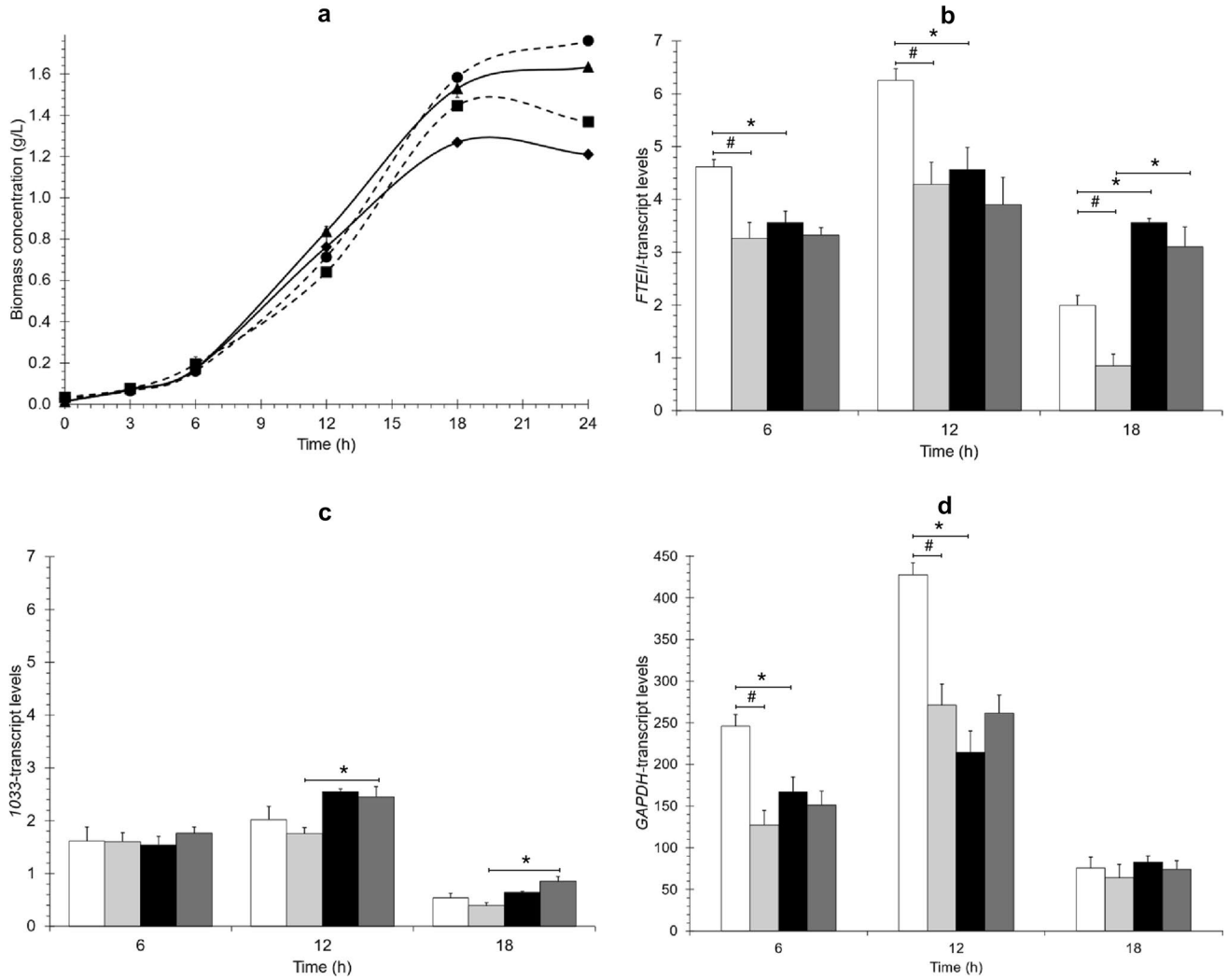

Fig. 3 Growth kinetics (**a**) in BMGlc (●, ▲) and BMGly (■, ♦) media for the KM71/P_{1033}-FTEII-T_{AOX1} (●, ■) and KM71/P_{1033}-FTEII-T_{1033} (▲, ♦) strains. Points represent the mean from three independent kinetics experiments (coefficient of variation less than 5%). Relative transcript levels of heterologous *FTEII* (**b**), *1033* (**c**), and *GAPDH* (**d**) genes for the KM71/P_{1033}-FTEII-T_{AOX1} (□, ■) and KM71/P_{1033}-FTEII-T_{1033} (▫, ▪) strains in BMGlc (□, ▫) and BMGly (■, ▪) media at different culture times. Data are presented as the mean ± standard error from three independent cultures. *Significant differences ($p < 0.05$) between glycerol- and glucose-grown cells from the same strain. #Significant differences ($p < 0.05$) between the two strains grown in the same carbon source

Activities of the regulatory elements (P_{1033}-T_{AOX1} and P_{1033}-T_{1033})

The transcript levels of the *FTEII* gene (i.e., the reporter gene), which is regulated by P_{1033}, were 1.4-times higher for glucose-grown KM71/P_{1033}-FTEII-T_{AOX1} cells at the exponential growth phase, compared to the KM71/P_{1033}-FTEII-T_{1033} cells (Fig. 3b). Nevertheless, when the two strains were grown in glycerol, no significant differences were seen between the *FTEII*-transcript levels at all sampling times. Moreover, the *FTEII* transcript-level profile for the KM71/P_{1033}-FTEII-T_{AOX1} strain was linked to cell growth in glucose. Although the KM71/P_{1033}-FTEII-T_{1033} strain showed an *FTEII* transcript-level profile linked to cell growth in glucose, it was not statistically significant. The *FTEII*-transcript levels were 1.3-times higher for glucose-grown, compared to glycerol-grown KM71/P_{1033}-FTEII-T_{AOX1} cells at the exponential phase, and no significant differences were seen for the KM71/P_{1033}-FTEII-T_{1033} cells in terms of the *FTEII*-transcript levels for cells grown in either of the carbon sources. The *FTEII* gene was more downregulated at the stationary phase for cells grown in glucose, compared to those grown in glycerol, since *FTEII*-transcript levels for KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033} cells were 1.8-times and 3.6-times greater, respectively, at the stationary phase when cells were grown in glycerol, compared to glucose.

Transcript levels of the endogenous *1033* gene did not show significant differences between the two strains grown in glucose or glycerol (Fig. 3c). The *1033*-transcript level profile was linked to cell growth in glycerol since the transcript levels increased significantly during the exponential growth phase (1.7-times and 1.4-times higher from 6 to 12 h of culture for KM71/P_{1033}-FTEII-T_{AOX1} and KM71P_{1033}-FTEII-T_{1033} cells, respectively). The same trend was observed in the *1033*-transcript levels for glucose-grown cells and for glycerol-grown cells between 6 and 12 h of culture, though it was not statistically significant. The *1033*-transcript levels were up to 1.4-times higher in glycerol-grown, compared to glucose-grown cells at 12 h of culture for both strains, though the increase was not statistically significant for the KM71/P_{1033}-FTEII-T_{AOX1} strain. The *1033* gene was downregulated in both strains when the carbon source was depleted at the stationary phase.

Transcript levels of the endogenous *GAPDH* gene did not show significant differences between the two strains grown with glycerol as the carbon source, but in glucose-grown cells, the *GAPDH*-transcript levels were up to 1.8-times higher at the exponential growth phase for the KM71/P_{1033}-FTEII-T_{AOX1} cells in comparison to the KM71/P_{1033}-FTEII-T_{1033} cells (Fig. 3d). The *GAPDH*-transcript level profile remained linked to cell growth for the two strains and carbon sources, with a transcript-level increase during the exponential growth phase and a decrease when the cells reached the stationary phase. The *GAPDH*-transcript levels of the KM71/P_{1033}-FTEII-T_{AOX1} strain at the exponential growth phase were 1.7-times higher for glucose-grown, compared to glycerol-grown cells, while *GAPDH*-transcript levels did not show significant differences for KM71/P_{1033}-FTEII-T_{1033} cells grown with either of the two carbon sources.

FTEII-transcript levels were on average 2.8- and 2.1-times higher, and 2.0- and 2.2-times higher than the *1033*-transcript levels for glucose-grown KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033} cells, and for glycerol-grown KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033} cells, respectively (Fig. 4a). During the stationary phase, the increased transcript-level ratio of *FTEII* to *1033* was enhanced in both carbon sources.

GAPDH-transcript levels were on average 50- and 60-times higher than *FTEII*-transcript levels in the glucose cultures, but in the glycerol cultures they were 34- and 38-times higher in KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033}, respectively (Fig. 4b). The *GAPDH*-transcript levels were on average 142- and 125-times higher than the *1033*-transcript levels for glucose-grown KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033} cells, respectively (Fig. 4c). For glycerol-grown KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033} cells, the transcript-level ratios of *GAPDH* to *1033* were 106 and 93, on average, respectively, at 6 h of culture.

$Y_{p/x}$ were, on average, 1.4-times higher for KM71/P_{1033}-FTEII-T_{AOX1} cultures, compared to KM71/P_{1033}-FTEII-T_{1033} cultures in glucose (Fig. 4d). This finding agrees with the higher (1.4-fold) *FTEII*-transcript levels of the KM71/P_{1033}-FTEII-T_{AOX1} strain, compared to the KM71/P_{1033}-FTEII-T_{1033} strain, both grown in glucose. Nevertheless, differences in the $Y_{p/x}$ were less pronounced among the glycerol-grown strains (1.1-times), which could be related to the absence of significant differences in the *FTEII*-transcript levels among the two strains grown in glycerol. Although the glucose-grown KM71/P_{1033}-FTEII-T_{AOX1} cells showed higher *FTEII*-transcript levels compared to glycerol-grown cells, the $Y_{p/x}$ was 1.3-fold higher in glycerol than in the glucose cultures. Although no significant differences were seen in the *FTEII*-transcript levels of the KM71/P_{1033}-FTEII-T_{1033} strain grown in glucose and glycerol, the $Y_{p/x}$ yield was 1.7-times higher in glycerol than in glucose cultures. These findings indicate that the translation and protein secretion pathway had a high impact on $Y_{p/x}$ when glycerol was used as a carbon source. The highest $Y_{p/x}$ in both strains occurred at the end of the exponential growth phase and the beginning of the stationary growth phase.

Discussion

Although a strong inducible or constitutive promoter is the first choice for expressing heterologous genes in *K. phaffii*, interest in weak promoters has recently increased due to their potential to produce host-toxic proteins (Ruth et al. 2010) or to simultaneously co-express several genes and maintain a metabolic balance in the host (Vogl et al. 2016; Jin et al. 2019; Duo et al., 2021). Few studies have been conducted on transcriptional terminators that can be used to tune gene expression, in comparison to the number of promoter studies (Ito et al. 2020). Therefore, in this work, we identified and functionally characterized for the first time the promoter (P_{1033}) and transcriptional terminator (T_{1033}) of a constitutive gene with weak, non-methanol-dependent transcriptional activity (i.e., the *1033* gene). This is the first report on the functional characterization of the promoter and terminal terminator of the *K. phaffii 1033* gene that expands the available toolbox for tuning gene expression in *K. phaffii*.

The 5′UTR and 3′UTR of the *1033* gene are not annotated in the *K. phaffii* genomes available in the databases. For this reason, we analyzed the RNA-seq data to locate subregions upstream and downstream of the *1033* CDS where the RNA reads were not aligned. In the upstream region, the region where RNA reads were not mapped had a length of 136 nucleotides on average, while the inter-CDS region between the *1033* gene and the adjacent upstream gene is annotated with 176 nucleotides. Since these two regions are similar in

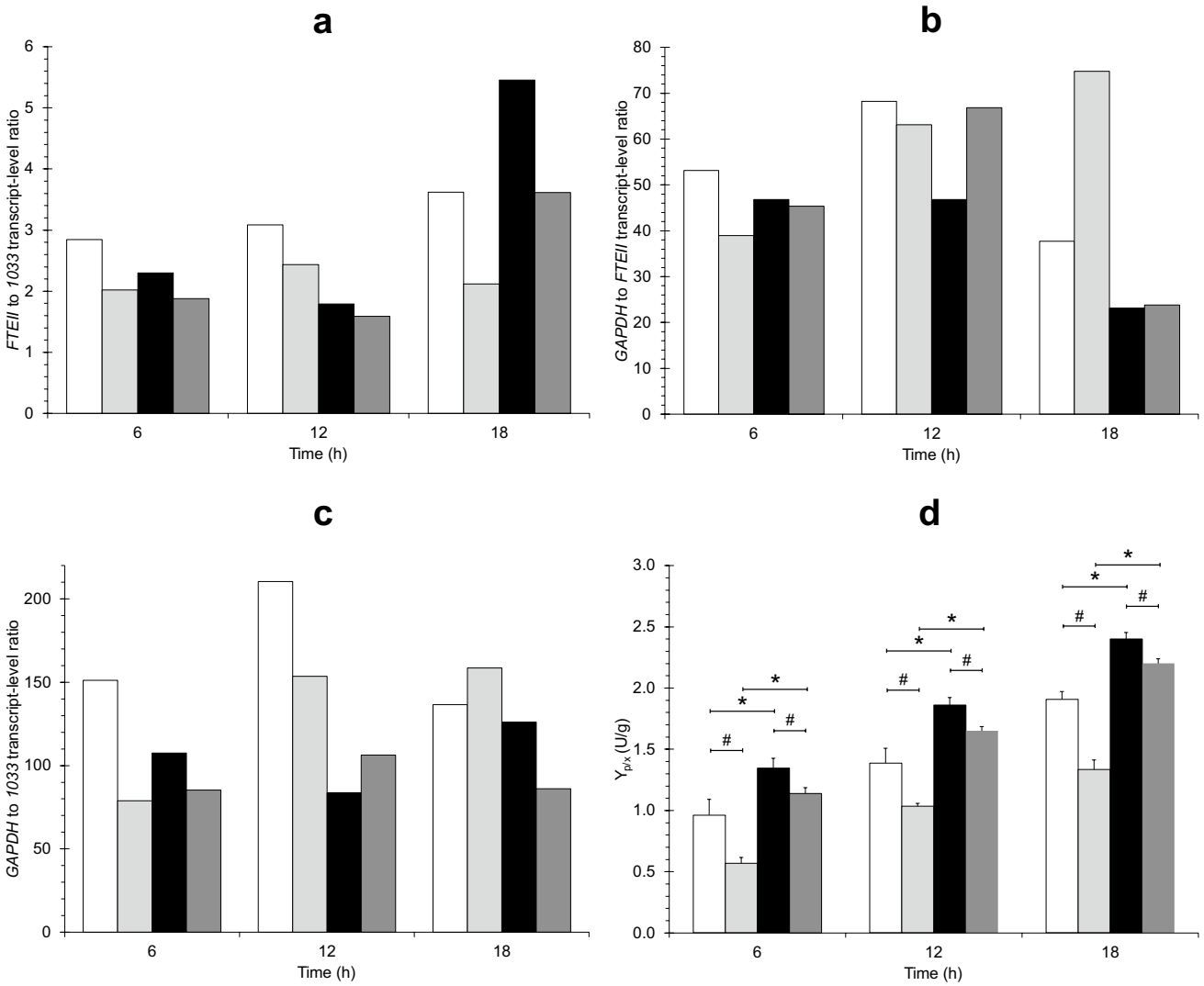

Fig. 4 *FTEII* to *1033* transcript-level ratio (**a**), and *GAPDH* to *FTEII* (**b**) or *1033* (**c**) transcript-level ratio for the KM71/P_{1033}-FTEII-T_{AOX1} (□, ■) and KM71/P_{1033}-FTEII-T_{1033} (▫, ▪) strains in BMGlc (□, ▫) and BMGly (■, ▪) media at different culture times. Data are presented as the mean from three independent cultures. (D) Extracellular product/biomass yields ($Y_{p/x}$) for the KM71/P_{1033}-FTEII-T_{AOX1} (□, ■) and KM71/P_{1033}-FTEII-T_{1033} (▫, ▪) strains in BMGlc (□, ▫) and BMGly (■, ▪) media at different culture times. Data are presented as the mean ± standard error from three independent cultures. *Significant differences ($p < 0.05$) between glycerol- and glucose-grown cells from the same strain. #Significant differences ($p < 0.05$) between the two strains grown in the same carbon source

size, we decided to use the sequence of the inter-CDS region as the P_{1033} sequence to be evaluated.

The P_{1033} contains a TATA-like sequence with two nucleotide changes from the TATA-box consensus sequence (TATAWAW), which is characteristic of constitutive promoters (Donczew and Hahn 2018). Moreover, the P_{1033} sequence harbors a mammalian-type INR element downstream of the TATA-like box. Mammalian-type INR sequences are present at, or adjacent to, the TSS of ~ 40% of yeast promoters (Yang et al. 2007). The putative TSS was located at a CA sequence on the coding strand, i.e., pyrimidine-purine (PyPu) dinucleotide at the − 1 and + 1 sites of the TSS, with two adenines at a region of 7 nucleotides immediately upstream of the CA sequence. This TSS nucleotide sequence has been described for yeast species that use the scanning model for transcription initiation (Lu and Lin 2021). The P_{1033} sequence also contains a homopolymeric stretch of 21 deoxythymidine nucleotides (Ts) in the coding strand. Poly (dA:dT) sequences are overabundant in eukaryotic genomes (Segal and Widom 2009) and they are considered as an upstream promoter element in constitutive yeast promoters that act bidirectionally to activate the transcription of two adjacent unrelated genes transcribed in opposite directions (Struhl 1985, 1986), as is the case for the *1033* gene and its adjacent upstream gene (i.e., *NOB1* gene).

Only 6 TFBS were identified in the P_{1033} sequence analysis, which is less than the 11 and 21 TFBS that we identified in the P_{AOX1} and P_{GAP} sequences, respectively, from the same *in silico* analysis (unpublished data). Among the putative 6 TF for P_{1033}, only one is shared between P_{1033} and P_{GAP} (i.e., Cat8-1). Cat8 is a zinc cluster transcriptional activator necessary for the derepression of a variety of genes under non-fermentative growth conditions. In *S. cerevisiae*, Cat8 is one of the most important TF that activates genes of the gluconeogenesis pathway (Schüller 2003; Turcotte et al. 2010). *K. phaffii* has two putative Cat8 homologs (i.e., Cat8-1 and Cat8-2) that are involved in activating the ethanol assimilation pathway (Barbay et al. 2021). Moreover, Cat8-1 is necessary for activating genes in the glyoxylate cycle that contribute to the gluconeogenesis pathway. The identification of this TFBS in the P_{1033} sequence could be related to the putative function of the encoded protein from the *1033* gene (i.e., D-amino-acid oxidase) that catalyzes the oxidation of neutral D-amino acids to α-keto acids, which can enter the gluconeogenesis, glycolysis, and Krebs cycle pathways.

The transcriptome analysis downstream of the *1033* CDS was used to define the 3′URT end of the *1033* gene (i.e., 107 nucleotides downstream from the *1033* CDS end). Therefore, we considered the sequence of the 3′UTR region with 50 additional nucleotides as the sequence of T_{1033} to be evaluated, following suggestions in the literature for yeast transcriptional terminators (Curran et al. 2013).

The low strength of P_{1033} was initially observed with the transformant cultures from the two constructed strains since the range in protein/biomass yield was 5- to 23-times lower than those previously obtained for similar cultures with P_{GAP}-T_{AOX1} clones using the same reporter gene (Herrera-Estala et al. 2022).

In previous works (Viader-Salvadó et al. 2010; Herrera-Estala et al. 2022) we saw that the expression cassette integration at the *HIS4* locus by electroporation could lead to multiple insertion events at this locus. A *K. phaffii* strain that contains multiple integrated copies of the expression cassette usually yields higher heterologous gene-transcript and protein levels than a single-copy strain (Cereghino and Cregg 2000; Looser et al. 2015; Mombeni et al. 2020). For this reason, for a better comparison of the effect of the two genetic rearrangements of the expression cassette on the gene expression levels, we evaluated the gene dosage of several transformants by qPCR and selected single-copy clones of the two constructed strains (KM71/P_{1033}-FTEII-T_{AOX1} and KM71/P_{1033}-FTEII-T_{1033}) to avoid the possible effect of gene dosage on the heterologous gene-transcript and protein levels and to compare it with the transcript levels of the endogenous *1033* and *GAPDH* genes in the two genetic rearrangements of the expression cassettes.

Generally, an intracellular reporter protein such as the green fluorescent protein is used to assess the strength of a promoter (Hartner et al. 2008; Qin et al. 2011; Prielhofer et al. 2013). Nevertheless, the fluorescent signal generated from this approach does not evaluate exclusively the promoter activity on the heterologous-gene transcription, otherwise, the transcription is measured together with the translation process. One of the greatest advantages of the *K. phaffii* expression system is its efficient protein secretion system. Hence, this host is often preferred for the extracellular production of recombinant proteins, which facilitates the downstream process. Consequently, we decided to evaluate the heterologous gene-transcript levels by RT-qPCR using an expression cassette harboring a reporter gene that codes for a protein secreted into the culture medium, and two combinations of promoter and transcriptional terminator pairs as regulatory DNA elements. With this approach, besides the promoter activity on the heterologous-gene transcription and the transcriptional terminator effect on transcript stability, we evaluated the impact of the regulatory elements on translation and protein secretion by measuring the extracellular product/biomass yield ($Y_{p/x}$).

The differences in the P_{1033} transcriptional activity on *FTEII* between the two strains grown in glucose were not seen for the *1033* gene. In contrast, *FTEII*- and *1033*-transcript levels did not show significant differences between the two strains in glycerol-grown cells. These results indicate that the T_{AOX1} and T_{1033} terminators did not contribute to differentiating the *FTEII*-transcript levels in glycerol-grown cells, but in glucose, T_{AOX1} increases the *FTEII*-transcript levels with respect to T_{1033}, which shows the influence of the transcriptional terminator in the expression cassette on the heterologous gene-transcript levels. These findings are similar to those described previously (Ramakrishnan et al. 2020; Ito et al. 2020; Herrera-Estala et al. 2022), where the T_{AOX1} activity was greater than other endogenous transcriptional terminators, which was said to be due to the higher stability of the 3′UTR region and a higher mRNA half-life of the T_{AOX1}. Our findings suggest that mRNA stability could also be dependent on the carbon source used for cell growth.

The endogenous *1033* gene was expressed with low transcript levels compared to the *GAPDH* gene, and its expression was modulated by cell growth and the carbon source. Both strains showed *1033*-transcript levels that were 1.4-times higher in glycerol- than in glucose-grown cells during the exponential growth phase. Similar P_{1033} transcriptional activity on the *1033* gene was seen for the two strains in the same carbon sources. In the stationary phase, the *1033*-transcript levels decreased, likely due to the absence of the carbon source. The decrease was more pronounced in glucose than in glycerol cultures. Although the *FTEII* and *1033* genes were driven by P_{1033} in both strains, the *FTEII*-transcriptional levels were higher than those of

the *1033* gene, being *FTEII* to *1033*-transcript level ratio highest in the KM71/P_{1033}-FTEII-T_{AOX1} strain compared to the KM71/P_{1033}-FTEII-T_{1033} strain in both carbon sources (Fig. 4). These findings indicate a competitive effect on *FTEII* and *1033* transcription when the same promoter was used two times in the same strain leading to a decrease on the transcript level of one of the genes, as was reported previously for P_{GAP}-driven heterologous genes and endogenous *GAPDH* gene (Dou et al. 2021). This phenomenon was also enhanced by the higher stability of the *FTEII* transcripts harboring the 3′UTR region from the *AOX1* gene than the *1033* transcripts harboring the 3′UTR region from the *1033* gene, as described previously for other 3′UTR regions (Ito et al. 2020).

Transcript levels of the endogenous *GAPDH* gene in the strain KM71/P_{1033}-FTEII-T_{AOX1} were different according to growth phase and carbon source, with higher transcript levels in glucose-grown cells than in glycerol-grown cells during the exponential phase. This is consistent with P_{GAP}, which has been traditionally cataloged as a strong constitutive promoter with greater activity at high specific growth rates (Looser et al. 2015) and in glucose cultures compared to glycerol cultures (Waterham et al. 1997). The *GAPDH*-transcript levels were higher in the KM71/P_{1033}-FTEII-T_{AOX1} strain than in the KM71/P_{1033}-FTEII-T_{1033} strain when cultures were grown in glucose. These results indicate that the presence of T_{AOX1} in the expression cassette renders a higher transcriptional activity for the endogenous *GAP*-promoter/*GAP*-terminator (P_{GAP}-T_{GAP}) combination in glucose cultures, in comparison to the T_{1033} in the expression cassette. The activity of P_{GAP} trans-regulatory elements in the glucose cultures was likely affected by the promoter-terminator combination of the expression cassette. Therefore, the heterologous gene regulation driven by the P_{1033}-T_{AOX1} combination in the expression cassette presented a lower competition with the endogenous *GAPDH* regulation (P_{GAP}-T_{GAP}) in glucose cultures, compared to the heterologous gene regulation driven by the P_{1033}-T_{1033} combination. In contrast, no differences in *GAPDH*-transcript levels in glycerol-grown cells were seen between the two strains, which indicates that the regulation driven by the two heterologous promoter–terminator combinations competed similarly with the *GAPDH* regulation. Thus, we speculate that the regulatory DNA elements (i.e., promoter and transcriptional terminator) of the *1033* gene are regulated similarly to those of the *GAPDH* gene in glucose cultures, affecting the *GAPDH* regulation when two copies of the promoter-terminator pair (i.e., endogenous and heterologous) of the *1033* gene are present in the yeast genome, as was reported for P_{GAP}-driven genes (Dou et al. 2021).

Although the highest heterologous gene transcription levels were attained in glucose-grown KM71/P_{1033}-FTEII-T_{AOX1} cells, the highest $Y_{p/x}$ values were obtained in the glycerol cultures for the two strains. These results agree with previous reports that concluded that transcript levels by themselves are not sufficient to predict protein expression levels (Liu et al. 2016; Dou et al. 2021). Our results indicate that the promoter-terminator pair along with the carbon source affect the translation and/or protein secretion pathway. We found similar results with a P_{GAP}-T_{AOX1} strain using the same reporter gene (Herrera-Estala et al. 2022), where the culture conditions that yielded low heterologous gene-transcript levels also yielded a high $Y_{p/x}$, which was correlated with an upregulation of the *KAR2* and *PSA1-1* genes of the secretion pathway rather than with increased heterologous gene transcription. Furthermore, a culture strategy of a glycerol batch phase followed by a glucose-fed batch phase for a *GAP*-promoter *K. phaffii* system only increased the $Y_{p/x}$ by 1.1-times, compared to the glycerol-glycerol strategy (Garcia-Ortega et al. 2013), even though P_{GAP} shows greater transcriptional activity in glucose than glycerol cultures (Waterham et al. 1997). Taken together, this data points to low heterologous gene-transcript levels along with glycerol cultures increase translation and/or protein secretion, at least for the reporter protein used in this work.

In conclusion, the findings indicate that the P_{1033} has 2–3% transcriptional activity of the P_{GAP} activity and is tunable by cell growth and the carbon source. The combinations of the regulatory DNA elements from the *K. phaffii 1033* and *AOX1* genes (i.e., P_{1033}-T_{AOX1} and P_{1033}-T_{1033} pairs) contributed to a differentiated transcriptional activity of the heterologous and endogenous genes that were dependent on carbon source, which shows the influence of the transcriptional terminator in the expression cassette on the heterologous and endogenous gene-transcript levels. The promoter-terminator pair and the carbon source affected the translation and/or protein secretion pathway. Moreover, low heterologous gene-transcript levels along with glycerol cultures increased translation and/or protein secretion. The P_{1033} in combination with the T_{AOX1} or T_{1033} could be used to produce host-toxic proteins, co-express genes of the secretory pathway, and maintain a metabolic balance in metabolic engineering or synthetic biology applications.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s11274-023-03682-5.

Acknowledgements The authors thank Glen D. Wheeler for his stylistic suggestions in the preparation of the manuscript. Y.R.-del-P. and A.L.H.-E. thank CONACYT for their fellowship.

Author contributions All authors contributed to the study conception and design. Material preparation and data collection were performed by Y.R.-del-P. and A.L.H.-E. All authors performed the data analysis and interpretation. The first draft of the manuscript was written by Y.R.-del-P., J.M.V.-S. and M.G.-O. All further improvements and the final version of the manuscript were prepared by J.M.V.-S. and M.G.-O. All authors read and approved the final manuscript.

Funding This work was supported by the Consejo Nacional de Ciencia y Tecnología (CONACYT), Mexico (Grant No CB-2016-286093) and the Universidad Autónoma de Nuevo León, Mexico (Grant Nos. CN1599-21 and CN604-22, PAICYT).

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests The authors declare no competing interests.

References

Barbay D, Mačáková M, Sützl L, De S, Mattanovich D, Gasser B (2021) Two homologs of the Cat8 transcription factor are involved in the regulation of ethanol utilization in *Komagataella phaffii*. Curr Genet 67:641–661. https://doi.org/10.1007/s00294-021-01165-4

Caballero-Pérez A, Viader-Salvadó JM, Herrera-Estala AL, Fuentes-Garibay JA, Guerrero-Olazarán M (2021) Buried Kex2 sites in glargine precursor aggregates prevent its intracellular processing in *Pichia pastoris* Mut^S strains and the effect of methanol-feeding strategy and induction temperature on glargine precursor production parameters. Appl Biochem Biotechnol 193:2806–2829. https://doi.org/10.1007/s12010-021-03567-z

Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol Rev 24:45–66. https://doi.org/10.1111/j.1574-6976.2000.tb00532.x

Curran KA, Karim AS, Gupta A, Alper HS (2013) Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. Metab Eng 19:88–97. https://doi.org/10.1016/j.ymben.2013.07.001

De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouzé P, Van de Peer Y, Callewaert N (2009) Genome sequence of the recombinant protein production host *Pichia pastoris*. Nat Biotechnol 27:561–566. https://doi.org/10.1038/nbt.1544

Donczew R, Hahn S (2018) Mechanistic differences in transcription initiation at TATA-less and TATA-containing promoters. Mol Cell Biol 38:e00448–e00417. https://doi.org/10.1128/MCB.00448-17

Dou W, Zhu Q, Zhang M, Jia Z, Guan W (2021) Screening and evaluation of the strong endogenous promoters in *Pichia pastoris*. Microb Cell Fact 20:156. https://doi.org/10.1186/s12934-021-01648-6

Garcia-Ortega X, Ferrer P, Montesinos JL, Valero F (2013) Fed-batch operational strategies for recombinant fab production with *Pichia pastoris* using the constitutive *GAP* promoter. Biochem Eng J 79:172–181. https://doi.org/10.1016/j.bej.2013.07.013

García-Ortega X, Cámara E, Ferrer P, Albiol J, Montesinos-Seguí JL, Valero F (2019) Rational development of bioprocess engineering strategies for recombinant protein production in *Pichia pastoris* (*Komagataella phaffii*) using the methanol-free *GAP* promoter. Where do we stand? N Biotechnol 53:24–34. https://doi.org/10.1016/j.nbt.2019.06.002

Graber JH, McAllister GD, Smith TF (2002) Probabilistic prediction of *Saccharomyces cerevisiae* mRNA 3′-processing sites. Nucleic Acids Res 30:1851–1858. https://doi.org/10.1093/nar/30.8.1851

Green MR, Sambrook J (2012) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York

Guerrero-Olazarán M, Rodríguez-Blanco L, Carreon-Treviño JG, Gallegos-López JA, Viader-Salvadó JM (2010) Expression of a *Bacillus* phytase C gene in *Pichia pastoris* and properties of the recombinant enzyme. Appl Environ Microbiol 76:5601–5608. https://doi.org/10.1128/AEM.00762-10

Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, Lin-Cereghino J, Kovar K, Cregg JM, Glieder A (2008) Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. Nucleic Acids Res 36:e76. https://doi.org/10.1093/nar/gkn369

Herrera-Estala AL, Fuentes-Garibay JA, Guerrero-Olazarán M, Viader-Salvadó JM (2022) Low specific growth rate and temperature in fed-batch cultures of a beta-propeller phytase producing *Pichia pastoris* strain under GAP promoter trigger increased *KAR2* and *PSA1-1* gene expression yielding enhanced extracellular productivity. J Biotechnol 352:59–67. https://doi.org/10.1016/j.jbiotec.2022.05.010

Hohenblum H, Gasser B, Maurer M, Borth N, Mattanovich D (2004) Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*. Biotechnol Bioeng 85:367–375. https://doi.org/10.1002/bit.10904

Ito Y, Terai G, Ishigami M, Hashiba N, Nakamura Y, Bamba T, Kumokita R, Hasunuma T, Asai K, Ishii J, Kondo A (2020) Exchange of endogenous and heterogeneous yeast terminators in *Pichia pastoris* to tune mRNA stability and gene expression. Nucleic Acids Res 48:13000–13012. https://doi.org/10.1093/nar/gkaa1066

Jin LQ, Jin WR, Ma ZC, Shen Q, Cai X, Liu ZQ, Zheng YG (2019) Promoter engineering strategies for the overproduction of valuable metabolites in microbes. Appl Microbiol Biotechnol 103:8725–8736. https://doi.org/10.1007/s00253-019-10172-y

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360. https://doi.org/10.1038/nmeth.3317

Kuersten S, Goodwin E (2003) The power of the 3′UTR: translational control and development. Nat Rev Genet 4:626–637. https://doi.org/10.1038/nrg1125

Liu Y, Beyer A, Aebersold R (2016) On the dependency of cellular protein levels on mRNA abundance. Cell 165:535–550. https://doi.org/10.1016/j.cell.2016.03.014

Looser V, Bruhlmann B, Bumbak F, Stenger C, Costa M, Camattari A, Fotiadis D, Kovar K (2015) Cultivation strategies to enhance productivity of *Pichia pastoris*: a review. Biotechnol Adv 33:1177–1193. https://doi.org/10.1016/j.biotechadv.2015.05.008

Lu Z, Lin Z (2021) The origin and evolution of a distinct mechanism of transcription initiation in yeasts. Genome Res 31(1):51–63. https://doi.org/10.1101/gr.264325.120

Matsuyama T (2019) Recent developments in terminator technology in Saccharomyces cerevisiae. J Biosci Bioeng 128:655–661. https://doi.org/10.1016/j.jbiosc.2019.06.006

Mattanovich D, Gasser B, Hohenblum H, Sauer M (2004) Stress in recombinant protein producing yeasts. J Biotechnol 113:121–135. https://doi.org/10.1016/j.jbiotec.2004.04.035

Mattanovich D, Jungo C, Wenger J, Dabros M, Maurer M (2014) Yeast suspension culture. In: Meyer H, Schmidhalter DR (eds) Industrial scale suspension culture of living cells. Wiley, Hoboken, NJ, pp 94–129

Mayr C (2019) What are 3'UTRs doing? Cold Spring Harb Perspect Biol 11:a034728. https://doi.org/10.1101/cshperspect.a034728

Mombeni M, Arjmand S, Siadat SOR, Alizadeh H, Abbasi A (2020) pMOX: a new powerful promoter for recombinant protein production in yeast *Pichia pastoris*. Enzyme Microb Technol 139:109582. https://doi.org/10.1016/j.enzmictec.2020.109582

Prielhofer R, Maurer M, Klein J, Wenger J, Kiziak C, Gasser B, Mattanovich D (2013) Induction without methanol: novel regulated promoters enable high-level expression in *Pichia pastoris*. Microb Cell Fact 12:1–10. https://doi.org/10.1186/1475-2859-12-5

Qin X, Qian J, Yao G, Zhuang Y, Zhang S, Chu J (2011) GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*. Appl Environ Microbiol 77:3600–3608. https://doi.org/10.1128/AEM.02843-10

Ramakrishnan K, Prattipati M, Samuel P, Sankaranarayanan M (2020) Transcriptional control of gene expression in *Pichia pastoris* by manipulation of terminators. Appl Microbiol Biotechnol 104:7841–7851. https://doi.org/10.1007/s00253-020-10785-8

Ruth C, Zuellig T, Mellitzer A, Weis R, Looser V, Kovar K, Glieder A (2010) Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in *Pichia pastoris*. Syst Synth Biol 4:181–191. https://doi.org/10.1007/s11693-010-9057-0

Schüller HJ (2003) Transcriptional control of nonfermentative metabolism in the yeast *Saccharomyces cerevisiae*. Curr Genet 43:139–160. https://doi.org/10.1007/s00294-003-0381-8

Sears IB, O'Connor J, Rossanese OW, Glick BS (1998) A versatile set of vectors for constitutive and regulated gene expression in *Pichia pastoris*. Yeast 14:783–790. https://doi.org/10.1002/(SICI)1097-0061(19980615)14:83.0.CO;2-Y

Segal E, Widom J (2009) Poly (dA:dT) tracts: major determinants of nucleosome organization. Curr Opin Struct Biol 19:65–71. https://doi.org/10.1016/j.sbi.2009.01.004

Struhl K (1985) Naturally occurring poly (dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. Proc Natl Acad Sci 82:8419–8423. https://doi.org/10.1073/pnas.82.24.8419

Struhl K (1986) Constitutive and inducible *Saccharomyces cerevisiae* promoters: evidence for two distinct molecular mechanisms. Mol Cell Biol 6:3847–3853. https://doi.org/10.1128/mcb.6.11.3847-3853.1986

Thomas-Chollier M, Hufton A, Heinig M, O'Keeffe S, Masri NE, Roider HG, Manke T, Vingron M (2011) Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. Nat Protoc 6:1860–1869. https://doi.org/10.1038/nprot.2011.409

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192. https://doi.org/10.1093/bib/bbs017

Turcotte B, Liang XB, Robert F, Soontorngun N (2010) Transcriptional regulation of nonfermentable carbon utilization in budding yeast. FEMS Yeast Res 10:2–13. https://doi.org/10.1111/j.1567-1364.2009.00555.x

van Helden JV, Olmo MLD, Pérez-Ortín JE (2000) Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. Nucleic Acids Res 28:1000–1010. https://doi.org/10.1093/nar/28.4.1000

Viader-Salvadó JM, Gallegos-López JA, Carreón-Trevino JG, Castillo-Galván M, Rojo-Domínguez A, Guerrero-Olazarán M (2010) Design of thermostable beta-propeller phytases with activity over a broad range of pHs and their overproduction by *Pichia pastoris*. Appl Environ Microbiol 76:6423–6430. https://doi.org/10.1128/AEM.00253-10

Vogl T, Glieder A (2013) Regulation of *Pichia pastoris* promoters and its consequences for protein production. N Biotechnol 30:385–404. https://doi.org/10.1016/j.nbt.2012.11.010

Vogl T, Sturmberger L, Kickenweiz T, Wasmayer R, Schmid C, Hatzl AM, Gerstmann MA, Pitzer J, Wagner M, Thallinger GG, Geier M, Glieder A (2016) A toolbox of diverse promoters related to methanol utilization: functionally verified parts for heterologous pathway expression in *Pichia pastoris*. ACS Synth Biol 5:172–186. https://doi.org/10.1021/acssynbio.5b00199

Ward N, Moreno-Hagelsieb G (2014) Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: how much do we miss? PLoS ONE 9:e101850. https://doi.org/10.1371/journal.pone.0101850

Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM (1997) Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene 186:37–44. https://doi.org/10.1016/s0378-1119(96)00675-0

Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E (2007) Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene 389:52–65. https://doi.org/10.1016/j.gene.2006.09.029