PROMOTER VARIANTS

TECHNICAL FIELD

The invention refers to an isolated artificial promoter, which is a functional variant or derivative of the carbon source regulatable pG1 promoter of *Pichia pastoris* identified by SEQ ID 1, which promoter is herein referred to as pG1-x that is characterized by specific promoter elements and features.

BACKGROUND

The methylotrophic yeast *Pichia pastoris* (syn. *Komagataella* sp.) is a well-established protein production host. Numerous strain engineering approaches for *P. pastoris* improved the productivity for various products and effort was also dedicated to promoters for production purposes (Prielhofer, R., M. Maurer, J. Klein, J. Wenger, C. Kiziak, B. Gasser & D. Mattanovich, (2013) Induction without methanol: novel regulated promoters enable high-level expression in *Pichia pastoris*. *Microb Cell Fact* **12**: 5). Gene promoters are key features for the expression of a gene of interest (GOI): transcription of RNA of a downstream (3´) GOI is driven by the upstream (5´) promoter sequence. RNA polymerase II (RNAPII) is responsible for transcription of mRNA in eukaryotes. RNAPII promoters consist of a core promoter and several cis-acting DNA elements: proximal promoter, enhancers, silencers and boundary/insulator elements. Yeast core promoters are typically located close (-75/+50 bp) to the main transcription initiation site, they frequently contain improper TATA boxes (up to 2 bases difference to the TATA consensus sequence) and lack promoter elements which are typically found in other organisms. Transcriptional regulation responds to different conditions and is conducted through by cis-acting elements and corresponding regulatory proteins (transcription factors (TFs)).

For biotechnological applications, promoters allowing either constitutive or regulated/inducible gene expression are used. Production processes utilizing *P. pastoris* favorably apply carbon source dependent promoters such as the methanol-inducible P_{AOX}. Thereby, the growth phase can be separated from the potentially burdening protein production phase. A set of promoters was recently reported (Prielhofer et al., 2013), which is also controlled by the carbon source, but does not

rely on methanol for induction: These promoters share the feature of repression by excess glycerol and induction by limiting glucose. pG1 (SEQ ID 1), the strongest out of these promoters, is fully induced below 0.05 g/L glucose; it natively controls the expression of fa high-affinity glucose transporter gene GTH1 . Glucose uptake characteristics are dependent on the presence of high and low affinity glucose transporters. Seventeen hexose transport (*HXT*) genes in *S. cerevisiae* (*HXT1-17*) are expressed depending on the glucose concentration, but only two *HXT* homologs are found in *P. pastoris* (PAS_chr1-4_0570 and PAS_chr2-1_0054, named PpHxt1 and PpHxt2). PpHxt1 was identified to be the major low-affinity transporter in *P. pastoris*, while high affinity glucose transport is facilitated by two other genes, namely PAS_chr3_0023 and PAS_chr1-3_0011 (*GTH1*, the gene controlled by pG1) Prielhofer et al., 2013).

While *S. cerevisiae* features a huge capacity of glucose uptake and (fermentative) glucose metabolism, *P. pastoris* has a lower glucose uptake rate and a respiratory metabolism of glucose. Furthermore, *P. pastoris* is able to take glucose at much lower extracellular concentrations than *S. cerevisiae* (K_M of high-affinity transporters in the µM range in *P. pastoris* vs. mM range in *S. cerevisiae*). The fundamental difference in glucose uptake behavior is also displayed at the transcriptional control of related genes and can also be seen in the evolved functions of transcriptional regulators e. g. *Pp*Aft1 and *Pp*Mxr1 (homolog of *Sc*Adr1).

P. pastoris promoter studies and random mutagenesis of P_{AOX1} and of the promoter of glyceraldehyde-3-phosphate dehydrogenase P_{GAP} resulted in libraries with promoter variants possessing different activities, altered induction behavior compared to the wild-type promoter and in the identification of several important transcription factor binding sites (TFBS) of P_{AOX1} (WO2006/089329 A2).

The pG1 promoter and fragments thereof are further described in WO2013/050551 A1.

WO2014067926A1 discloses the expression of a protein of interest employing specific leader sequences. The leader were used with various promoter. As an exemplary promoter, the pG1 promoter is used.

Struhl K. (Proceedings of the National Academy of Sciences of the United States of America 1982, 78(7):4461-4465) describes deletion mapping of the yeast *his3* promoter region. He concluds that the T-A-T-A box, a sequence in front of most eukaryotic genes is not sufficient for wild-type promotr function and suggests that the

yeast promoter appears to be more complex than a simple site of interaction between RNA polymerase and DNA.

Quandt et al. (Nucleic Acids Research 1995, 23(23)4878-4884) describe tools for detection of consensus matches in nucleotide sequence data to identify regulatory motifs based on sequence data analysis. A library of consensus patterns was created and potential sequence matches were detected using a software tool (MatInspector).

SUMMARY OF THE INVENTION

It is the object of the invention to provide improved regulatable promoters with respect to carbon source regulation and promoter strength. It is the further object to provide such promoter for enhanced POI production and/or POI production within a reduced time period.

The object is solved by the subject matter as claimed.

According to the invention there is provided an isolated and/or artificial pG1-x promoter, which is a functional variant of the carbon source regulatable pG1 promoter of *Pichia pastoris* identified by SEQ ID 1, which pG1-x promoter consists of or comprises at least a part of SEQ ID 1 with a length of at least 293 bp, characterized by the following promoter regions:

a) at least one core regulatory region comprising the nucleotide sequences SEQ ID 2 and SEQ ID 3; and

b) a non-core regulatory region, which is any region within the pG1-x promoter sequence other than the core regulatory region;

wherein the pG1-x promoter comprises at least one mutation in any of the promoter regions and a sequence identity of at least 80% in SEQ ID 2 and SEQ ID 3, and a sequence identity of at least 50% in any region other than SEQ ID 2 or SEQ ID 3; and further

wherein the pG1-x promoter is characterized by the same or an increased promoter strength and induction ratio as compared to the pG1 promoter, wherein

- the promoter strength is at least 1.1-fold increased in the induced state as compared to the pG1 promoter, and/or

- the induction ratio is at least 1.1-fold increased as compared to the pG1 promoter.

Specifically, the pG1 promoter of *Pichia pastoris* identified by SEQ ID 1 is any of SEQ ID 7, 8, or 9, more specifically SEQ ID 9 as used herein as a reference in the Examples.

Specifically, the pG1-x promoter is not any of the prior art promoter named pG1 (SEQ ID 264), or any of pG1a (SEQ ID 265), pG1b (SEQ ID 266), pG1c (SEQ ID 267), pG1d (SEQ ID 268), pG1e (SEQ ID 269), or pG1f (SEQ ID 270), as described in WO2013050551 A1.

According to a specific embodiment, the pG1-x promoter according to the invention is a carbon source regulatable promoter which is characterized by

- an at least 1.1-fold, or at least 1.2-fold, or at least 1.3-fold, or at least 1.4-fold, or at least 1.5-fold, or at least 1.6-fold, or at least 1.7-fold, or at least 1.8-fold, or at least 1.9-fold, or at least 2-fold, or at least 2.1-fold, or at least 2.2-fold, or at least 2.3-fold, or at least 2.4-fold, or at least 2.5-fold, or at least 2.6-fold, or at least 2.7-fold, or at least 2.8-fold increased, or at least 2.9-fold, or at least 3-fold, or at least 3.3-fold, or at least 3.5-fold, or at least 3.8-fold, or at least 4-fold, or at least 4,.5-fold, or at least 5-fold, or at least 5.5-fold, or at least 6-fold increased promoter strength in the induced state as compared to the pG1 promoter, and

- the capability of being carbon source regulated as determined by an induction ratio which is the same or higher as compared to the induction ratio achieved with the pG1 promoter.

According to a specific further embodiment, the pG1-x promoter according to the invention is a carbon source regulatable promoter which is characterized by

- the same or higher promoter strength in the induced state as compared to the pG1 promoter, and

- the capability of being carbon source regulated as determined by an induction ratio which is at least 1.1-fold, or at least 1.2-fold, or at least 1.3-fold, or at least 1.4-fold, or at least 1.5-fold, or at least 1.6-fold, or at least 1.7-fold, or at least 1.8-fold, or at least 1.9-fold, or at least 2-fold, or at least 2.1-fold, or at least 2.2-fold, or at least 2.3-fold, or at least 2.4-fold, or at least 2.5-fold, or at least 2.6-fold, or at least 2.7-fold, or at least 2.8-fold increased, or at least 2.9-fold, or at least 3-fold, or at least 3.3-fold, or at least 3.5-fold, or at least 3.8-fold, or at least 4-fold, or at least 4,.5-fold, or at least 5-fold, or at least 5.5-fold, or at least 6-fold increased as compared to the induction ratio achieved with the pG1 promoter.

According to a specific further embodiment, the pG1-x promoter according to the invention is a carbon source regulatable promoter which is characterized by

- an at least 1.1-fold, or at least 1.2-fold, or at least 1.3-fold, or at least 1.4-fold, or at least 1.5-fold, or at least 1.6-fold, or at least 1.7-fold, or at least 1.8-fold, or at least 1.9-fold, or at least 2-fold, or at least 2.1-fold, or at least 2.2-fold, or at least 2.3-fold, or at least 2.4-fold, or at least 2.5-fold, or at least 2.6-fold, or at least 2.7-fold, or at least 2.8-fold increased, or at least 2.9-fold, or at least 3-fold, or at least 3.3-fold, or at least 3.5-fold, or at least 3.8-fold, or at least 4-fold, or at least 4,.5-fold, or at least 5-fold, or at least 5.5-fold, or at least 6-fold increased promoter strength in the induced state as compared to the pG1 promoter, and

- the capability of being carbon source regulated as determined by an induction ratio which is at least 1.1-fold, or at least 1.2-fold, or at least 1.3-fold, or at least 1.4-fold, or at least 1.5-fold, or at least 1.6-fold, or at least 1.7-fold, or at least 1.8-fold, or at least 1.9-fold, or at least 2-fold, or at least 2.1-fold, or at least 2.2-fold, or at least 2.3-fold, or at least 2.4-fold, or at least 2.5-fold, or at least 2.6-fold, or at least 2.7-fold, or at least 2.8-fold increased, or at least 2.9-fold, or at least 3-fold, or at least 3.3-fold, or at least 3.5-fold, or at least 3.8-fold, or at least 4-fold, or at least 4,.5-fold, or at least 5-fold, or at least 5.5-fold, or at least 6-fold increased as compared to the induction ratio achieved with the pG1 promoter.

Specifically, the promoter strength is determined by the expression level of a protein of interest (POI), such as a model protein (e.g., Green Fluorescence Protein, GFP, including e.g., enhanced GFP, eGFP, Gene Bank Accession no. U57607), and/or the transcription rate, as compared to the pG1 promoter. The promoter strength of pG1-x is specifically at least 1.2-fold, or at least 1.3-fold, or at least 1.4-fold, or 1.5-fold, or at least 1.6-fold, or at least 1.7-fold, or at least 1.8-fold, or at least 1.9-fold, or at least 2-fold, or at least 2.1-fold, or at least 2.2-fold, or at least 2.3-fold, or at least 2.4-fold, or at least 2.5-fold, or at least 2.6-fold, or at least 2.7-fold, or at least 2.8-fold increased, or at least 2.9-fold, or at least 3-fold, or at least 3.5-fold, or at least 4-fold, or at least 4.5-fold, or at least 5-fold, or at least 5.5-fold, or at least 6-fold, or at least 6.5-fold, or at least 7-fold, or at least 7.5-fold, or at least 8-fold, or at least 8.5-fold, or at least 9-fold, or at least 9.5-fold, or at least 10-fold increased as compared for example to the pG1 promoter.

Herein, the pG1 promoter may serve as a reference or control to determine the improved promoter function. Such control pG1 promoter may be used in parallel

control experiments using the same host cell and expression system, or as internal control within the same host cell culture. Such control experiments to qualify the promoter function as compared to the pG1 promoter are preferably carried out in *P. pastoris* host cell cultures, in particular recombinant *P. pastoris* expressing a model protein, such as GFP or eGFP.

The pG1-x promoter induction specifically refers to induction of transcription, specifically including further translation and optional expression of said POI.

Said transcription rate is determined as a measure of the promoter strength and specifically refers to the amount of transcripts obtained upon fully inducing said promoter.

Said transcription rate may be determined by the transcription strength in the fully induced state, which is e.g., obtained under conditions of glucose-limited chemostat cultivations and expressed relative to the transcription rate of the pG1 promoter.

Preferably the transcription analysis is quantitative or semi-quantitative, preferably employing qRT-PCR, DNA microarrays, RNA sequencing and transcriptome analysis.

The promoter strength as compared to the pG1 promoter strength can be determined by the following standard assay: *P. pastoris* strains expressing eGFP under the control of the promoter to be tested are screened in 24- deep well plates at 25°C with shaking at 280 rpm with 2 mL culture per well. Glucose feed beads (6mm, Kuhner, CH) are used to generate glucose-limiting growth conditions. Cells are analysed for eGFP expression in the induced state (YP + 1 feed bead, for 20-28 hours).

Said promoter is considered as de-repressed and fully induced, if the culture conditions provide for about maximum induction, e.g. at glucose concentrations of less than 0.4 g/L, preferably less than 0.04 g/L, specifically less than 0.02 g/L. The fully induced promoter preferably shows a transcription rate of at least 20%, more preferred at least 30%, 40%, 50%, 60%, 70%, 80%, 90% and at least 100% or even higher transcription rate of at least 150% or at least 200% as compared to the native pGAP promoter. The transcription rate may, for example, be determined by the amount of transcripts of a reporter gene, such as eGFP, such as described in the Example section below, upon cultivating a clone in liquid culture. Alternatively, the transcription rate may be determined by the transcription strength on a microarray, where

microarray data show the difference of expression level between repressed and de-repressed state and a high signal intensity in the fully induced state as compared to a control.

Said native pGAP promoter specifically of is a promoter endogenous or homologous to the eukaryotic cell which may be used as a host cell to determine the expression of a POI, and serves as a standard or reference promoter for comparison purposes.

For example, a native pGAP promoter of *P. pastoris* is the unmodified, endogenous promoter sequence in *P. pastoris*, as used to control the expression of GAPDH in *P. pastoris*, e.g. having the sequence shown in Figure 7: native pGAP promoter sequence of *P. pastoris* (GS115) (SEQ ID 260). If *P. pastoris* is used as a host for producing a POI according to the invention, the transcription strength or rate of the pG1-x promoter according to the invention is compared to such native pGAP promoter of *P. pastoris*, and/or compared to the native pG1 promoter.

As another example, a native pGAP promoter of *S. cerevisiae* is the unmodified, endogenous promoter sequence in *S. cerevisiae*, as used to control the expression of GAPDH in *S. cerevisiae*. If *S. cerevisiae* is used as a host for producing a POI, the transcription strength or rate of the pG1-x promoter is compared to such native pGAP promoter of *S. cerevisiae.*

Therefore, the relative transcription strength or rate of a promoter according to the invention is usually compared to the native pGAP promoter of a cell of the same species or strain that is used as a host for producing a POI.

The induction ratio is a key parameter to determine the regulation of the present pG1-x promoter, and sets the promoter activity or strength in the induced state in relation to the promoter activity or strength in the repressed state. For example, the expression level of a model protein (e.g., GFP or eGFP) and/or the transcription rate in the repressed state is determined upon repression by excess glycerol, and the expression level of the model protein and/or the transcription rate is determined in the induced state upon induction by limiting glucose feeding.

Specifically, the induction ratio is determined by the ratio of expression level (e.g. GFP or eGFP) in the induced vs. the repressed state. The induction ratio of the pG1-x promoter is specifically the same or higher as compared to the pG1 promoter. In specific cases, the induction ratio is at least 2-fold, or at least 3-fold, or at least 4-fold,

at least 5-fold, or at least 6-fold, or at least 7-fold, at least 8-fold, or at least 9-fold, or at least 10-fold increased, as compared to the pG1 promoter.

The induction ratio as compared to the pG1 promoter strength can be determined by the following standard assay: *P. pastoris* strains expressing eGFP under the control of the promoter to be tested are screened in 24- deep well plates at 25°C with shaking at 280 rpm with 2 mL culture per well. Glucose feed beads (6mm, Kuhner, CH) are used to generate glucose-limiting growth conditions. Cells are analyzed for eGFP expression during repression (YP + 1% glycerol, exponential phase) and induction (YP + 1 feed bead, for 20-28 hours).

Specifically, the pG1-x promoter has a promoter activity or strength (e.g., transcriptional activity or transcription strength) in the de-repressed (induced) state, which is at least 2.5-fold, or at least 3 fold, or at least 4-fold, at least 5-fold, or at least 6-fold, or at least 7-fold, at least 8-fold, or at least 9-fold, or at least 10-fold higher than in the repressed state.

Specifically, the core regulatory region incorporates the nucleotide sequences SEQ ID 2 and SEQ ID 3, meaning that the sequences SEQ ID 2 and 3 are comprised in the pG1-x promoter sequence in any order, preferably in close proximity to each other, e.g. with up to 10, 20, 50 or 100 bp between the sequences SEQ ID 2 and 3.

Specifically, the SEQ ID 2 and/or SEQ ID 3 contain one or more transcription factor binding sites (TFBS).

Specifically, the SEQ ID 2 and SEQ ID 3 nucleotide sequences, each of which or both sequences together represents a TFBS or at least a part thereof which is considered functional being recognized by the respective transcription factor. Such SEQ ID 2 or SEQ ID 3 nucleotide sequence (or a functional variant thereof) is considered essential and is incorporated in the pG1-x promoter either in unmodified form or as a functional variant thereof with at least 80% sequence identity, or at least 85%, or at least 90%, or at least 95%, up to 100% sequence identity.

Specifically, the pG1-x promoter comprises a nucleotide sequence other than SEQ ID 2 and SEQ ID 3, which has at least 50% sequence identity to a corresponding region in the pG1 promoter, specifically, at least 60%, or at least 70%, or at least 80%, or at least 90% sequence identity in the core regulatory region or in the non-core regulatory region. Specifically, the nucleotide sequence within the core-regulatory region which is any other than SEQ ID 2 and SEQ ID 3 has at least at least 60%, or at least 70%, or at least 80%, or at least 90%, or at least 95%, or at least 98% sequence

identity to the corresponding region in the pG1 promoter. Specifically, the nucleotide sequence in the non-core regulatory region can have less than 90%, or less than 80%, or less than 70%, or less than 60% sequence identity to a corresponding region in the pG1 promoter.

Specifically, the core regulatory region comprises or consists of the nucleotide sequence SEQ ID 4, or a functional variant thereof comprising the TFBS, preferably a functional variant with at least 80%, or at least 90%, or at least 95%, or at least 98% sequence identity.

Specifically, the core regulatory region is incorporated into a main regulatory region represented by SEQ ID 5, or a functional variant thereof comprising the TFBS, preferably a functional variant with at least 80%, or at least 90%, or at least 95%, or at least 98% sequence identity.

Specifically, the one or more TFBS is a TFBS for any of the transcription factors selected from the group consisting of Rgt1, Cat8-1 and Cat8-2.

Specifically, the TFBS are recognized by the transcription factors Rgt1 and/or Cat8-1 and/or Cat8-2. TFBS are characterized by certain consensus sequences, which can vary for the same factor. The specific transcription factors are identified as follows:

Rgt1 is a glucose-responsive transcriptional activator and repressor and it regulates the expression of several glucose transporter (HXT) genes. Rgt1 of *P. pastoris* is characterized by the amino acid sequence SEQ ID 261 (Figure 7).

Cat8-1 and Cat8-2 are zinc cluster transcriptional activators binding to carbon source response elements, necessary for derepression of a variety of genes under non-fermentative growth conditions. Cat8-1 and Cat8-2 of *P. pastoris* are characterized by the amino acid sequences SEQ ID 262 and 263, respectively (Figure 7).

Specifically, the core regulatory region comprises a deletion of one or more nucleotides between the nucleotide sequences SEQ ID 2 and SEQ ID 3. The deletion may be one or more point mutations, and refer to 1, 2, 3, 4, 5, 6, 7, 8, or all 9 nucleotides positioned between SEQ ID 2 and SEQ ID 3.

Specifically, the core regulatory region comprises an insertion of one or more nucleotides between the nucleotide sequences SEQ ID 2 and SEQ ID 3. The insertion may be one or more point mutations, and refer to at least 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides positioned between SEQ ID 2 and SEQ ID 3.

Specifically, the core regulatory region comprises a substitution of one or more nucleotides between the nucleotide sequences SEQ ID 2 and SEQ ID 3. The

substitution may be one or more point mutations, and refer to 1, 2, 3, 4, 5, 6, 7, 8, or all 9 nucleotides positioned between SEQ ID 2 and SEQ ID 3.

Any of the specific deletions, insertions or substitutions may be combined to obtain the pG1-x promoter.

According to a specific aspect, the pG1-x promoter comprises at least two copies of the core regulatory region or the main regulatory region, either the original core regulatory region or the functional variant comprising at least one mutation. Specifically, the pG1-x promoter may comprise at least two, three or four copies of the core regulatory region and/or at least two, three or four copies of the main regulatory region.

According to a another specific aspect, the pG1-x promoter comprises at least two, three, four, five, six, seven or eight copies of the one or more TFBS selected from the group consisting of Rgt1, Cat8-1 and Cat8-2.

Specifically, the pG1-x promoter is an improved functional variant of the pG1 promoter comprising a deletion of one or more nucleotides at the 5'-end of the pG1 sequence, preferably leaving at least 280 nucleotides of the 3´ region of the pG1 sequence or a functional variant of the 3' region.

According to a specific embodiment, the pG1-x promoter comprises at least one or at least two T motifs identified by any of SEQ ID 12-29. The T motif specifically consists of any of

a) a sequence of contiguous T (thymine) which is herein referred to as T_n or $(T)_n$, preferably wherein n=13-20; preferably wherein the T motif is T14, T15, or T16;

b) a sequence characterized by A (adenine) at the first position, followed by a sequence of contiguous T (thymine), which is herein referred to as ATn or $A(T)_n$, preferably wherein n=13-20, in some cases preferably wherein n=13-22;

c) a sequence characterized by T (thymine) at the first position, and A (adenine) at the second position, followed by a sequence of contiguous T (thymine), which is herein referred to as TATn or $TA(T)_n$, preferably wherein n=13-20;

d) a sequence characterized by a sequence of contiguous T (thymine) and A (adenine) at the last position, which is herein referred to as TnA or $(T)_nA$, preferably wherein n=13-20;

e) a sequence characterized by a sequence of contiguous T (thymine) followed by A (adenine) at the last but one position, and T (thymine) at the last position, which is herein referred to as TnAT or $(T)_nAT$, preferably wherein n=13-20; or

d) a sequence of c) or e) wherein the A (adenine) is substituted by T (thymine), which is herein referred to as TTTn or TnTT or T(A/T)Tn or T(A/T)$(T)_n$, or Tn(A/T)T or $(T)_n$(A/T)T, preferably wherein n=13-20, e.g. resulting in a T motif which consists of a sequence of $(T)_n$ wherein n=15-22.

Any of the T motifs specified under a) to d) above may be combined in one promoter sequence e.g., such that the promoter sequence comprises one T motif which is a TA$(T)_n$ motif wherein n=13-20, and another T motif which is a $(T)_n$ motif, wherein n=13-22.

The T motif optionally comprises an extension, such that it is extended by one or more "A" (e.g., 1, 2, or 3 adenine) and optionally further extended by "T" (e.g., 1, 2, or 3 thymine) at the 3'-end and/or at the 5'-end of the T motif, which extension is herein also referred to as an extended T motif.

Herein the term "T motif" shall always include the T motif which is extended or not, thus, the term specifically includes both, the T motif that does not comprise the extension, or the extended T motif.

Specifically, the T motif comprises or consists of the nucleotide sequence which is any of SEQ ID 12-29. Any one, two, or more of the T motifs may be incorporated into the pG1-x promoter with or without the motif extension.

According to one specific aspect, the T motif extension is a "TA" sequence elongation at its 5'-end, to obtain a "TAT" 5'-end.

According to another specific aspect, the T motif extension is a "TAA" sequence elongation at its 5'-end, to obtain a "TAAT" 5'-end.

According to another specific aspect, the T motif extension is a "AT" sequence elongation at its 3'-end, to obtain a "TAT" 3'-end.

According to another specific aspect, the T motif extension is a "AAT" sequence elongation at its 3'-end, to obtain a "TAAT" 3'-end.

According to a specific aspect, the T motif is located upstream the core regulatory region, and optionally upstream the main regulatory region.

According to another specific aspect, the T motif is located downstream the core regulatory region, and optionally downstream the main regulatory region.

Specifically, the pG1-x promoter comprises a 3'-terminal nucleotide sequence incorporating at least part of a translation initiation site. A translation initiation site is specifically known as Kozak consensus sequence in eukaryotes, and a suitable sequence to support gene expression.

Specifically, the translation initiation site is

a) originating from the pG1 promoter and consists of or comprises the nucleotide sequence SEQ ID 6, or a functional variant thereof with at least 90% sequence identity; or

b) originating from any other promoter of *Pichia pastoris*, or a functional variant thereof with at least 90% sequence identity.

Exemplary alternative 3'-terminal promoter regions which can be used instead of the 3'-terminal region of the pG1 promoter, or instead of the nucleotide sequence SEQ ID 6, are e.g., derived from any of the following promoter: pAOX1, pAOX2, pDAS1, pDAS2, pFLD, pGAP, or pTEF2.

According to a specific embodiment, the promoter has a length up to 2000 bp. Specific pG1-x promoter have a length which is shorter than the pG1 promoter, such as with a length of at least 293 bp or 300 bp, or of at least 328 bp, or at least 350 bp or at least 400 bp, or at least 500 bp.

Specifically, the pG1-x promoter may comprise a sequence originating from a fragment of the pG1 promoter. According to a specific aspect, the pG1-x promoter is a variant or derivative of a parent fragment of pG1, which comprises at least the 3'-region of SEQ ID 1 which extends to at least 50%, or 60%, or 70%, or 80%, or at least 90% of the pG1 sequence.

Specifically, the pG1-x nucleotide sequence is derived from the pG1 promoter nucleotide sequence which comprises a deletion of or in the 5' terminal region, e.g. a cut-off of the nucleotide sequence at the 5' end, so to obtain a specific length with a range from the 3' end to a varying 5' end, such as with a length of the nucleotide sequence length of at least 293 bp or 300 bp, or of at least 328 bp, or at least 350 bp, or at least 400 bp, or at least 500 bp up to the length of the pG1 promoter fragment which comprises a deletion of at least 1, or at least 10, or at least 100 bp.

However, the promoter length can as well be increased, such as to obtain a length which is longer than the length of the pG1 promoter, specifically a length of up to 1500 bp, or up to 2000 bp. Specifically, the length may be within any of the ranges: 293 bp – 1500 bp, 293 bp – 2000 bp, 328 bp -1500 bp, or 328- 2000 bp.

According to a specific aspect, the invention provides for an isolated and/or artificial pG1-x promoter, comprising or consisting of the nucleotide sequence selected from the group consisting of any of

a) SEQ ID 37–44, preferably any of SEQ ID 45-76;

b) SEQ ID 77–80, preferably any of SEQ ID 81-112;

c) SEQ ID 113-114, preferably any of SEQ ID 115-130;

d) SEQ ID 131-132, preferably any of SEQ ID 133-148;

e) SEQ ID 149-150, preferably any of SEQ ID 151-166;

f) SEQ ID 167-168, preferably any of SEQ ID 169-184;

g) SEQ ID 185-186, preferably any of SEQ ID 187-202;

h) SEQ ID 203-204, preferably any of SEQ ID 205-220;

i) SEQ ID 221-222, preferably any of SEQ ID 223-238;

j) SEQ ID 239-240, preferably any of SEQ ID 241-256; and

k) SEQ ID 32-36 or SEQ ID 257-259;

or

l) a functional variant of any of a) – k) above, preferably, wherein the pG1-x promoter is characterized by the same or an increased promoter strength and induction ratio as compared to the pG1 promoter, wherein

- the promoter strength is at least 1.1-fold increased in the induced state as compared to the pG1 promoter, and/or

- the induction ratio is at least 1.1-fold increased as compared to the pG1 promoter..

A functional variant of such pG1-x promoter of a) – k) above is preferably characterized by any of the specific features as described for the functional variant of the pG1 promoter as described herein.

Specifically, the functional variant of any of the pG1-x promoter of a) – k) above, preferably a functional variant of any of SEQ ID 45-76, is characterized by one or more of the following features

a) the sequence is a functional variant of the promoter sequence of any of the pG1-x promoter of a) – k) above comprising a deletion of one or more nucleotides at the 5'-end of the promoter sequence, preferably leaving at least 280 nucleotides of the 3´ region of the promoter sequence or a functional variant of the 3' region, preferably comprising a 5' deletion of the promoter sequence of 50, 100, 150, 200, 250, or 300 nucleotides up to but not including the main regulatory region together with any

sequence downstream or 3' of said main regulatory region, in case of more than 1 main regulatory regions the 5'-end deletion of the promoter sequence is up to but not including the first or most 5' main regulatory region;

b) the sequence comprises one or more TFBS, preferably wherein the TFBS is for any of the transcription factors selected from the group consisting of Rgt1, Cat8-1, and Cat8-2;

c) the core regulatory region comprises the nucleotide sequence SEQ ID 4, or a functional variant thereof comprising one or more TFBS, preferably a functional variant with at least 80% sequence identity,

d) the core regulatory region is incorporated into a main regulatory region represented by SEQ ID 5, or a functional variant thereof comprising the TFBS, preferably a functional variant with at least 80% sequence identity;

e) the core regulatory region comprises a deletion of one or more nucleotides between the nucleotide sequences SEQ ID 2 and SEQ ID 3;

f) the sequence comprises at least two copies of the core regulatory region or of the main regulatory region;

g) the sequence further comprises at least one or at least two T motifs identified by any of SEQ ID 12-29; preferably wherein the T motif is located either upstream or downstream the core regulatory region, and optionally upstream or downstream the main regulatory region;

h) the sequence comprises a 3'-terminal nucleotide sequence comprising at least part of a translation initiation site;

i) the sequence is elongated to a length up to 2000 bp.

The invention further provides for the pG1-x promoter in the isolated form.

Specifically, the isolated pG1-x promoter nucleic acid is provided which comprises the pG1-x promoter as described herein, or a nucleic acid comprising the complementary sequence. Specifically, the complementary sequence is a sequence which hybridizes under stringent conditions to the pG1-x promoter.

Specifically, the nucleic acid is operably linked to a nucleotide sequence encoding a protein of interest (POI), which nucleic acid is not natively associated with the nucleotide sequence encoding the POI. The POI is specifically a heterologous polypeptide or protein.

Specifically, the nucleotide sequence further comprises a nucleotide sequence encoding a signal peptide enabling the secretion of the POI, preferably wherein nucleotide sequence encoding the signal peptide is located adjacent to the 5'-end of the nucleotide sequence encoding the POI.

Specifically, the signal peptide is selected from the group consisting of signal sequences from *S. cerevisiae* alpha-mating factor prepro peptide, the signal peptides from the *P. pastoris* acid phosphatase gene (PHO1) and the extracellular protein X (EPX1) (Heiss, S., V. Puxbaum, C. Gruber, F. Altmann, D. Mattanovich & B. Gasser, (2015) Multistep processing of the secretion leader of the extracellular protein Epx1 in *Pichia pastoris* and implications on protein localization. Microbiology).

Specifically, the POI is a eukaryotic protein, preferably a mammalian protein.

In specific cases, a POI is a multimeric protein, specifically a dimer or tetramer.

According to specific embodiments, the POI is a heterologous protein, preferably selected from therapeutic proteins, including antibodies or fragments thereof, enzymes and peptides, protein antibiotics, toxin fusion proteins, carbohydrate - protein conjugates, structural proteins, regulatory proteins, vaccines and vaccine like proteins or particles, process enzymes, growth factors, hormones and cytokines, or a metabolite of a POI, specifically including a cell metabolite of the recombinant cell culture that expresses a gene of interest under the transcriptional control of a promoter of the invention.

A specific POI is an antigen-binding molecule such as an antibody, or a fragment thereof. Among specific POIs are antibodies such as monoclonal antibodies (mAbs), immunoglobulin (Ig) or immunoglobulin class G (IgG), heavy-chain antibodies (HcAb's), or fragments thereof such as fragment-antigen binding (Fab), Fd, single-chain variable fragment (scFv), or engineered variants thereof such as for example Fv dimers (diabodies), Fv trimers (triabodies), Fv tetramers, or minibodies and single-domain antibodies like VH or VHH or V-NAR. Further antigen-binding molecules may be selected from (alternative) scaffold proteins such as e.g. engineered Kunitz domains, Adnectins, Affibodies, Anticalins, and DARPins. The term "scaffold" describes a multifaceted group of compact and stably folded proteins – differing in size, structure, and origin – that serve as a starting point for the generation of antigen-binding molecules. Inspired by the structure–function relationships of antibodies (immunoglobulins), such an alternative protein scaffold provides a robust, conserved

structural framework that supports an interaction site which can be reshaped for the tight and specific recognition of a given (bio)molecular target.

According to a specific embodiment, a fermentation product is manufactured using the POI, a metabolite or a derivative thereof.

The invention further provides for an expression construct comprising the nucleic acid as described herein, preferably an autonomously replicating vector or plasmid, or a vector or plasmid which integrates into the chromosomal DNA of a host cell.

Specifically, the expression construct comprises the pG1-x promoter, operably linked to a nucleotide sequence encoding a POI under the transcriptional control of said promoter, which promoter is not natively associated with the coding sequence of the POI. Specifically, the expression construct is a vector.

The invention further provides for a recombinant host cell which comprises the expression construct as described herein, preferably a eukaryotic cell, such as a mammalian, insect, yeast, filamentous fungi or plant cells, preferably a yeast or filamentous fungal cell, more preferably a yeast cell of the Saccharomyces or Pichia genus.

Specifically, the yeast is selected from the group consisting of Pichia, Candida, Torulopsis, Arxula, Hansenula, Yarrowia, Kluyveromyces, Saccharomyces, Komagataella, preferably a methylotrophic yeast.

A specifically preferred yeast is *Pichia pastoris, Komagataella pastoris*, *K. phaffii*, or *K. pseudopastoris*, such as e.g., any of the *P. pastoris* strains CBS 704, CBS 2612, CBS 7435, CBS 9173-9189, DSMZ 70877, X-33, GS115, KM71 and SMD1168.

According to a specific aspect, the recombinant host cell comprises multiple copies of the nucleic acid sequence, and/or multiple copies of the expression construct. For example, the recombinant cell comprises 2, 3, 4, or more copies (gene copy number, GCN).

The invention further provides for a stable culture of the recombinant host cell as described herein.

According to a specific embodiment, a cell is employed, which has a higher specific growth rate in the presence of a surplus of carbon source relative to conditions of limited carbon source.

The invention further provides for a method of producing a POI by culturing a recombinant host cell line as described herein, comprising the steps of

a) cultivating the cell line under conditions to express said POI, and

b) recovering the POI.

Specifically, said method is carried out under the transcriptional control of the carbon source regulatable pG1-x promoter, wherein said pG1-x promoter has at least one of the promoter strength and regulatable features improved as compared to the pG1 promoter.

According to a specific embodiment, the cell line is cultivated under batch, fed-batch or continuous cultivation conditions, and/or in media containing limited carbon substrate.

Specifically, the cultivation is performed in a bioreactor starting with a batch phase as the first step, followed by a fed-batch phase or a continuous cultivation phase as the second step.

Specifically, the host cells are grown in a carbon source rich medium during the phase of high growth rate (e.g. at least 50%, or at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99%, or up to the maximum growth rate) and producing the POI during a phase of low growth rate (e.g. less than 90%, preferably less than 80%, less than 70%, less than 60%, less than 50%, or less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, less than 3%, less than 2%, less than 1%, less than 0.5%, less than 0.4%, less than 0.3%, or less than 0.2% of the maximum growth rate) e.g. while limiting the carbon source, preferably by feeding a defined minimal medium.

Specifically, the POI is expressed under growth-limiting conditions, e.g. by cultivating the cell line at a growth rate of less than the maximal growth rate, typically less than 90%, preferably less than 80%, less than 70%, less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, less than 3%, less than 2%, less than 1%, less than 0.5%, less than 0.4%, less than 0.3%, or less than 0.2% of the maximum growth rate of the cells. Typically the maximum growth rate is individually determined for a specific host cell.

Specifically, the cultivation method comprises

a) a first step using a basal carbon source repressing the pG1-x promoter, followed by

b) a second step using no or a limited amount of a supplemental carbon source de-repressing or inducing the pG1-x promoter to induce production of the POI.

Specifically, the batch phase is performed until the basal carbon source that is initially added to the cell culture is consumed by the cell line. The dissolved oxygen (DO) spike method can be used to determine basal carbon source consumption during batch phase.

According to a specific embodiment, the batch phase is characterized by a continuous decrease in oxygen partial pressure (pO2) signal and wherein the end of the batch phase is characterized by an increase of pO2. Typically, while consuming the basal carbon source during the batch phase and without adding further carbon sources as typical for batch phases, the oxygen partial pressure (pO2) signal will continuously decrease until for example below 65% such as for example 30%.Upon consumption of the basal carbon source, the pO2 may increase to e.g. above 30% such as for example above 65%, or more indicating the appropriate time point to switch to the fed-batch system using feed medium to add further carbon source under carbon source limited conditions.

Specifically, the pO2 is decreased to less than 65% or less saturation during batch phase followed by an increase of above 65% or more saturation at the end of the batch. Specifically, the batch phase is performed until an increase of the oxygen partial pressure (pO2) signal above 65% saturation, specifically above any of 70%, 75%, 80%, or 85%.

Specifically, the batch phase is performed for around 20 to 36h.

The term “around” with respect to cultivation time shall mean +/-5% or +/-10%.

For example, the specific batch performance time of around 20 to 36h means a duration of 18 to 39.6h, specifically 19 to 37.8h.

According to a specific embodiment, the batch phase is performed using 40 to 50 g/L glycerol, specifically 45 g/L glycerol as a basal carbon source in batch media, and cultivation is performed at 25°C for around 27 to 30h, or at 30°C for around 23 to 36h, or at any temperature between 25°C and 30°C during a cultivation time of 23 to 36h. Lowering the glycerol concentration in the batch medium would decrease the length of the batch phase, while increasing the glycerol in the batch medium would

even prolong the batch phase. As an alternative to glycerol, glucose can be used, e.g. in about the same amounts.

In a typical system of cell culture and POI expression, wherein a batch phase is followed by a fed-batch phase, specifically, the cultivation in the fed-batch phase is performed for any of, around 15 to 80h, around 15 to 70h, around 15 to 60h, around 15 to 50h, around 15 to 45h, around 15 to 40h, around 15 to 35h, around 15 to 30h, around 15 to 35h, around 15 to 25h, or around 15 to 20h; preferably around 20 to 40h. Specifically, the cultivation in the fed-batch phase is performed for any of around 80h, around 70h, around 60h, around 55h, around 50h, around 45h, around 40h, around 35h, around 33h, around 30h, around 25h, around 20h, or around 15h.

Any such fed-batch cultivation of less than 120h or less than 100h or up to 80h, which results in a successful POI production thereby obtaining a high yield is herein referred to as "speed fermentation". Specifically, the volume specific product formation rate (rP) is the amount of product (mg) formed per Unit Volume (L) and Unit time (h) (mg (L h)$^{-1}$). Volume specific product formation rate is also called space time yield (STY) or volumetric productivity.

Specifically, the fed-batch cultivation is performed such that a space time yield of around 30 mg (L h)$^{-1}$ (meaning 30 mg (L h)$^{-1}$ +/-5% or +/-10%). Specifically a space time yield of around 30 mg (L h)$^{-1}$ is achieved within around 30h fed batch, specifically at least any of 27, 28, 29, 30, 31, 32, or 33 mg (L h)$^{-1}$ within less than any of 33h, 32h, 31h, 30h, 29h, 28h, 27h, 26h, or 25h fed batch time can be achieved.

Specifically, the batch phase is performed as a first step a), and the fed-batch phase is performed as a second step b).

Specifically, the second step b) employs a feed medium in a fed-batch phase that provides for the supplemental carbon source in a growth limiting amount to keep the specific growth rate within the range of 0.04 h^{-1} to 0.2 h^{-1}, preferably less than any of 0.2, 0.15, 0.1 h^{-1} or 0.15 h^{-1}.

Specifically, the method of batch and fed-batch cultivation employs a yeast host cell, e.g. a yeast of any of the Saccharomyces genus or Pichia genus or *Komagataella* genus, or yeast from a genus other than Pichia, such as from *K. lactis*, *Z. rouxii*, *P. stipitis*, *H. polymorpha*, or *Y. lipolytica,* preferably *Pichia pastoris* or *Komagataella pastoris*. Specifically, the yeast is used in a speed fermentation.

Specifically, the method of batch and fed-batch cultivation employs the pG1-x promoter which is any of SEQ ID 37–44, preferably any of SEQ ID 45-76. In particular, the pG1-x promoter is characterized by SEQ ID 39, preferably SEQ ID 49.

Specifically, the POI is produced at a transcription rate of at least 15% as compared to the native pGAP promoter of the cell.

According to a specific embodiment, the basal carbon source is different from the supplemental carbon source, e.g. quantitatively and/or qualitatively different. The quantitative difference may provide for the different conditions to repress or de-repress the promoter activity.

According to a further specific embodiment the basal and the supplemental carbon sources comprise the same type of molecules or carbohydrates, preferably in different concentrations. According to a further specific embodiment, the carbon source is a mixture of two or more different carbon sources.

Any type of organic carbon suitable used for eukaryotic cell culture may be used. According to a specific embodiment, the carbon source is a hexose, such as glucose, fructose, galactose or mannose, a disaccharide, such as saccharose, an alcohol, such as glycerol or ethanol, or a mixture thereof.

According to a specifically preferred embodiment, the basal carbon source is selected from the group consisting of glucose, glycerol, ethanol, or mixtures thereof, and complex nutrient material. According to a preferred embodiment, the basal carbon source is glycerol.

According to a further specific embodiment, the supplemental carbon source is a hexose such as glucose, fructose, galactose and mannose, a disaccharide, such as saccharose, an alcohol, such as glycerol or ethanol, or a mixture thereof. According to a preferred embodiment, the supplemental carbon source is glucose.

Specifically,

a) the basal carbon source is selected from the group consisting of glucose, glycerol, ethanol, a mixture thereof, and complex nutrient material; and

b) the supplemental carbon source is a hexose such as glucose, fructose, galactose or mannose, a disaccharide, such as saccharose, an alcohol, such as glycerol or ethanol, or a mixture of any of the foregoing.

Said cultivating steps specifically comprise cultivating the cell line in the presence of said carbon sources, thus, in a culture medium comprising said carbon sources, or in step b) also in the absence of a supplemental carbon source.

The de-repressing (or inducing) conditions suitably may be achieved by specific means. The second step b) optionally employs a feed medium that provides for no or the supplemental carbon source in a limited amount.

Specifically, the feed medium is chemically defined and methanol-free.

The feed medium may be added to the culture medium in the liquid form or else in an alternative form, such as a solid, e.g. as a tablet or other sustained release means, or a gas, e.g. carbon dioxide. Yet, according to a preferred embodiment the limited amount of a supplemental carbon source added to the cell culture medium, may even be zero. Preferably, under conditions of a limited carbon substrate, the concentration of a supplemental carbon source in the culture medium is 0-1 g/L, preferably less than 0.6 g/L, more preferred less than 0.3 g/L, more preferred less than 0.1 g/L, preferably 1-50 mg/L, more preferred 1-10 mg/L, specifically preferred 1 mg/L or even below, such as below the detection limit as measured with a suitable standard assay, e.g. determined as a residual concentration in the culture medium upon consumption by the growing cell culture.

In a preferred method, the limited amount of the supplemental source provides for a residual amount in the cell culture which is below the detection limit as determined in the fermentation broth at the end of a production phase or in the output of a fermentation process, preferably upon harvesting the fermentation product.

Specifically, the second step b) employs a feed medium that provides for the supplemental carbon source in a growth limiting amount to keep the specific growth rate within the range of 0.001 h^{-1} to 0.2 h^{-1}, preferably 0.005 h^{-1} to 0.15 h^{-1}.

FIGURES

Figure 1: pG1 sequence analysis for carbon source-related TFBS using MatInspector. pG1 (also referred to as P_{GTH1}), was initially amplified and cloned from position -965 to -1 (length of 965 bp, sequence is provided in **Figure 6** (SEQ ID 1, in particular SEQ ID 9 has been used). Numbers indicate TFBS which were selected for deletion (**listed in Table 2**). Associated matrix families are F$CSRE (carbon source response elements, striped boxes), F$ADR (Yeast metabolic regulator, dotted boxes), F$MGCM (Monomeric Gal4-class motifs, filled boxes) and F$YMIG (Yeast GC-Box Proteins, white boxes). Other TFBS might be affected by the deletions (matrix match detail information is given in **Table 1)**. The black dashed box indicates the main

regulatory region of pG1 which was identified by the screening of shortened pG1 variants. The asterisk indicates the position of the prominent TAT (position -390 to -374) motif which was also selected for deletion and for mutation. Alternative 5´-starts of the shortened pG1 promoter variants are labeled with arrows and the length of the corresponding variant.

Figure 2: Screening data of the shortened pG1 promoter variants

The geometric mean of the population´s specific eGFP fluorescence (fluorescence related to cell volume) is shown for clones expressing eGFP under control of pG1 (clone #8, verified GCN of 1) or a shortened pG1 variant (each 2 clones cultivated in triplicates, selected in pre-screenings) in repressing and inducing growth conditions. Non-expressing wild type *P. pastoris* cells were used as negative control. Samples were taken during the repressing pre-culture and after 24 and 48 hours induction with feed beads.

Figure 3: Screening data of the TFBSdeletion and -TAT mutation variants

The geometric mean of the population´s specific eGFP fluorescence (fluorescence related to cell volume) is shown for clones expressing eGFP under the control of pG1 (clone #8, verified GCN of 1) or a pG1 variant (up to 9 clones were pool cultivated in 3 wells) in repressing and inducing growth conditions. Wild type *P. pastoris* cells were used as negative control.

Figure 4: Screening data of the pG1 duplication variants

The geometric mean of the population´s specific eGFP fluorescence (fluorescence related to cell volume) is shown for clones expressing eGFP under the control of pG1 (clone #8, verified GCN of 1) or a pG1 variant (up to 9 clones were pool cultivated in 3 wells, selected in pre-screenings) in repressing and inducing growth conditions. Wild type *P. pastoris* cells were used as negative control.

Figure 5: Fed batch cultivation of pG1 and pG1 variants expressing eGFP

Relative eGFP fluorescence was measured from bioreactor samples (diluted to similar biomass densities) using a plate reader and is shown over the feed time (batch end set to 0) in batch (A) and fed batch cultivation (B). A clone expressing eGFP under control of pG1 (#8) was compared to clones expressing under control of a pG1 deletion variant (pG1-Δ2, SEQ ID 211), a TAT mutation (pG1-T16, SEQ ID 257, and a duplication (pG1-D1240) variant (SEQ ID 49).

Figure 6: pG1 and pG1-x promoter sequences

Fig. 6a: Reference sequences

Fig. 6b: Sequences of pG1-x promoter

Individual Sequence elements:

Position 8 (SEQ ID 2): *ATAAATGGA* (e.g. position -293 to -285 in SEQ ID 8):

Position 9 (SEQ ID 3): *CATATTTTTCCGGTT* (e.g. position -275 to-261 in SEQ ID 8)

Core region: (SEQ ID 4): *ATAAATGGACGCCTGCTCCATATTTTTCCGGTT* (e.g. position -293 to -261 in SEQ ID 8)

Main regulatory region: (SEQ ID 5): (e.g. position -328 to -211 in SEQ ID 8):

CCGGATAAGAGAATTTTGTTTGATTATCCGTTCGG*ATAAATGGACGCCTGC TCCATATTTTTCCGGTT*ATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATT ACGGATAATAC

3'-terminal nucleotide sequence (SEQ ID 6): TTCCACCCTT

Indications in sequences:

- Main regulatory region: **bold**
- Core regulatory region: ***bold, italic and underlined***, SEQ ID 2 and 3 ***double underlined***
- T motif: *italic and underlined*, may be optionally extended (at the 5'-terminal end of the T motif) by a preceding TA sequence, or (at the 3'-terminal end of the T motif) by a succeeding AT sequence
- 3'-terminal region: underlined with dotted line
- Region less relevant for promoter activity in the reference pG1 (P_{GTH1}) sequences: underlined with a dash-dot line: one or more nucleotides up to all nucleotides within the region ranging from the 5'-terminal end to -328 (region underlined in Fig. 6a with a dash-dot line) may be substituted, or deleted, or further nucleotides may be inserted within such region, however, preferred embodiments still comprise at least one T motif which is (T)n (n=13-20) with or without preceding A or TA nucleotides; or with or without succeeding A or AT nucleotides. Such a less relevant region which can be partially or fully deleted is the region ranging from the 5'-terminal end to the first or 5' main regulatory region (**bold**) in any one of SEQ ID 37 to SEQ ID 202; preferably, up to 50, 100, 150, 200, 250, 300, 320, or 325 nucleotides of the 5'-terminal end of any one of SEQ ID 37 to SEQ ID 202 can be deleted.
- Deletion: del (underlined)

$(T)_n$ (n=13-20) motifs: may be optionally extended at its 5' end, e.g. by "A" or "TA"; or at its 3' end, e.g. by "A" or "AT"

- $(T)_{13}$: SEQ ID 12: *TTTTTTTTTTTTT*
- $(T)_{14}$: SEQ ID 13: *TTTTTTTTTTTTTT*
- $(T)_{15}$: SEQ ID 14: *TTTTTTTTTTTTTTT*
- $(T)_{16}$: SEQ ID 15: *TTTTTTTTTTTTTTTT*
- $(T)_{17}$: SEQ ID 16: *TTTTTTTTTTTTTTTTT*
- $(T)_{18}$: SEQ ID 17: *TTTTTTTTTTTTTTTTTT*
- $(T)_{19}$: SEQ ID 18: *TTTTTTTTTTTTTTTTTTT*
- $(T)_{20}$: SEQ ID 19: *TTTTTTTTTTTTTTTTTTTT*

$TA(T)_n$ (n=13-20) motifs, may be optionally mutated to substitute the "A" at position 2 for a "T" (A/T)

- $TA(T)_{13}$: SEQ ID 20: *TATTTTTTTTTTTTT*
- $TA(T)_{13}$ (substituted A/T), SEQ ID 14 (see $(T)_{15}$): *TTTTTTTTTTTTTTT*
- $TA(T)_{14}$: SEQ ID 21: *TATTTTTTTTTTTTTT*
- $TA(T)_{14}$ (substituted A/T), SEQ ID 15 (see $(T)_{16}$): *TTTTTTTTTTTTTTTT*
- $TA(T)_{15}$: SEQ ID 22: *TATTTTTTTTTTTTTTT*
- $TA(T)_{15}$ (substituted A/T), SEQ ID 16 (see $(T)_{17}$): *TTTTTTTTTTTTTTTTT*
- $TA(T)_{16}$: SEQ ID 23: *TATTTTTTTTTTTTTTTT*
- $TA(T)_{16}$ (substituted A/T), SEQ ID 17 (see $(T)_{18}$): *TTTTTTTTTTTTTTTTTT*
- $TA(T)_{17}$: SEQ ID 24: *TATTTTTTTTTTTTTTTTT*
- $TA(T)_{17}$ (substituted A/T), SEQ ID 18 (see $(T)_{19}$): *TTTTTTTTTTTTTTTTTTT*
- $TA(T)_{18}$: SEQ ID 25: *TATTTTTTTTTTTTTTTTTT*
- $TA(T)_{18}$ (substituted A/T), SEQ ID 19 (see $(T)_{20}$): *TTTTTTTTTTTTTTTTTTTT*
- $TA(T)_{19}$: SEQ ID 26: *TATTTTTTTTTTTTTTTTTTT*
- $TA(T)_{19}$ (substituted A/T), SEQ ID 28 (i.e. $(T)_{21}$): *TTTTTTTTTTTTTTTTTTTTT*
- $TA(T)_{20}$: SEQ ID 27: *TATTTTTTTTTTTTTTTTTTTT*
- $TA(T)_{20}$ (substituted A/T), SEQ ID 29 (i.e. $(T)_{22}$): *TTTTTTTTTTTTTTTTTTTTTT*

Figure 7:

Native pGAP promoter sequence of *P. pastoris* (GS115) (SEQ ID 260)

#	Name	PAS*	PIPA*	GS115 description
pGAP	TDH3	PAS_chr2-1_0437	PIPA02510	Glyceraldehyde-3-phosphate dehydrogenase

*PAS: ORF name in *P. pastoris* GS115; PIPA: ORF name in *P. pastoris* type strain DSMZ70382

Figure 7 continued : Transcription factor sequences

Rgt1 (PAS_chr1-3_0233) (SEQ ID 261)

Cat8-2(PAS_chr4_0540) (SEQ ID 262)

Cat8-1(PAS_chr2-1_0757) (SEQ ID 263)

Figure 8 : Prior art sequences

pG1 (SEQ ID 264), pG1a (SEQ ID 265), pG1b (SEQ ID 266), pG1c (SEQ ID 267), pG1d (SEQ ID 268), pG1e (SEQ ID 269), or pG1f (SEQ ID 270), as described in WO2013050551 A1

Figure 9: Fed batch cultivation of the selected pG1-3 embodiment of SEQ ID 39 (pG1-D1240 (SEQ ID 49)) expressing an alternative scaffold protein as a model protein using (**A**) the standard fed batch protocol, (**B**) the space-time yield optimized fed batch protocol ("speed fermentation") adapted from Maurer et al. (Microbial Cell Factories, 2006, 5:37*)*

DETAILED DESCRIPTION OF THE INVENTION

Specific terms as used throughout the specification have the following meaning.

The term "carbon source" also referred as "carbon substrate" as used herein shall mean a fermentable carbon substrate, typically a source carbohydrate, suitable as an energy source for microorganisms, such as those capable of being metabolized by host organisms or production cell lines, in particular sources selected from the group consisting of monosaccharides, oligosaccharides, polysaccharides, alcohols including glycerol, in the purified form, in minimal media or provided in raw materials, such as a complex nutrient material. The carbon source may be used according to the invention as a single carbon source or as a mixture of different carbon sources.

A "basal carbon source" such as used according to the invention typically is a carbon source suitable for cell growth, such as a nutrient for eukaryotic cells. The

basal carbon source may be provided in a medium, such as a basal medium or complex medium, but also in a chemically defined medium containing a purified carbon source. The basal carbon source typically is provided in an amount to provide for cell growth, in particular during the growth phase in a cultivation process, for example to obtain cell densities of at least 5 g/L cell dry mass, preferably at least 10 g/L cell dry mass, or at least 15 g/L cell dry mass, e.g. exhibiting viabilities of more than 90% during standard sub-culture steps, preferably more than 95%.

According to the invention the basal carbon source is typically used in an excess or surplus amount, which is understood as an excess providing energy to increase the biomass, e.g. during the cultivation of a cell line with a high specific growth rate, such as during the growth phase of a cell line in a batch or fed-batch cultivation process. This surplus amount is particularly in excess of the limited amount of a supplemental carbon source (as used under growth-limited conditions) to achieve a residual concentration in the fermentation broth that is measurable and typically at least 10 fold higher, preferably at least 50 fold or at least 100 fold higher than during feeding the limited amount of the supplemental carbon source.

A "supplemental carbon source" such as used according to the invention typically is a supplemental substrate facilitating the production of fermentation products by production cell lines, in particular in the production phase of a cultivation process. The production phase specifically follows a growth phase, e.g. in batch, fed-batch and continuous cultivation process. The supplemental carbon source specifically may be contained in the feed of a fed-batch process. The supplemental carbon source is typically employed in a cell culture under carbon substrate limited conditions, i.e. using the carbon source in a limited amount.

A "limited amount" of a carbon source or a "limited carbon source" is herein understood to specifically refer to the type and amount of a carbon substrate facilitating the production of fermentation products by production cell lines, in particular in a cultivation process with controlled growth rates of less than the maximum growth rate. The production phase specifically follows a growth phase, e.g. in batch, fed-batch and continuous cultivation process. Cell culture processes may employ batch culture, continuous culture, and fed-batch culture. Batch culture is a culture process by which a small amount of a seed culture solution is added to a medium and cells are grown without adding an additional medium or discharging a culture solution during culture. Continuous culture is a culture process by which a medium is continuously added and

discharged during culture. The continuous culture also includes perfusion culture. Fed-batch culture, which is an intermediate between the batch culture and the continuous culture and also referred to as semi-batch culture, is a culture process by which a medium is continuously or sequentially added during culture but, unlike the continuous culture, a culture solution is not continuously discharged.

Specifically preferred is a fed-batch process which is based on feeding of a growth limiting nutrient substrate to a culture. The fed-batch strategy, including single fed-batch or repeated fed-batch fermentation, is typically used in bio-industrial processes to reach a high cell density in the bioreactor. The controlled addition of the carbon substrate directly affects the growth rate of the culture and helps to avoid overflow metabolism or the formation of unwanted metabolic byproducts. Under carbon source limited conditions, the carbon source specifically may be contained in the feed of a fed-batch process. Thereby, the carbon substrate is provided in a limited amount.

Also in chemostat or continuous culture as described herein, the growth rate can be tightly controlled.

The limited amount of a carbon source is herein particularly understood as the amount of a carbon source necessary to keep a production cell line under growth-limited conditions, e.g. in a production phase or production mode. Such a limited amount may be employed in a fed-batch process, where the carbon source is contained in a feed medium and supplied to the culture at low feed rates for sustained energy delivery, e.g. to produce a POI, while keeping the biomass at low specific growth rates. A feed medium is typically added to a fermentation broth during the production phase of a cell culture.

The limited amount of a carbon source may, for example, be determined by the residual amount of the carbon source in the cell culture broth, which is below a predetermined threshold or even below the detection limit as measured in a standard (carbohydrate) assay. The residual amount typically would be determined in the fermentation broth upon harvesting a fermentation product.

The limited amount of a carbon source may as well be determined by defining the average feed rate of the carbon source to the fermenter, e.g. as determined by the amount added over the full cultivation process, e.g. the fed-batch phase, per cultivation time, to determine a calculated average amount per time. This average feed rate is kept low to ensure complete usage of the supplemental carbon source by the cell

culture, e.g. between 0.6 g L^{-1} h^{-1} (g carbon source per L initial fermentation volume and h time) and 25 g L^{-1} h^{-1}, preferably between 1.6 g L^{-1} h^{-1} and 20 g L^{-1} h^{-1}.

The limited amount of a carbon source may also be determined by measuring the specific growth rate, which specific growth rate is kept low, e.g. lower than the maximum specific growth rate, during the production phase, e.g. within a predetermined range, such as in the range of 0.001 h^{-1} to 0.20 h^{-1}, or 0.005 h^{-1} to 0.20 h^{-1}, preferably between 0.01 h^{-1} and 0.15 h^{-1}.

Specifically, a feed medium is used which is chemically defined and methanol-free.

The term "chemically defined" with respect to cell culture medium, such as a minimal medium or feed medium in a fed-batch process, shall mean a cultivation medium suitable for the *in vitro* cell culture of a production cell line, in which all of the chemical components and (poly)peptides are known. Typically, a chemically defined medium is entirely free of animal-derived components and represents a pure and consistent cell culture environment.

The term "cell line" as used herein refers to an established clone of a particular cell type that has acquired the ability to proliferate over a prolonged period of time. The term "host cell line" refers to a cell line as used for expressing an endogenous or recombinant gene or products of a metabolic pathway to produce polypeptides or cell metabolites mediated by such polypeptides. A "production host cell line" or "production cell line" is commonly understood to be a cell line ready-to-use for cultivation in a bioreactor to obtain the product of a production process, such as a POI. The term "eukaryotic host" or "eukaryotic cell line" shall mean any eukaryotic cell or organism, which may be cultivated to produce a POI or a host cell metabolite. It is well understood that the term does not include human beings.

The term "cell culture" or "cultivation", also termed "fermentation", with respect to a host cell line is meant the maintenance of cells in an artificial, e.g., an *in vitro* environment, under conditions favoring growth, differentiation or continued viability, in an active or quiescent state, of the cells, specifically in a controlled bioreactor according to methods known in the industry.

When cultivating a cell culture using the culture media of the present invention, the cell culture is brought into contact with the media in a culture vessel or with substrate under conditions suitable to support cultivation of the cell culture. In certain embodiments, a culture medium as described herein is used to culture cells according

to standard cell culture techniques that are well-known in the art. In various aspects of the invention, a culture medium is provided that can be used for the growth of eukaryotic cells, specifically yeast or filamentous fungi.

Cell culture media provide the nutrients necessary to maintain and grow cells in a controlled, artificial and *in vitro* environment. Characteristics and compositions of the cell culture media vary depending on the particular cellular requirements. Important parameters include osmolality, pH, and nutrient formulations. Feeding of nutrients may be done in a continuous or discontinuous mode according to methods known in the art. The culture media used according to the invention are particularly useful for producing recombinant proteins.

Whereas a batch process is a cultivation mode in which all the nutrients necessary for cultivation of the cells are contained in the initial culture medium, without additional supply of further nutrients during fermentation, in a fed-batch process, after a batch phase, a feeding phase takes place in which one or more nutrients are supplied to the culture by feeding. The purpose of nutrient feeding is to increase the amount of biomass in order to increase the amount of recombinant protein as well. Although in most cultivation processes the mode of feeding is critical and important, the present invention employing the promoter of the invention is not restricted with regard to a certain mode of cultivation.

In certain embodiments, the method of the invention is a fed-batch process. Specifically, a host cell transformed with a nucleic acid construct encoding a desired recombinant POI, is cultured in a growth phase medium and transitioned to a production phase medium in order to produce a desired recombinant POI.

In another embodiment, host cells of the present invention are cultivated in continuous mode, e.g. a chemostat. A continuous fermentation process is characterized by a defined, constant and continuous rate of feeding of fresh culture medium into the bioreactor, whereby culture broth is at the same time removed from the bioreactor at the same defined, constant and continuous removal rate. By keeping culture medium, feeding rate and removal rate at the same constant level, the cultivation parameters and conditions in the bioreactor remain constant.

A stable cell culture as described herein is specifically understood to refer to a cell culture maintaining the genetic properties, specifically keeping the POI production level high, e.g. at least at a µg level, even after about 20 generations of cultivation, preferably at least 30 generations, more preferably at least 40 generations, most

preferred of at least 50 generations. Specifically, a stable recombinant host cell line is provided which is considered a great advantage when used for industrial scale production.

The cell culture of the invention is particularly advantageous for methods on an industrial manufacturing scale, e.g. with respect to both the volume and the technical system, in combination with a cultivation mode that is based on feeding of nutrients, in particular a fed-batch or batch process, or a continuous or semi-continuous process (e.g. chemostat).

The term "expression" or "expression system" or "expression cassette" refers to nucleic acid molecules containing a desired coding sequence and control sequences in operable linkage, so that hosts transformed or transfected with these sequences are capable of producing the encoded proteins or host cell metabolites. In order to effect transformation, the expression system may be included in a vector; however, the relevant DNA may also be integrated into the host chromosome. Expression may refer to secreted or non-secreted expression products, including polypeptides or metabolites.

"Expression constructs" or "vectors" or "plasmid" used herein are defined as DNA sequences that are required for the transcription of cloned recombinant nucleotide sequences, i.e. of recombinant genes and the translation of their mRNA in a suitable host organism. Expression vectors or plasmids usually comprise an origin for autonomous replication in the host cells, selectable markers (e.g. an amino acid synthesis gene or a gene conferring resistance to antibiotics such as zeocin, kanamycin, G418 or hygromycin), a number of restriction enzyme cleavage sites, a suitable promoter sequence and a transcription terminator, which components are operably linked together. The terms "plasmid" and "vector" as used herein include autonomously replicating nucleotide sequences as well as genome integrating nucleotide sequences.

The expression construct of the invention specifically comprises a promoter of the invention, operably linked to a nucleotide sequence encoding a POI under the transcriptional control of said promoter, which promoter is not natively associated with the coding sequence of the POI.

The term "heterologous" as used herein with respect to a nucleotide or amino acid sequence or protein, refers to a compound which is either foreign, i.e. "exogenous", such as not found in nature, to a given host cell; or that is naturally found in a given host cell, e.g., is "endogenous", however, in the context of a heterologous

construct, e.g. employing a heterologous nucleic acid. The heterologous nucleotide sequence as found endogenously may also be produced in an unnatural, e.g. greater than expected or greater than naturally found, amount in the cell. The heterologous nucleotide sequence, or a nucleic acid comprising the heterologous nucleotide sequence, possibly differs in sequence from the endogenous nucleotide sequence but encodes the same protein as found endogenously. Specifically, heterologous nucleotide sequences are those not found in the same relationship to a host cell in nature. Any recombinant or artificial nucleotide sequence is understood to be heterologous. An example of a heterologous polynucleotide is a nucleotide sequence not natively associated with the promoter according to the invention, e.g. to obtain a hybrid promoter, or operably linked to a coding sequence, as described herein. As a result, a hybrid or chimeric polynucleotide may be obtained. A further example of a heterologous compound is a POI encoding polynucleotide operably linked to a transcriptional control element, e.g., a promoter of the invention, to which an endogenous, naturally-occurring POI coding sequence is not normally operably linked.

The term "variant" as used herein in the context of the present invention shall refer to any sequence with a specific sequence identity or homology to a comparable parent sequence. A variant is specifically any sequence derived from a parent sequence e.g., by size variation, such as (terminal or non-terminal, such as "interstitional" i.e. with deletions or insertions within the nucleotide sequence) elongation, or fragmentation, mutation, hybridization (including combination of sequences).

The pG1-x promoter as described herein is specifically an artificial variant of the native (wild-type) pG1 promoter. Though there is a certain degree of sequence identity to the native structure, it is well understood that the materials, methods and uses of the invention, e.g. specifically referring to isolated nucleic acid sequences, amino acid sequences, expression constructs, transformed host cells and recombinant proteins, are "man-made" or synthetic, and are therefore not considered as a result of "law of nature".

The promoter herein referred to as "pG1-x promoter" is a variant of the pG1 promoter and its nucleotide sequence may be produced by mutagenesis of the pG1 promoter which is used as a "parent" sequence for producing a variant. A pG1-x promoter includes a promoter comprising two, three, four or more copies of SEQ ID 2, SEQ ID 3, SEQ ID 4 or SEQ ID 5.

A series of pG1-x promoters is e.g., exemplified by the promoter comprising or consisting of any of the sequences exemplified in Figure 6b, in particular any of the following sequences:

a) SEQ ID 37–44, preferably any of SEQ ID 45-76;

b) SEQ ID 77–80, preferably any of SEQ ID 81-112;

c) SEQ ID 113-114, preferably any of SEQ ID 115-130;

d) SEQ ID 131-132, preferably any of SEQ ID 133-148;

e) SEQ ID 149-150, preferably any of SEQ ID 151-166;

f) SEQ ID 167-168, preferably any of SEQ ID 169-184;

g) SEQ ID 185-186, preferably any of SEQ ID 187-202;

h) SEQ ID 203-204, preferably any of SEQ ID 205-220;

i) SEQ ID 221-222, preferably any of SEQ ID 223-238;

j) SEQ ID 239-240, preferably any of SEQ ID 241-256; and

k) SEQ ID 32-36 or SEQ ID 257-259.

A pG1-x promoter also includes 3' fragments of any one of SEQ ID 37 to SEQ ID 202 wherein part or all of the 5'-terminal end up to the first or 5' main regulatory region has been deleted; preferably, up to 50, 100, 150, 200, 250, 300, 320, or 325 nucleotides of the 5'-terminal end of any one of SEQ ID 37 to SEQ ID 202 is deleted.

The pG1-x promoter is characterized by having the same or an increased promoter strength and induction ratio as compared to the pG1 promoter, wherein

- the promoter strength is at least 1.1-fold increased in the induced state as compared to the pG1 promoter, and/or

- the induction ratio is at least 1.1-fold increased as compared to the pG1 promoter.

Further pG1-x variants are feasible e.g., using the exemplified pG1-x promoter of Figure 6b, or size variants, in particular elongated variants or fragments thereof, as "parent" sequences to produce variants by mutagenesis of certain regions, in particular such, that the essential elements and functions of the promoter be maintained or even improved. The pG1-x promoter variants may e.g., be derived from any of the exemplified pG1-x promoter sequences by mutagenesis to produce sequences suitable for use as a promoter in recombinant cell lines. Such variant promoter may be obtained from a library of mutant sequences by selecting those library members with predetermined properties. Variant promoters may have the same or even improved properties, e.g. improved in the promoter strength, the induction of POI production,

with increased differential effect under repressing and de-repressing conditions (in particular the induction ratio). The variant promoter may also comprise a nucleotide sequence from analogous sequences, e.g. from eukaryotic species other than *Pichia pastoris* or from a genus other than Pichia, such as from *K. lactis*, *Z. rouxii*, *P. stipitis*, *H. polymorpha*.

The term "functionally active" as used herein with respect to e.g., a promoter variant, the pG1-x promoter or variant of a pG1-x promoter as described herein or variant of the pG1 promoter, means a variant sequence resulting from modification of a parent sequence by mutagenesis, specifically by insertion, deletion or substitution of one or more nucleotides within the sequence or at either or both of the distal ends of the sequence, and which modification does not affect (in particular impair) the activity of this sequence. Regarding the pG1-x promoter as described herein, the function and activity is specifically characterized by the promoter activity and strength as well as the induction ratio.

Functionally active promoter variants as described herein are specifically characterized by exhibiting substantially the same promoter activity as the pG1 promoter (+/-10%, or +/- 5%), or even higher.

Functionally active promoter variants as described herein are specifically characterized by exhibiting substantially the same regulatable properties as the pG1 promoter e.g., measured by the induction ratio (+/- 10%, or +/- 5%), or an even higher induction ratio.

The term "promoter" as used herein refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. Promoter activity may be assessed by its transcriptional efficiency. This may be determined directly by measurement of the amount of mRNA transcription from the promoter, e.g. by Northern Blotting or indirectly by measurement of the amount of gene product expressed from the promoter.

The pG1-x promoter as described herein specifically initiates, regulates, or otherwise mediates or controls the expression of a coding DNA. Promoter DNA and coding DNA may be from the same gene or from different genes, and may be from the same or different organisms.

The pG1-x promoter as described herein is specifically understood as a regulatable promoter, in particular a carbon source regulatable promoter with different promoter strength in the repressed and induced state.

The strength of the promoter of the invention specifically refers to its transcription strength, represented by the efficiency of initiation of transcription occurring at that promoter with high or low frequency. The higher transcription strength the more frequently transcription will occur at that promoter. Promoter strength is important, because it determines how often a given mRNA sequence is transcribed, effectively giving higher priority for transcription to some genes over others, leading to a higher concentration of the transcript. A gene that codes for a protein that is required in large quantities, for example, typically has a relatively strong promoter. The RNA polymerase can only perform one transcription task at a time and so must prioritize its work to be efficient. Differences in promoter strength are selected to allow for this prioritization.

According to the invention the regulatable promoter is relatively strong in the fully induced state, which is typically understood as the state of about maximal activity.

The relative strength is commonly determined with respect to a comparable promoter, such as the pG1 promoter, or a standard promoter, such as the respective pGAP promoter of the cell as used as the host cell. The frequency of transcription is commonly understood as the transcription rate, e.g. as determined by the amount of a transcript in a suitable assay, e.g. RT-PCR or Northern blotting. For example, the transcription strength of a promoter according to the invention is determined in the host cell which is *P. pastoris* and compared to the native pGAP promoter of *P. pastoris*.

The strength of a promoter to express a gene of interest is commonly understood as the expression strength or the capability of support a high expression level/rate. For example, the expression and/or transcription strength of a promoter of the invention is determined in the host cell which is *P. pastoris* and compared to the native pGAP promoter of *P. pastoris*.

The comparative transcription strength employing the pGAP promoter as a reference (standard) may be determined by standard means, such as by measuring the quantity of transcripts, e.g. employing a microarray, or else in a cell culture, such as by measuring the quantity of respective gene expression products in recombinant cells. An exemplary test is illustrated in the Examples section.

In particular, the transcription rate may be determined by the transcription strength on a microarray, or with quantitative real time PCR (qRT-PCR) where microarray or qRT-PCR data show the difference of expression level between conditions with high growth rate and conditions with low growth rate, or conditions

employing different media composition, and a high signal intensity as compared to the native pGAP promoter.

The expression rate may, for example, be determined by the amount of expression of a reporter gene, such as eGFP.

The pG1-x promoter as described herein exerts a relatively high transcription strength, reflected by a transcription rate or transcription strength of at least 15% as compared to the native pGAP promoter in the host cell, sometimes called "homologous pGAP promoter". Preferably the transcription rate or strength is at least 20%, in specifically preferred cases at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90% and at least 100% or even higher, such as at least 150% or at least 200% as compared to the native pGAP promoter, e.g. determined in the eukaryotic cell selected as host cell for producing the POI.

The native pGAP promoter typically initiates expression of the *gap* gene encoding glyceraldehyde-3-phosphate dehydrogenase (GAPDH), which is a constitutive promoter present in most living organisms. GAPDH (EC 1\2\1\12), a key enzyme of glycolysis and gluconeogenesis, plays a crucial role in catabolic and anabolic carbohydrate metabolism.

The native pGAP promoter specifically is active in a recombinant eukaryotic cell in a similar way as in a native eukaryotic cell of the same species or strain, including the unmodified (non-recombinant) or recombinant eukaryotic cell. Such native pGAP promoter is commonly understood to be an endogenous promoter, thus, homologous to the eukaryotic cell, and serves as a standard or reference promoter for comparison purposes.

For example, a native pGAP promoter of *P. pastoris* is the unmodified, endogenous promoter sequence in *P. pastoris*, as used to control the expression of GAPDH in *P. pastoris*, e.g. having the sequence shown in Figure 13: native pGAP promoter sequence of *P. pastoris* (GS115) (SEQ ID 260). If *P. pastoris* is used as a host for producing a POI according to the invention, the transcription strength or rate of the promoter according to the invention is compared to such native pGAP promoter of *P. pastoris.*

As another example, a native pGAP promoter of *S. cerevisiae* is the unmodified, endogenous promoter sequence in *S. cerevisiae*, as used to control the expression of GAPDH in *S. cerevisiae*. If *S. cerevisiae* is used as a host for producing a POI

according to the invention, the transcription strength or rate of the promoter according to the invention is compared to such native pGAP promoter of *S. cerevisiae.*

Therefore, the relative expression or transcription strength of a promoter according to the invention is usually compared to the native pGAP promoter of a cell of the same species or strain that is used as a host for producing a POI.

The term "regulatable" with respect to a pG1-x promoter or pG1 promoter as used herein shall refer to a promoter that is repressed in a eukaryotic cell in the presence of an excess amount of a carbon source (nutrient or basal substrate) in the growth phase of a batch culture, and de-repressed to exert strong promoter activity in the production phase of a production cell line, e.g. upon reduction of the amount of carbon, such as upon feeding of a growth limiting carbon source (nutrient or supplemental substrate) to a culture according to the fed-batch strategy. In this regard, the term "regulatable" is understood as "carbon source-limit regulatable" or "glucose-limit regulatable", referring to the de-repression of a promoter by carbon consumption, reduction, shortcoming or depletion, or by limited addition of the carbon source so that it is readily consumed by the cells.

The functionally active pG1-x promoter as described herein is a relatively strong regulatable promoter that is silenced or repressed under cell growth conditions (growth phase), and activated or de-repressed under production condition (production phase), and therefore suitable for inducing POI production in a production cell line by limiting the carbon source.

Specifically, the promoter as described herein is carbon source regulatable with a differential promoter strength as determined in a test comparing its strength in the presence of glucose and glucose limitation, showing that it is still repressed at relatively high glucose concentrations, preferably at concentrations of at least 10 g/L, preferably at least 20 g/L. Specifically the promoter according to the invention is fully induced at limited glucose concentrations and glucose threshold concentrations fully inducing the promoter, which threshold is less than 20 g/L, preferably less than 10 g/L, less than 1 g/L, even less than 0.1 g/L or less than 50 mg/L, preferably with a full transcription strength of e.g. at least 50% of the native, homologous pGAP promoter, at glucose concentrations of less than 40 mg/L.

Preferably the induction ratio is understood as a differential promoter strength which is determined by the initiation of POI production upon switching to inducing conditions below a predetermined carbon source threshold, and compared to the

strength in the repressed state. The transcription strength commonly is understood as the strength in the fully induced state, i.e. showing about maximum activities under de-repressing conditions. The differential promoter strength is, e.g. determined according to the efficiency or yield of POI production in a recombinant host cell line under de-repressing conditions as compared to repressing conditions, or else by the amount of a transcript. The regulatable promoter according to the invention has a preferred differential promoter strength, which is at least 2 fold, more preferably at least 5 fold, even more preferred at least 10 fold, more preferred at least 20 fold, more preferably at least 30, 40, 50, or 100 fold in the de-repressed state compared to the repressed state, also understood as fold induction.

The term “sequence identity” of a variant as compared to a parent sequence indicates the degree of identity (or homology) in that two or more nucleotide sequences have the same or conserved base pairs at a corresponding position, to a certain degree, up to a degree close to 100%. A homologous sequence typically has at least about 50% nucleotide sequence identity, preferably at least about 60% identity, more preferably at least about 70% identity, more preferably at least about 80% identity, more preferably at least about 90% identity, more preferably at least about 95% identity.

"Percent (%) identity" with respect to the nucleotide sequence e.g., of a promoter or a gene, is defined as the percentage of nucleotides in a candidate DNA sequence that is identical with the nucleotides in the DNA sequence, after aligning the sequence and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent nucleotide sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared. For purposes of the present invention, the sequence identity between two nucleotide sequences is determined using the NCBI BLAST program version 2.2.29 (Jan-06-2014) with blastn set at the following exemplary parameters: Word Size: 11; Expect value: 10; Gap costs: Existence = 5 , Extension = 2; Filter = low complexity activated; Match/Mismatch Scores: 2,-3; Filter String: L; m.

The term "mutagenesis" as used in the context of the present invention shall refer to a method of providing mutants of a nucleotide sequence, e.g. through insertion, deletion and/or substitution of one or more nucleotides, so to obtain variants thereof with at least one change in the non-coding or coding region. Mutagenesis may be through random, semi-random or site directed mutation. Specific pG1-x promoter variants are derived from the pG1 promoter sequence by a mutagenesis method using the pG1 nucleotide sequence as a parent sequence. Such mutagenesis method encompass those methods of engineering the nucleic acid or *de novo* synthesizing a nucleotide sequence using the pG1 promoter sequence information as a template. Specific mutagenesis methods apply rational promoter engineering.

The pG1-x promoter may be produced by mutagenesis of the pG1 promoter, and variants of the pG1-x promoter as described herein may further be produced, including functionally active variants, employing standard techniques. The promoter may e.g. be modified to generate promoter variants with altered expression levels and regulatory properties. For instance, a promoter library may be prepared by mutagenesis of selected promoter sequences, which may be used as parent molecules, e.g. to fine-tune the gene expression in eukaryotic cells by analyzing variants for their expression under different fermentation strategies and selecting suitable variants. A synthetic library of variants may be used, e.g. to select a promoter matching the requirements for producing a selected POI. Such variants may have increased expression efficiency in eukaryotic host cells and differential expression under carbon source rich and limiting conditions. Typically large randomized gene libraries are produced with a high gene diversity, which may be selected according to a specifically desired genotype or phenotype.

Some of the preferred pG1-x promoter as described herein are size variants of the pG1 promoter and comprise more than one copy of certain elements or regions of the promoter, or comprise one or more (the same or different) fragments of the pG1 promoter.

Specific mutagenesis methods provide for point mutations of one or more nucleotides in a sequence, in particular tandem point mutations, such as to change at least 2, 3, 4, 5, 6, 7, 8, 9, 10, or even more continuous nucleotides within the nucleotide sequence of the promoter. Such mutation is typically at least one of a deletion, insertion, and/or substitution of one or more nucleotides. The promoter sequence may be mutated at the distal ends, in particular within the 5'-region which

amounts to up to 50% of the nucleotide sequence, which may be highly variable without substantially losing the promoter activity. The promoter sequence may specifically be mutated within the main regulatory region, yet, it is preferred that the sequence identity to the pG1 parent main regulatory region and in particular to the parent core regulatory region is high, such as e.g. at least 80%. Within the main regulatory region, but outside the core regulatory region the variability of the sequence may be higher so to obtain a sequence identity of less than 80%.

The core regulatory region specifically incorporates the SEQ ID 2 and SEQ ID 3, which represent transcription factor binding sites (TFBS) and an interstitional region between SEQ ID 2 and SEQ ID 3.

The nucleotide sequence identified as SEQ ID 2 comprises at least part of the TFBS recognized by Rgt1, Cat8-1 and Cat8-2.

The nucleotide sequence identified as SEQ ID 3 comprises at least part of the TFBS recognized by Rgt1, Cat8-1 and Cat8-2.

Specifically, the nucleotide sequence between SEQ ID 2 and SEQ ID 3 (the interstitional sequence) may be mutated to a non-homologous sequence (e.g., with a sequence identity of less than 50%) or even be deleted.

Any mutations within the SEQ ID 2 and SEQ ID 3 are specifically conservative, i.e. such as to maintain (or improve) the recognition by the respective transcription factor. Upon engineering such conservative mutants, the sequence identity within the SEQ ID 2 and/or SEQ ID 3 nucleotide sequence is at least 90%, preferably at least 95%.

The main regulatory region comprises or consists of the nucleotide sequence identified by SEQ ID 5. Such region comprises the core regulatory region and further non-core regulatory region, which comprises essential elements of the pG1 promoter and which may be mutated to a certain extent to produce the pG1-x promoter as described herein.

Specific regions of site directed mutagenesis are e.g., the non-core regulatory region of the pG1 or the pG1-x promoter (inside or outside the main regulatory region). However, specific mutants may as well be prepared by mutagenesis methods directed to the core regulatory region of the promoter, keeping a certain degree of sequence identity to maintain the promoter function. Further specific regions are outside or within the main regulatory region. Specifically, the promoter may comprise a hybrid nucleotide sequence e.g. comprising the core regulatory region of the pG1 promoter

and one or more regions or alternative (native or artificial) promoter, such as the translation initiation site at the 3'-region (specifically the 3'-end which comprises at least 10 terminal nucleotides, or at least 15 terminal nucleotides) of a promoter which is any other than the pG1 promoter may be used to substitute the translation initiation site of the pG1 promoter.

Specific mutations refer to the duplication of selected regions (or motifs) of the pG1 promoter e.g., the T motif or the extended T motif. Such selected motifs may be elongated by additional nucleotides or shortened at one or both distal ends of the motif, or within the motif. The native pG1 sequence comprises a TAT motif consisting of the nucleotides "T" followed by "A" followed by T15 (SEQ ID 14). Such TAT motif 5´-TATTTTTTTTTTTTTTT-3´ (SEQ ID 22) has turned out to have a positive effect on the promoter strength, which may even be increased by duplicating the TAT motif, or inserting at least 2, or 3, or 4 copies of the TAT motif, either the same TAT motif or using an alternative T motif, extended T motif (e.g. a TAT motif), which comprises at least the T13 motif (SEQ ID 12).

The invention further encompasses a nucleotide sequence which hybridizes under stringent conditions to the pG1-x promoter.

As used in the present invention, the term "hybridization" or "hybridizing" is intended to mean the process during which two nucleic acid sequences anneal to one another with stable and specific hydrogen bonds so as to form a double strand under appropriate conditions. The hybridization between two complementary sequences or sufficiently complementary sequences depends on the operating conditions that are used, and in particular the stringency. The stringency may be understood to denote the degree of homology; the higher the stringency, the higher percent homology between the sequences. The stringency may be defined in particular by the base composition of the two nucleic sequences, and/or by the degree of mismatching between these two nucleic sequences. By varying the conditions, e.g. salt concentration and temperature, a given nucleic acid sequence may be allowed to hybridize only with its exact complement (high stringency) or with any somewhat related sequences (low stringency). Increasing the temperature or decreasing the salt concentration may tend to increase the selectivity of a hybridization reaction.

As used herein, the phrase "hybridizing under stringent hybridizing conditions" is preferably understood to refer to hybridizing under conditions of certain stringency. In a preferred embodiment the "stringent hybridizing conditions" are conditions where

homology of the two nucleic acid sequences is at least 70%, preferably at least 80%, preferably at least 90%, i.e. under conditions where hybridization is only possible if the double strand obtained during this hybridization comprises preferably at least 70%, preferably at least 80%, preferably at least 90% of A-T bonds and C-G bonds.

The stringency may depend on the reaction parameters, such as the concentration and the type of ionic species present in the hybridization solution, the nature and the concentration of denaturing agents and/or the hybridization temperature. The appropriate conditions can be determined by those skilled in the art, e.g. as described in Sambrook et al. (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, 1989).

The term “isolated” or “isolation” as used herein with respect to a nucleic acid, a POI or other compound shall refer to such compound that has been sufficiently separated from the environment with which it would naturally be associated, so as to exist in "substantially pure" form. "Isolated" does not necessarily mean the exclusion of artificial or synthetic mixtures with other compounds or materials, or the presence of impurities that do not interfere with the fundamental activity, and that may be present, for example, due to incomplete purification. In particular, isolated nucleic acid molecules of the present invention are also meant to include those chemically synthesized,”, and in particular those not naturally-occurring in *P. pastoris* or any other organism, herein referred to as “artificial”. With reference to nucleic acids of the invention, the term "isolated nucleic acid" or “isolated nucleic acid sequence” is sometimes used. This term, when applied to DNA, refers to a DNA molecule that is separated from sequences with which it is immediately contiguous in the naturally occurring genome of the organism in which it originated. For example, an "isolated nucleic acid" may comprise a DNA molecule inserted into a vector, such as a plasmid or virus vector, or integrated into the genomic DNA of a prokaryotic or eukaryotic cell or host organism. An "isolated nucleic acid" (either DNA or RNA) may further represent a molecule produced directly by biological or synthetic means and separated from other components present during its production.

The term "operably linked" as used herein refers to the association of nucleotide sequences on a single nucleic acid molecule, e.g. a vector, in a way such that the function of one or more nucleotide sequences is affected by at least one other nucleotide sequence present on said nucleic acid molecule. For example, a promoter is operably linked with a coding sequence of a recombinant gene, when it is capable of

effecting the expression of that coding sequence. As a further example, a nucleic acid encoding a signal peptide is operably linked to a nucleic acid sequence encoding a POI, when it is capable of expressing a protein in the secreted form, such as a preform of a mature protein or the mature protein. Specifically, such nucleic acids operably linked to each other may be immediately linked, i.e. without further elements or nucleic acid sequences in between the nucleic acid encoding the signal peptide and the nucleic acid sequence encoding a POI.

A promoter sequence is typically understood to be operably linked to a coding sequence, if the promoter controls the transcription of the coding sequence. If a promoter sequence is not natively associated with the coding sequence, its transcription is either not controlled by the promoter in native (wild-type) cells or the sequences are recombined with different contiguous sequences.

The term "protein of interest (POI)" as used herein refers to a polypeptide or a protein that is produced by means of recombinant technology in a host cell. More specifically, the protein may either be a polypeptide not naturally occurring in the host cell, i.e. a heterologous protein, or else may be native to the host cell, i.e. a homologous protein to the host cell, but is produced, for example, by transformation with a self-replicating vector containing the nucleic acid sequence encoding the POI, or upon integration by recombinant techniques of one or more copies of the nucleic acid sequence encoding the POI into the genome of the host cell, or by recombinant modification of one or more regulatory sequences controlling the expression of the gene encoding the POI, e.g. of the promoter sequence. In some cases the term POI as used herein also refers to any metabolite product by the host cell as mediated by the recombinantly expressed protein.

The POI may specifically be recovered from the cell culture in the purified form, e.g. substantially pure.

The term "substantially pure" or “purified” as used herein shall refer to a preparation comprising at least 50% (w/w), preferably at least 60%, 70%, 80%, 90% or 95% of a compound, such as a nucleic acid molecule or a POI. Purity is measured by methods appropriate for the compound (e.g. chromatographic methods, polyacrylamide gel electrophoresis, HPLC analysis, and the like).

The term “recombinant” as used herein shall mean “being prepared by or the result of genetic engineering”. Thus, a “recombinant microorganism” comprises at least one “recombinant nucleic acid”. A recombinant microorganism specifically comprises

an expression vector or cloning vector, or it has been genetically engineered to contain a recombinant nucleic acid sequence. A “recombinant protein” is produced by expressing a respective recombinant nucleic acid in a host. A “recombinant promoter” is a genetically engineered non-coding nucleotide sequence suitable for its use as a functionally active promoter as described herein.

In general, the recombinant nucleic acids or organisms as referred to herein may be produced by recombination techniques well known to a person skilled in the art. In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Maniatis, Fritsch & Sambrook, "Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, (1982).

According to a preferred embodiment of the present invention, a recombinant construct is obtained by ligating the promoter and relevant genes into a vector or expression construct. These genes can be stably integrated into the host cell genome by transforming the host cell using such vectors or expression constructs.

Expression vectors may include but are not limited to cloning vectors, modified cloning vectors and specifically designed plasmids. The preferred expression vector as used in the invention may be any expression vector suitable for expression of a recombinant gene in a host cell and is selected depending on the host organism. The recombinant expression vector may be any vector which is capable of replicating in or integrating into the genome of the host organisms, also called host vector.

Appropriate expression vectors typically comprise further regulatory sequences suitable for expressing DNA encoding a POI in a eukaryotic host cell. Examples of regulatory sequences include operators, enhancers, ribosomal binding sites, and sequences that control transcription and translation initiation and termination. The regulatory sequences may be operably linked to the DNA sequence to be expressed.

To allow expression of a recombinant nucleotide sequence in a host cell, the expression vector may provide the promoter according to the invention adjacent to the 5’ end of the coding sequence, e.g. upstream from the gene of interest (GOI) or a signal peptide gene enabling secretion of the POI. The transcription is thereby regulated and initiated by this promoter sequence.

The term “signal peptide” as used herein shall specifically refer to a native signal peptide, a heterologous signal peptide or a hybrid of a native and a heterologous signal peptide, and may specifically be heterologous or homologous to the host

organism producing a POI. The function of the signal peptide is to allow the POI to be secreted to enter the endoplasmic reticulum. It is usually a short (3-60 amino acids long) peptide chain that directs the transport of a protein outside the plasma membrane, thereby making it easy to separate and purify a heterologous protein. Some signal peptides are cleaved from the protein by signal peptidase after the proteins are transported.

Exemplary signal peptides are signal sequences from *S. cerevisiae* alpha-mating factor prepro peptide and the signal peptides from the *P. pastoris* acid phosphatase gene (PHO1) and the extracellular protein X (EPX1) (Heiss et al., 2015; WO2014067926A1).

Expression vectors comprising one or more of the regulatory elements (such as the pG1-x promoter and optionally a signal sequence) may be constructed to drive expression of a POI, and the expressed yield is compared to constructs with conventional regulatory elements, such as to prove the function of the relevant sequences. The identified nucleotide sequences may be amplified by PCR using specific nucleotide primers, cloned into an expression vector and transformed into a eukaryotic cell line, e.g. using a yeast vector and a strain of *P. pastoris,* for high level production of various different POI. To estimate the effect of the pG1-x promoter as described herein on the amount of recombinant POI so produced, the eukaryotic cell line may be cultured in shake flask experiments and fedbatch or chemostat fermentations in comparison with strains comprising a conventional pG1 promoter or the pGAP promoter, in the respective cell. In particular, the choice of the promoter has a great impact on the recombinant protein production.

The POI can be produced using the recombinant host cell line by culturing a transformant, thus obtained in an appropriate medium, isolating the expressed product or metabolite from the culture, and optionally purifying it by a suitable method.

Transformants according to the present invention can be obtained by introducing such a vector DNA, e.g. plasmid DNA, into a host and selecting transformants which express the POI or the host cell metabolite with high yields. Host cells are treated to enable them to incorporate foreign DNA by methods conventionally used for transformation of eukaryotic cells, such as the electric pulse method, the protoplast method, the lithium acetate method, and modified methods thereof. *P. pastoris* is preferably transformed by electroporation. Preferred methods of transformation for the uptake of the recombinant DNA fragment by the microorganism include chemical

transformation, electroporation or transformation by protoplastation. Transformants according to the present invention can be obtained by introducing such a vector DNA, e.g. plasmid DNA, into a host and selecting transformants which express the relevant protein or host cell metabolite with high yields.

Several different approaches for the production of the POI according to the method of the invention are preferred. Substances may be expressed, processed and optionally secreted by transforming a eukaryotic host cell with an expression vector harboring recombinant DNA encoding a relevant protein and at least one of the regulatory elements as described above, preparing a culture of the transformed cell, growing the culture, inducing transcription and POI production, and recovering the product of the fermentation process.

The host cell according to the invention is preferably tested for its expression capacity or yield by the following test: ELISA, activity assay, HPLC, or other suitable tests.

The invention specifically allows for the fermentation process on a pilot or industrial scale. The industrial process scale would preferably employ volumina of at least 10 L, specifically at least 50 L, preferably at least 1 m^3, preferably at least 10 m^3, most preferably at least 100 m^3.

Production conditions in industrial scale are preferred, which refer to e.g. fed batch cultivation in reactor volumes of 100 L to 10 m^3 or larger, employing typical process times of several days, or continuous processes in fermenter volumes of approximately 50 – 1000 L or larger, with dilution rates of approximately 0.02 – 0.15 h^{-1}.

The suitable cultivation techniques may encompass cultivation in a bioreactor starting with a batch phase, followed by a short exponential fed batch phase at high specific growth rate, further followed by a fed batch phase at a low specific growth rate. Another suitable cultivation technique may encompass a batch phase followed by a continuous cultivation phase at a low dilution rate.

A preferred embodiment includes a batch culture to provide biomass followed by a fed-batch culture for high yields POI production.

It is preferred to cultivate the host cell line as described herein in a bioreactor under growth conditions to obtain a cell density of at least 1 g/L cell dry weight, more preferably at least 10 g/L cell dry weight, preferably at least 20 g/L cell dry weight. It is

advantageous to provide for such yields of biomass production on a pilot or industrial scale.

A growth medium allowing the accumulation of biomass, specifically a basal growth medium, typically comprises a carbon source, a nitrogen source, a source for sulphur and a source for phosphate. Typically, such a medium comprises furthermore trace elements and vitamins, and may further comprise amino acids, peptone or yeast extract.

Preferred nitrogen sources include $NH_4H_2PO_4$, or NH_3 or $(NH_4)_2SO_4$;

Preferred sulphur sources include $MgSO_4$, or $(NH_4)_2SO_4$ or K_2SO_4;

Preferred phosphate sources include $NH_4H_2PO_4$, or H_3PO_4 or NaH_2PO_4, KH_2PO_4, Na_2HPO_4 or K_2HPO_4;

Further typical medium components include KCl, $CaCl_2$, and Trace elements such as: Fe, Co, Cu, Ni, Zn, Mo, Mn, I, B;

Preferably the medium is supplemented with vitamin B_7;

A typical growth medium for *P. pastoris* comprises glycerol, sorbitol or glucose, $NH_4H_2PO_4$, $MgSO_4$, KCl, $CaCl_2$, biotin, and trace elements.

In the production phase a production medium is specifically used with only a limited amount of a supplemental carbon source.

Preferably the host cell line is cultivated in a mineral medium with a suitable carbon source, thereby further simplifying the isolation process significantly. An example of a preferred mineral medium is one containing an utilizable carbon source (e.g. glucose, glycerol, sorbitol or methanol), salts containing the macro elements (potassium, magnesium, calcium, ammonium, chloride, sulphate, phosphate) and trace elements (copper, iodide, manganese, molybdate, cobalt, zinc, and iron salts, and boric acid), and optionally vitamins or amino acids, e.g. to complement auxotrophies.

Specifically, the cells are cultivated under conditions suitable to effect expression of the desired POI, which can be purified from the cells or culture medium, depending on the nature of the expression system and the expressed protein, e.g. whether the protein is fused to a signal peptide and whether the protein is soluble or membrane-bound. As will be understood by the skilled artisan, cultivation conditions will vary according to factors that include the type of host cell and particular expression vector employed.

A typical production medium comprises a supplemental carbon source, and further $NH_4H_2PO_4$, $MgSO_4$, KCl, $CaCl_2$, biotin, and trace elements.

For example the feed of the supplemental carbon source added to the fermentation may comprise a carbon source with up to 50 wt % utilizable sugars. The low feed rate of the supplemental medium will limit the effects of product or byproduct inhibition on the cell growth, thus a high product yield based on substrate provision will be possible.

The fermentation preferably is carried out at a pH ranging from 3 to 7.5.

Typical fermentation times are about 24 to 120 hours with temperatures in the range of 20 °C to 35°C, preferably 22-30°C.

The POI is preferably expressed employing conditions to produce yields of at least 1 mg/L, preferably at least 10 mg/L, preferably at least 100 mg/L, most preferred at least 1 g/L.

It is understood that the methods disclosed herein may further include cultivating said recombinant host cells under conditions permitting the expression of the POI, preferably in the secreted form or else as intracellular product. A recombinantly produced POI or a host cell metabolite can then be isolated from the cell culture medium and further purified by techniques well known to a person skilled in the art.

The POI produced according to the invention typically can be isolated and purified using state of the art techniques, including the increase of the concentration of the desired POI and/or the decrease of the concentration of at least one impurity.

If the POI is secreted from the cells, it can be isolated and purified from the culture medium using state of the art techniques. Secretion of the recombinant expression products from the host cells is generally advantageous for reasons that include facilitating the purification process, since the products are recovered from the culture supernatant rather than from the complex mixture of proteins that results when yeast cells are disrupted to release intracellular proteins.

The cultured transformant cells may also be ruptured sonically or mechanically, enzymatically or chemically to obtain a cell extract containing the desired POI, from which the POI is isolated and purified.

As isolation and purification methods for obtaining a recombinant polypeptide or protein product, methods, such as methods utilizing difference in solubility, such as salting out and solvent precipitation, methods utilizing difference in molecular weight, such as ultrafiltration and gel electrophoresis, methods utilizing difference in electric charge, such as ion-exchange chromatography, methods utilizing specific affinity, such as affinity chromatography, methods utilizing difference in hydrophobicity, such as

reverse phase high performance liquid chromatography, and methods utilizing difference in isoelectric point, such as isoelectric focusing may be used.

The highly purified product is essentially free from contaminating proteins, and preferably has a purity of at least 90%, more preferred at least 95%, or even at least 98%, up to 100%. The purified products may be obtained by purification of the cell culture supernatant or else from cellular debris.

As isolation and purification methods the following standard methods are preferred: Cell disruption (if the POI is obtained intracellularly), cell (debris) separation and wash by Microfiltration or Tangential Flow Filter (TFF) or centrifugation, POI purification by precipitation or heat treatment, POI activation by enzymatic digest, POI purification by chromatography, such as ion exchange (IEX), hydrophobic interaction chromatography (HIC), Affinity chromatography, size exclusion (SEC) or HPLC Chromatography, POI precipitation of concentration and washing by ultrafiltration steps.

The isolated and purified POI can be identified by conventional methods such as Western blot, HPLC, activity assay, or ELISA.

The POI can be any eukaryotic, prokaryotic or synthetic polypeptide. It can be a secreted protein or an intracellular protein. The present invention also provides for the recombinant production of functional homologs, functional equivalent variants, derivatives and biologically active fragments of naturally occurring proteins. Functional homologs are preferably identical with or correspond to and have the functional characteristics of a sequence.

A POI referred to herein may be a product homologous to the eukaryotic host cell or heterologous, preferably for therapeutic, prophylactic, diagnostic, analytic or industrial use.

The POI is preferably a heterologous recombinant polypeptide or protein, produced in a eukaryotic cell, preferably a yeast cell, preferably as secreted proteins. Examples of preferably produced proteins are immunoglobulins, immunoglobulin fragments, aprotinin, tissue factor pathway inhibitor or other protease inhibitors, and insulin or insulin precursors, insulin analogues, growth hormones, interleukins, tissue plasminogen activator, transforming growth factor a or b, glucagon, glucagon-like peptide 1 (GLP-1), glucagon-like peptide 2 (GLP-2), GRPP, Factor VII, Factor VIII, Factor XIII, platelet-derived growth factor1, serum albumin, enzymes, such as lipases or proteases, or a functional homolog, functional equivalent variant, derivative and

biologically active fragment with a similar function as the native protein. The POI may be structurally similar to the native protein and may be derived from the native protein by addition of one or more amino acids to either or both the C- and N-terminal end or the side-chain of the native protein, substitution of one or more amino acids at one or a number of different sites in the native amino acid sequence, deletion of one or more amino acids at either or both ends of the native protein or at one or several sites in the amino acid sequence, or insertion of one or more amino acids at one or more sites in the native amino acid sequence. Such modifications are well known for several of the proteins mentioned above.

A POI can also be selected from substrates, enzymes, inhibitors or cofactors that provide for biochemical reactions in the host cell, with the aim to obtain the product of said biochemical reaction or a cascade of several reactions, e.g. to obtain a metabolite of the host cell. Exemplary products can be vitamins, such as riboflavin, organic acids, and alcohols, which can be obtained with increased yields following the expression of a recombinant protein or a POI according to the invention.

In general, the host cell, which expresses a recombinant product, can be any eukaryotic cell suitable for recombinant expression of a POI.

Examples of preferred mammalian cells are BHK, CHO (CHO-DG44, CHO-DUXB11, CHO-DUKX, CHO-K1, CHOK1SV, CHO-S), HeLa, HEK293, MDCK, NIH3T3, NS0, PER.C6, SP2/0 and VERO cells.

Examples of preferred yeast cells used as host cells according to the invention include but are not limited to the *Saccharomyces* genus (e.g. *Saccharomyces cerevisiae)*, the *Pichia* genus (e.g. *P. pastoris, or P. methanolica*), the *Komagataella* genus (*K. pastoris*, *K. pseudopastoris* or *K. phaffii*), *Hansenula polymorpha, Yarrowia lipolytica, Schefferomyces stipitis* or *Kluyveromyces lactis*.

Newer literature divides and renames *Pichia pastoris* into *Komagataella pastoris*, *Komagataella phaffii* and *Komagataella pseudopastoris*. Herein *Pichia pastoris* is used synonymously for all, *Komagataella pastoris*, *Komagataella phaffii* and *Komagataella pseudopastoris*.

The preferred yeast host cells are derived from methylotrophic yeast, such as from *Pichia* or *Komagataella*, e.g. *Pichia pastoris*, or *Komagataella pastoris*, or *K. phaffii*, or *K. pseudopastoris*. Examples of the host include yeasts such as *P. pastoris*. Examples of *P. pastoris* strains include CBS 704 (=NRRL Y-1603 = DSMZ 70382), CBS 2612 (=NRRL Y-7556), CBS 7435 (=NRRL Y-11430), CBS 9173-9189 (CBS

strains: CBS-KNAW Fungal Biodiversity Centre, Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands), and DSMZ 70877 (German Collection of Microorganisms and Cell Cultures), but also strains from Invitrogen, such as X-33, GS115, KM71 and SMD1168. Examples of *S. cerevisiae* strains include W303, CEN.PK and the BY-series (EUROSCARF collection). All of the strains described above have been successfully used to produce transformants and express heterologous genes.

A preferred yeast host cell according to the invention, such as a *P. pastoris* or *S. cerevisiae* host cell, contains a heterologous or recombinant promoter sequences, which may be derived from a *P. pastoris* or *S. cerevisiae* strain, different from the production host. In another specific embodiment the host cell according to the invention comprises a recombinant expression construct according to the invention comprising the promoter originating from the same genus, species or strain as the host cell.

According to the invention it is preferred to provide a *P. pastoris* host cell line comprising a pG1-x promoter sequence as described herein operably linked to the nucleotide sequence coding for the POI.

If the POI is a protein homologous to the host cell, i.e. a protein which is naturally occurring in the host cell, the expression of the POI in the host cell may be modulated by the exchange of its native promoter sequence with a promoter sequence according to the invention.

This purpose may be achieved e.g. by transformation of a host cell with a recombinant DNA molecule comprising homologous sequences of the target gene to allow site specific recombination, the promoter sequence and a selective marker suitable for the host cell. The site specific recombination shall take place in order to operably link the promoter sequence with the nucleotide sequence encoding the POI. This results in the expression of the POI from the promoter sequence according to the invention instead of from the native promoter sequence.

It is specifically preferred that the pG1-x promoter has an increased promoter activity relative to the native promoter sequence of the POI.

According to a specific embodiment, the POI production method employs a recombinant nucleotide sequence encoding the POI, which is provided on a plasmid suitable for integration into the genome of the host cell, in a single copy or in multiple copies per cell. The recombinant nucleotide sequence encoding the POI may also be provided on an autonomously replicating plasmid in a single copy or in multiple copies per cell.

The preferred method as described herein employs a plasmid, which is a eukaryotic expression vector, preferably a yeast expression vector. Expression vectors may include but are not limited to cloning vectors, modified cloning vectors and specifically designed plasmids. The preferred expression vector as used in the invention may be any expression vector suitable for expression of a recombinant gene in a host cell and is selected depending on the host organism. The recombinant expression vector may be any vector which is capable of replicating in or integrating into the genome of the host organisms, also called host vector, such as a yeast vector, which carries a DNA construct according to the invention. A preferred yeast expression vector is for expression in yeast selected from the group consisting of methylotrophic yeasts represented by the genera Hansenula, Pichia, Candida and Torulopsis.

In the present invention, it is preferred to use plasmids derived from pPICZ, pGAPZ, pPIC9, pPICZalfa, pGAPZalfa, pPIC9K, pGAPHis or pPUZZLE as the vector.

According to a preferred embodiment of the present invention, a recombinant construct is obtained by ligating the relevant genes into a vector. These genes can be stably integrated into the host cell genome by transforming the host cell using such vectors. The polypeptides encoded by the genes can be produced using the recombinant host cell line by culturing a transformant, thus obtained in an appropriate medium, isolating the expressed POI from the culture, and purifying it by a method appropriate for the expressed product, in particular to separate the POI from contaminating proteins.

Expression vectors may comprise one or more phenotypic selectable markers, e.g. a gene encoding a protein that confers antibiotic resistance or that supplies an autotrophic requirement. Yeast vectors commonly contain an origin of replication from a yeast plasmid, an autonomously replicating sequence (ARS), or alternatively, a sequence used for integration into the host genome, a promoter region, sequences for polyadenylation, sequences for transcription termination, and a selectable marker.

The procedures used to ligate the DNA sequences and regulatory elements, e.g. the pG1-x promoter and the gene(s) coding for the POI, the promoter and the terminator, respectively, and to insert them into suitable vectors containing the information necessary for integration or host replication, are well-known to persons skilled in the art, e.g. described by J. Sambrook et al., (A Laboratory Manual, Cold Spring Harbor, 1989).

It will be understood that the vector, which uses the regulatory elements according to the invention and/or the POI as an integration target, may be constructed either by first preparing a DNA construct containing the entire DNA sequence coding for the regulatory elements and/or the POI and subsequently inserting this fragment into a suitable expression vector, or by sequentially inserting DNA fragments containing genetic information for the individual elements, followed by ligation.

Also multicloning vectors, which are vectors having a multicloning site, can be used according to the invention, wherein a desired heterologous gene can be incorporated at a multicloning site to provide an expression vector. In expression vectors, the promoter is placed upstream of the gene of the POI and regulates the expression of the gene. In the case of multicloning vectors, because the gene of the POI is introduced at the multicloning site, the promoter is placed upstream of the multicloning site.

The DNA construct as provided to obtain a recombinant host cell according to the invention may be prepared synthetically by established standard methods, e.g. the phosphoramidite method. The DNA construct may also be of genomic or cDNA origin, for instance obtained by preparing a genomic or cDNA library and screening for DNA sequences coding for all or part of the polypeptide of the invention by hybridization using synthetic oligonucleotide probes in accordance with standard techniques (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, 1989). Finally, the DNA construct may be of mixed synthetic and genomic, mixed synthetic and cDNA or mixed genomic and cDNA origin prepared by annealing fragments of synthetic, genomic or cDNA origin, as appropriate, the fragments corresponding to various parts of the entire DNA construct, in accordance with standard techniques.

In another preferred embodiment, the yeast expression vector is able to stably integrate in the yeast genome, e. g. by homologous recombination.

A transformant host cell according to the invention obtained by transforming the cell with the regulatory elements according to the invention and/or the POI genes may preferably first be cultivated at conditions to grow efficiently to a large cell number. When the cell line is prepared for the POI expression, cultivation techniques are chosen to produce the expression product.

The foregoing description will be more fully understood with reference to the following examples. Such examples are, however, merely representative of methods of

practicing one or more embodiments of the present invention and should not be read as limiting the scope of invention.

EXAMPLES

Example 1: 5'-shortening of pG1 reveals the main regulatory region of pG1

The native (wild-type) pG1 promoter has been isolated from *P. pastoris* (*Komagatella phaffii*) strain CBS2612 (CBS strains: CBS-KNAW Fungal Biodiversity Centre, Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands). As determined by Sanger sequencing and subsequent BLAST analysis, the pG1 promoter sequence of CBS2612 had more than 95% sequence identity to the respective regions in the genomic sequences of the strains GS115 (Invitrogen) (upstream of PAS_chr1-3_0011) and CBS7435 (upstream of P7435_Chr1-0007) or *K. pastoris* DSMZ 70382 (DSMZ strains: German Collection of Microorganisms and Cell Cultures) (upstream of PIPA00372). During the analysis of the genomic region of pG1, it was realized that its gene *GTH1* has a different start annotation in the strains CBS7435 (P7435_Chr1-0007) and DSMZ 70382 (PIPA00372) than in GS115 (PAS_chr1-3_0011). In contrast to GS115 and CBS2612, the coding sequence is annotated to start 36 bp further downstream in the genomic sequences of the other two strains.

In order to identify the relevant regulatory region of pG1 8 shortened pG1 variants were cloned from CBS2612 starting from the alternative 5´ positions -858, -663, -492, -371, -328, -283, -211 and -66 to position -1 (see **Figure 1**, numbering based on the start of the GTH1 gene locus PAS_chr1-3_0011). These shortened promoter variants were screened for eGFP expression in deep well plates as described in Example 8 to test for the repression- (glycerol) and induction properties (glucose feed beads) in comparison to the original 965 bp version of pG1 (**Figure 2**). No difference in eGFP signal was found for all length variants in the repressing condition, showing that promoter repression was not restricted in any of the shortened variants. After 48 hours of induction, the expression capacity remained fully functional for the promoter variants down to a length of 328 bp. The 283 bp- variant was only about two thirds strong compared to the original pG1 promoter. The two shortest length variants (211 and 66 bp) appeared to be almost nonfunctional. These results that the region between position -400 and -200 contains important regulatory features.

Example 2: A high density of predicted carbon source related TFBS marks the main regulatory region of the pG1 promoter

The pG1 promoter sequence (1000 bp upstream of the gene PAS_chr1-3_0011) was searched for matrix families belonging to the matrix groups 'fungi' and 'general core promoter elements' using the MatInspector from Genomatix. 111 putative TFBS belonging to 46 different matrix families were found (**Table 1**). The most common matrix families in the analyzed sequence were monomeric Gal4-class motifs (F$MGCM, 12 binding sites), homeodomain-containing transcriptional regulators (F$HOMD, 6 binding sites), fungal basic leucine zipper family (F$BZIP, 5 binding sites) and yeast GC-Box Proteins (F$YMIG, 5 binding sites). A very high TFBS binding site density was noticed between position -400 to -200 with about two thirds of the mentioned TFBS (most common matrix families) occurring there (18 out of 28). Regarding general core promoter elements, no yeast- or fungi-related motifs were identified by the MatInspector, but a TATA box can be found starting at position -26.

A prominent motif was identified e.g. at position -390 to-375, which was termed TAT14 due to its sequence 5'-TATTTTTTTTTTTTTT-3' (SEQ ID 21) or TAT15 due to its sequence 5'-TATTTTTTTTTTTTTTT-3' (SEQ ID 22). Such poly(A:T) tracts in promoter regions are known to negatively affect nucleosome binding and to stimulate TF binding at nearby sites in yeast.

Example 3: The carbon source-related transcription factors Mxr1, Rgt1, Cat8-1, Cat8-2 and Mig1 were revealed to be important for the regulatory properties of pG1

Transcription factor binding sites with predicted glucose- or carbon source dependency were selected for further analysis (see **Figure 1** and **Table 2**). pG1 variants with deletions of the respective regions were generated using overlap-extension PCR. **Table 3** lists all selected TFBS and indicates all TFBS which are (partially) affected by the deletion (detailed list in Table 2). For some deletions (e.g. Δ9 and Δ10), some nucleotides of the respective TFBS were left untouched in order to keep close neighboring TFBS functional and to separately examine their effect.

All TFBS deletion and TAT mutation variants were screened for eGFP expression as described in Example 8 in repressing (glycerol) and inducing conditions (glucose feed bead) (**Figure 3**). It is important to consider that individual TF/TFBS are usually not sufficient to fulfill a promoter's regulation. TFBS deletions also imply that

the promoter sequence can be affected by the newly formed adjoined sequence, by altered distances between TFBS or by changes of higher order properties (chromatin organization). The same TFBS at different positions of the promoter can have different functions, also because of other adjacent TFBS. At closely neighbouring TFBS, TFs might either act synergistically or restrict binding of other TFs due to steric hindrance.

Four different carbon source-related TF families were deleted in the pG1 promoter variants (see **Table 2** and **Table 3**): Yeast metabolic regulator (F$ADR; matrixes: F$ADR1.01), Monomeric Gal4-class motifs (F$MGCM; matrixes: F$RGT1.01, F$RGT1.02), Carbon source-responsive elements (F$CSRE, matrixes: F$CSRE.01, F$SIP4.01) and Yeast GC-Box Proteins (F$YMIG; matrixes: F$MIG1.01 and F$MIG1.02). The corresponding transcription factors in *S. cerevisiae* are Adr1, Rgt1, Sip4/Cat8 and Mig1, respectively.

Carbon source dependent promoters are controlled by glucose repression and/or induction by carbohydrates or other non-sugar carbon sources. Glucose repression is mainly conducted by the Snf1 protein kinase complex, the transcriptional repressor Mig1 and protein phosphatase 1. Downstream factors regulate e.g. respiratory genes (Hap4), gluconeogenesis genes (Cat8, Sip4) and glucose transporters (Rgt1) in *S. cerevisiae.*

P. pastoris has two Mig1 homologs, called Mig1-1 and Mig1-2, the second of which possibly acts as carbon catabolite repressor. When glucose is available, Mig1 acts as a repressor, while Rgt1 acts as transcriptional activator. To fulfill repressor function, Mig1 gets dephosphorylated and imported into the nucleus where it recruits the corepressors Ssn6 and Tup1.

In limiting glucose, Rgt1 gets dephosphorylated and acts as transcriptional repressor. Rgt1 function is controlled by its phosphorylation state (Rgt1 has four phosphorylation sites), and induction of regulatedpromoters does not require Rgt1 dissociation in *S. cerevisiae*, as typically seen for transcriptional repressors.

The carbon source-responsive zinc-finger transcription factor Adr1 is required for transcriptional activation of the glucose-repressible alcohol dehydrogenase (*ADH2*) gene in *S. cerevisae*. The Adr1 homolog in *P. pastoris* is Mxr1 (PAS_chr4_0487), the key regulator of methanol metabolism, and it was reported to be a positive acting transcription factor being essential for strong P_{AOX} induction on methanol. The reported TFBS core motif 5′ CYCC 3′ for Mxr1 matches with both F$ADR1.01 sites found in the pG1 promoter sequence.

The carbon source response element (CSRE) is bound by the transcriptional activators Sip4 and Cat8 and functions to induce the expression of gluconeogenesis genes in *S. cerevisiae*. Two *P. pastoris* homologs of *Sc*Cat8 can be found: Cat8-1 (PAS_chr2-1_0757) and Cat8-2 (PAS_chr4_0540), both also being the best blastp hits for *Sc*Sip4. Cat8-2 is weakly similar to *Sc*Cat8, and it potentially plays an important role in derepressing conditions.

Example 4: Deletion variants of the pG1 promoter reveal TFBS responsible for its repression and induction

Out of the 5 deletion variants residing upstream (5′) of the main regulatory region of pG1 identified before (see dashed box in **Figure 1 and Table 2**), the variants pG1-Δ1, -Δ2 and -Δ4 appear to have a beneficial effect on promoter strength while the deletion variants pG1-Δ3 and Δ5 had no effect on GFP expression compared to the original pG1 promoter (SEQ ID 9). This result suggests that 5′ shortening of the promoter might be beneficial for the engineering of pG1. TFBS deletions within the main regulatory region of pG1 (pG1-Δ6 to -Δ12, see **Figure 1 and Table 2**) had different impacts on eGFP expression, but none showed increased induction without losing the repression properties. Therefore, it is assumed that the main regulatory region of pG1 needs to be maintained in engineered pG1 promoter variants in order to retain its tight regulation. Accordingly, without this region, much lower induction in limiting glucose was observed in Example 1 (pG1-328 and pG1-283, **Figure 2**).

Mig1 binding sites were deleted in pG1-Δ3, -Δ4, -Δ10 and -Δ11 (F$MIG1.02 in Δ3, F$MIG1.01 in Δ4, Δ10 and Δ11), whereat pG1-Δ10 and pG1-Δ11 also include F$ADR1.01 and F$RGT1.02 deletions, respectively. Slightly tighter repression was found for Δ3, while Δ4 had unchanged repression but enhanced eGFP levels after induction.

Liberated repression seen for Δ10 and weaker promoter induction of Δ10 and Δ11 could also be connected to F$RGT1 binding sites in this region (F$RGT1.01 and F$RGT1.02 deleted in Δ9 and Δ11). Also, Mig1 could play a bifunctional role in pG1 regulation: two *MIG1* genes are found in *P. pastoris* (*MIG1-1, MIG1-2*) and they were shown to be regulated contrariwise upon glucose availability.

The deletion of F$ADR1.01 increased eGFP levels in the variant pG1-Δ1, although Mxr1 (positive regulator of methanol metabolism in *Pp,* homolog of *ScADR1*) binding site deletion would be expected to rather weaken the promoter. Combined

deletion of F$ADR1.01 with F$MIG1.01 in pG1-Δ10 liberated promoter repression on glycerol and weakened its induction, which is a conclusive response for Mig1 TFBS deletion.

In the main regulatory region, the binding site F$RGT1.02 was deleted in the variants pG1-Δ6 (two sites), -Δ7, -Δ8, -Δ11 and -Δ12, and F$RGT1.01 was deleted in Δ9. The variant harboring the deletion of the paired F$RGT1.02 site (Δ6, binding sites on opposite strands with a shift of 7 bp) showed a slightly liberated repression and reduced induction. The variants Δ7 and Δ8 contain very close F$RGT1.02 sites, whereat the first lies on the negative- and the second on the positive strand; also Δ8 contains the deletion of an F$SIP4.01 site. The first (Δ7) showed a slightly liberated repression and increased induction, while the second (Δ8) was much weaker induced (but had unchanged promoter repression). This indicates a strong role for the transcriptional activator Cat8-1 and/or Cat8-2 (strongest homologs for *Sc*Sip4) for pG1 induction. The variant Δ9 was created to delete closely located F$RGT1.01 and F$CSRE.01 TFBS (binding sites on opposite strands) and the drastic loss of repression indicates a strong role of these TFBS to tightly control pG1, most likely through binding of Rgt1, Cat8-1 and/or Cat8-2. The deletion of F$RGT1.02 in the variant pG1-Δ12 did not have an effect on eGFP expression performance. Interestingly, *CAT8-2* transcription is strongly upregulated in limiting glucose compared to glucose surplus, while RGT1 and CAT8-2 were not transcriptionally regulated in the tested conditions.

Example 5: pG1 promoter strength is dependent on the poly(A:T) tract TAT14

The TAT motif is located about 80 bp upstream (5′, e.g. position -390 to -374) of the main regulatory region of pG1. Repeated sequencing of the 5'-region of GTH1 in *P. pastoris* CBS2612, CBS7435 or GS115 resulted in the detection of 15 +/- 1 Ts in the TAT motif. To elucidate its impact on promoter performance, the TAT14 motif was selected for deletion (pG1-ΔTAT14) and mutation (to T16, T18 and T20; pG1-T16, pG1-T18, pG1-T20). Primers (see primers #37-42 in **Table 4**) were initially designed to obtain T18, T20 and T22, but variants with different lengths (T16, T20 and T18, respectively) were obtained and used. Deletion of the TAT14 motif resulted in lower GFP signals, whereas its prolongation increased the expression strength of pG1. This

indicates that the use of a prolonged TAT14 motif would be beneficial for pG1 engineering.

Example 6: Partial sequence duplications of pG1´s main regulatory region significantly improve its expression strength

Two duplication variants (pG1-D1240 (SEQ ID 49) and pG1-D1427 (SEQ ID 85), the numbers state the lengths of the respective promoter variants) of the pG1 promoter were generated by PCR amplification of two sequence fragments (-472 to –188 and -472 to -1) and insertion using the restriction sites PstI and BglII (positions 509-514 and 525-530). The duplication sections start upstream of TFBS deleted in pG1-Δ5 and end after the main regulatory region of pG1 for the first variant (pG1-D1240), while the second duplication (pG1-D1427) reaches until the 3´-end of the pG1 promoter. These variants were screened for eGFP expression in the same way as described for the TFBS deletion and TAT14 mutation variants (see Example 8). Both duplication variants showed more tight repression in excess glycerol and stronger induction upon limiting glucose (**Figure 4**).

The post-transformational stability of the duplication variant clone pG1-D1240 #3 was tested by performing three consecutive batch cultivations without selection pressure, which is equal to about 20 generations. eGFP expression was stable over the whole cultivation time (data not shown). In comparison, a typical *P. pastoris* bioreactor process starts with OD_{600}=1 (~0.2-0.4 g/L YDM) in the batch phase and ends with ~100 g/L YDM after the fed batch phase and thereby takes about 10 generations.

Example 7: Verification of pG1 promoter variant performance in fed batch bioreactor cultivation

In order to verify the performance of the generated promoter variants in bioprocess conditions, some variants were selected for fed batch cultivation based on their altered eGFP expression performance: pG1-Δ2 (SEQ ID 211) was the most enhanced variant upstream of the main regulatory region, and pG1-T16 (SEQ ID 257) and pG1-D1240 (SEQ ID 49) showed higher eGFP expression levels in limiting glucose without losing promoter repression in the glycerol condition. A bioreactor cultivation, which was started with a glycerol batch phase followed by a space-time yield optimized fed batch (Prielhofer et al., 2013), was performed for one clone each

and compared to the control strain pG1 #8 for eGFP expression (see **Figure 5** and **Table 5**).

Fed batch fermentations were performed in DASGIP reactors with a final working volume of 0.7 L.

Following media were used:

PTM_1 trace salts stock solution contained per liter

6.0 g $CuSO_4 \cdot 5H_2O$, 0.08 g NaI, 3.36 g $MnSO_4 \cdot H_2O$, 0.2 g $Na_2MoO_4 \cdot 2H_2O$, 0.02 g H_3BO_3, 0.82 g $CoCl_2$, 20.0 g $ZnCl_2$, 65.0 g $FeSO_4 \cdot 7H_2O$, 0.2 g biotin and 5.0 ml H_2SO_4 (95 %-98 %).

Glycerol Batch medium contained per liter

2 g Citric acid monohydrate ($C_6H_8O_7 \cdot H_2O$), 39.2 g Glycerol, 12.6 g $NH_4H_2PO_4$, 0.5 g $MgSO_4 \cdot 7H_2O$, 0.9 g KCl, 0.022 g $CaCl_2 \cdot 2H_2O$, 0.4 mg biotin and 4.6 ml PTM1 trace salts stock solution. HCl was added to set the pH to 5.

Glucose fed batch medium contained per liter

464 g glucose monohydrate, 5.2 g $MgSO_4 \cdot 7H_2O$, 8.4 g KCl, 0.28 g $CaCl_2 \cdot 2H_2O$, 0.34 mg biotin and 10.1 mL PTM1 trace salts stock solution.

The dissolved oxygen was controlled at DO = 20 % with the stirrer speed (400 – 1200 rpm). Aeration rate was 24 L h^{-1} air, the temperature was controlled at 25°C and the pH setpoint of 5 was controlled with addition of NH_4OH (25 %).

To start the fermentation, 400 mL batch medium was sterile filtered into the fermenter and was inoculated from a selective pre-culture of the respective *P. pastoris* clone with a starting optical density (OD600) of 1. The batch phase of approximately 25 h (reaching a dry biomass concentration of approximately 20 g/L) was followed by a glucose-limited fed batch starting with an exponential feed for 7 h and a constant feed rate of 15 g/L for 13 h, leading to a final dry biomass concentration of approximately 100 g/L. Samples were taken during batch and fed batch phase, and analyzed for eGFP expression using a plate reader (Infinite 200, Tecan, CH). Therefore, samples were diluted to an optical density (OD600) of 5. Results are shown in Figure 5 as relative fluorescence per bioreactor (FL/r).

The gene copy number of these three clones was analyzed using Real-time PCR and resulted in one GCN for all of them (data not shown). All pG1-variants displayed good repression in the batch phase and strong expression in the induced state (Table 5). The strong improvement of the duplication variant pG1-D1240 could be verified in bioreactor conditions, the clone pG1-D1240 #3 showed a 50% increase in

GFP fluorescence at the fed batch end compared to pG1. Although the signal was already increased at the batch end, the induction ratio was even a bit higher than for the original pG1 Other than in the screening, the clone pG1-Δ2 #3 had a slightly increased signal at the batch end, and about 10% weakened signal at the fed batch end. The TAT14 mutation variant clone pG1-T16 #3 showed the strongest signal at the batch end, and fell behind the duplication variant at the fed batch end, reaching about 20% improvement over the control pG1 #8, similar to the screening result. The different induction behavior of the clones in the batch phase is explained by derepression due to decreasing glycerol concentration throughout the batch phase (see **Figure 5A**). Overall, the fed batch cultivations could largely confirm the results obtained in small scale screening.

Achievements and Conclusions

Gene promoters with carbon source-dependent regulation are favorable for bioprocess application because the production phase can be separated from growth. Potential promoter-based protein production improvement can be accomplished by finding the optimal growth conditions (e. g. growth rate, feeding strategy) or by directly manipulating the promoter sequence (e. g. mutations, deletions).

Several pG1 promoter variants were constructed with shortened length, TFBS deletions, TAT motif mutations and fragment duplications. Thereby, the main regulatory region of pG1, including its important TFBS was identified. The analysis of TFBS deletions indicates that the transcription factors Rgt1 and Cat8-1 and/or Cat8-2 play an essential role for pG1 repression and induction: two motifs consisting of F$RGT1 and F$CSRE binding at the same position on the opposite strands were deleted. Deletion of the first part (pG1-Δ8, position -293 to -285; RGT1: (+)-310 to -299, CSRE: (-) -299 to-285) caused weakened promoter induction, while deletion of the second part (pG1-Δ9, position -275 to -261; RGT1: (-) -275 to -259, CSRE: (+) -276 to -260) lead to decreased promoter repression. Thereby, regulatory motifs were identified which are essential and characteristic for pG1 regulation.

The role of the transcriptional regulators Mig1 (F$MIG1) and Mxr1 (F$ADR1) might be more important in other conditions such as excess glucose or methanol induction. Other transcription factors which bind in or close to that region might also contribute to pG1´s regulation.

The poly(A:T) tracts are known to play a role in promoter sequences, and the TAT motif in pG1, which is located upstream (e.g. position -390 to -375) of the main regulator region, could be shown to be essential for its strength. Elongation of this motif to T16, T18 and T20 had a positive effect on promoter performance.

Deletion variants of pG1 revealed that 5´shortening might be beneficial for promoter engineering as well. TFBS for Mxr1, Mig1, Rgt1 and Cat8 deleted upstream of the main regulatory region of pG1 improved eGFP expression, although this effect was not seen for the 5´shortened promoter variants.

Two variants with partial sequence duplications reached greatly enhanced expression capacities compared to the wild type pG1.

Distinct features of pG1 good expression performance could be assigned, which is a solid basis for rational promoter engineering: 5´shortening, TAT motif use and optional mutation/elongation and fragment duplication. pG1 variant performance in small scale screening could successfully be verified in fed batch cultivations.

Abbreviations:

CSRE: carbon source response element, F$: fungi specific TF matrix, GCN: gene copy number, GOI: gene of interest, *Pp*: *Pichia pastoris*, *Sc*: *Saccharomyces cerevisiae*, TF: transcription factor(s), TFBS: transcription factor binding site(s), YDM: yeast dry mass

Example 8: Determining the Repression, Induction, pG1-x Expression level (expression level compared to pG1), Induction ratio

The promoter strength as compared to the pG1 promoter and the induction ratio can be determined by the following standard assay: *P. pastoris* strains are screened in 24- deep well plates at 25°C with shaking at 280 rpm with 2 mL culture per well. Glucose feed beads (6mm, Kuhner, CH) are used to generate glucose-limiting growth conditions. Cells are analyzed for eGFP expression during repression (YP + 1% glycerol, exponential phase) and induction (YP + 1 feed bead, for 20-28 hours) using flow cytometry. The specific eGFP fluorescence is calculated from fluorescence intensity and forward scatter for at least 3000 data points of the flow cytometry data. Forward scatter is a relative measure for the cell volume. Specific eGFP fluorescence equals fluorescence intensity (FI) divided by forward scatter (FSC) to the 1.5, that is $FI/FSC^{1.5}$ (Hohenblum, H., N. Borth & D. Mattanovich, (2003) Assessing viability and

cell-associated product of recombinant protein producing *Pichia pastoris* with flow cytometry. *J Biotechnol* **102**: 281-290). From this data, the geometric mean of the population´s specific fluorescence is used, and normalized by subtracting background signal of non-producing *P. pastoris* wild type cells. The specific eGFP fluorescence of the glycerol condition is termed „Repression", and the specific eGFP fluorescence of the limited glucose condition (glucose feed beads) is termed „Induction".. Therefore, only Repression and Induction values of the same screening and flow cytometry measurement can be compared and used for calculations. To determine relative pG1-x promoter strength, the eGFP expression levels in the induced state of the pG1-x promoters were compared to the original pG1 promoter by dividing the Induction value of a strain comprising the pG1-x promoter by the Induction value of a strain comprising the original pG1 promoter. The Induction ratio is calculated by dividing the Induction value by the Repression value of the same strain/promoter. Repression, Induction, relative pG1-x promoter strength and Induction ratio are shown in Table 6 for several promoter variants

Further examples have proven that by using a pG1-x promoter comprising or consisting of the nucleotide sequence SEQ ID 49 a model protein (POI) was produced in *P. pastoris* at much higher yields (a fold increase of more than 3.5 fold), fed-batch experiments) as compared to the unmodified pG1 promoter (reference SEQ ID 7).

Example 9: Comparison of "Speed fermentation" and standard fermentation

Summary: Significantly reduced fermentation times could be obtained for the expression of an alternative scaffold protein as model protein under control of a pG1-3 embodiment of SEQ ID 39 (pG1-D1240 (SEQ ID 49)) promoter by employing a space-time yield optimized fed batch protocol instead of using a standard fed batch regime.

A clone expressing a model protein under control of pG1-D1240 (SEQ ID 49) was selected for the fed batch cultivations. Fed batch cultivations were performed in DASGIP reactors (Eppendorf, Germany) with a final working volume of 0.5 L. Media and trace element solution were prepared as previously described in Example 7, except for the glycerol concentration in the glycerol batch medium which was 45 g/L. During cultivation the dissolved oxygen level was controlled at DO = 30 % with the stirrer speed (400 –1200 rpm). Aeration rate was 1 vvm air, the temperature was controlled at 25°C and the pH set-point of 5.0 was controlled with addition of NH_4OH

(25 %). To start the bioreactor cultivation, 250 mL batch medium were inoculated from a pre-culture of the respective *P. pastoris* clone with a starting optical density (OD600) of 1.0. The batch phase on glycerol took approximately 30 h and reached a dry biomass concentration of 25 - 29 g/L. The glycerol batch phase was followed by a glucose-limited fed batch. Two different fed batch cultivation modes were compared: (**A**) a standard fed batch protocol using a constant feed rate, (**B**) a space-time yield optimized fed batch protocol ("Speed fermentation"), where the glucose feed rate was optimized to maximize the volumetric productivity of the fermentation.

For the standard cultivation, a constant glucose feed rate of 1.25 mL h^{-1} was selected. The fed batch cultivation was maintained for 100 h (126 h total cultivation time) resulting in a final dry biomass concentration of approximately 90 g L^{-1}. For the "Speed fermentation", a model-based optimization algorithm (Maurer et al., Microbial Cell Factories, 2006, 5:37) was adopted, where the optimized volumetric glucose feed rate F(t) was approximated by a linearly increasing function: F(t) [mL h^{-1}]= 0.3234 mL h^{-2} * t + 3.3921 mL h^{-1}. The fed batch phase was maintained for t= 33 h (60 h total cultivation time), which resulted in a final dry biomass concentration of approximately 140 g L^{-1}.

Samples were taken at the end of the batch and during the fed batch phase. Product titers were analyzed from clarified supernatants using a HT low MW protein express reagent kit and the Caliper LabChip GXI system (Perkin Elmer, USA). As a reference standard for absolute quantification a purified standard of alternative scaffold protein was used.

Figure 9 shows the product and biomass generation over the total cultivation time for the standard cultivation (**A**) and the "Speed fermentation" (**B**). In comparison, final product titers of 6.4 g L^{-1} and 4.3 g L^{-1} could be reached after 60 h and 126 h for the "Speed fermentation" and the standard fermentation, respectively. In other words, a 1.4-fold higher titer (resp. 1.2- fold higher broth titers) could be found in significantly shorter fermentation time (-66 h) when supplementing the glucose feed during expression under the pG1-D1240 (SEQ ID 49) promoter as described for the "Speed fermentation" instead of using the described standard feed regime.

Tables

Table 1: TFBS identified in the pG1 promoter sequence using MatInspector. Targeted carbon source-related TFBS of the pG1 deletion variants are shown in bold.

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Start position	End position	Strand	Sequence SEQ ID NO.
F$TEAF	TEA/ATTS DNA binding domain factors	F$ABAA.01	Aspergillus spore/developmental regulator	-985	-969	-	accctaCATTctactgg (SEQ ID 271)
F$NRGF	NRG zinc finger factors	F$NRG1.01	Transcriptional repressor Nrg1	-976	-964	+	tgtAGGGtcccca (SEQ ID 272)
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	-956	-942	-	gagactaGGGGgagc (SEQ ID 273)
F$PDRE	Pleiotropic drug resistance responsive elements	F$PDRE.01	Pleiotropic drug resistance responsive element (yeast)	-944	-936	-	TCCCtggag (SEQ ID 274)
F$YMAT	Yeast mating factors	F$HMRA2.01	Hidden Mat Right A2, a2 is one of two genes encoded by the a mating type cassette in S. cerevisiae	-939	-927	+	gggaaaTGTAaaa (SEQ ID 275)
F$MADS	Yeast MADS-Box factors	F$RLM1.01	Yeast MADS-Box RLM1 transcription factor	-926	-908	-	gtttTCTAttagcagtata (SEQ ID 276)
O$INRE	Core promoter initiator elements	O$DINR.01	Drosophila initiator motifs	-899	-889	+	gcTCAGttgtc (SEQ ID 277)
F$RFXP	Regulatory factor X protein, homologous to mammalian RFX1-5	F$RFX1.02	RFX1 (CRT1), acts by recruiting Ssn6 and Tup1, general repressors to the promoters of damage-inducible genes	-896	-882	-	ttatcctgaCAACtg (SEQ ID 278)

F$HOMD	Homeodomain-containing transcriptional regulators	F$YOX1.02	Yeast homeobox 1, homeodomain-containing transcriptional repressor	-889	-875	-	aacgtaATTAtcctg (SEQ ID 279)
F$HOMD	Homeodomain-containing transcriptional regulators	F$YOX1.02	Yeast homeobox 1, homeodomain-containing transcriptional repressor	-888	-874	+	aggataATTAcgttc (SEQ ID 280)
O$MTEN	Core promoter motif ten elements	O$DMTE.01	Drosophila motif ten element	-888	-868	-	acagtcgAACGtaattatcct (SEQ ID 281)
F$BZIP	Fungal basic leucine zipper family	F$CST6.01	Chromosome stability, bZIP transcription factor of the ATF/CREB family (ACA2)	-885	-865	-	actacagtcgaACGTaattat (SEQ ID 282)
F$MADS	Yeast MADS-Box factors	F$RLM1.01	Yeast MADS-Box RLM1 transcription factor	-855	-837	-	tcttTCTAacaatacagat (SEQ ID 283)
F$YMAT	Yeast mating factors	F$MATALPHA2.02	Homeodomain transcriptional repressor Matalpha2	-853	-841	+	ctgtaTTGTtaga (SEQ ID 284)
F$MMAT	M-box interacting with Mat1-Mc	F$MAT1MC.01	HMG-BOX protein interacts with M-box site, cooperativity with HMG-Box STE11 protein	-852	-842	+	tgtATTGttag (SEQ ID 285)
F$STPF	STP gene family	F$STP2.01	Proteolytically activated transcription factor	-828	-814	-	gcggcGCCGtaaaaa (SEQ ID 286)
F$STPF	STP gene family	F$STP2.01	Proteolytically activated transcription factor	-823	-809	+	acggcGCCGccatat (SEQ ID 287)
F$YADR	**Yeast metabolic regulator**	**F$ADR1.01**	**Alcohol Dehydrogenase Regulator, carbon source-responsive zinc-finger transcription factor**	**-785**	**-777**	**+**	**aaCCCCact** (SEQ ID 288)

F$RFXP	Regulatory factor X protein, homologous to mammalian RFX1-5	F$RFX1.01	RFX1 (CRT1) is a DNA-binding protein that acts by recruiting Ssn6 and Tup1, general repressors to the promoters of damage-inducible genes	-763	-749	-	cgtgtataGCAAcag (SEQ ID 289)
F$YMCB	Yeast Mlu I cell cycle box	F$SWI4.01	DNA binding component of the SBF(SCB binding factor) complex (Swi4p-Swi6p)	-756	-744	+	tatacaCGAAcca (SEQ ID 290)
F$CYTO	Activator of cytochrome C	F$HAP1.01	HAP1, S. cerevisiae member of GAL family, regulates heme dependent cytochrome expression	-715	-701	+	ctgaagtcATCGgtt (SEQ ID 291)
F$FKHD	Fungal fork head transcription factors	F$FKH1.01	Fork head transcription factor Fkh1	-709	-693	+	tcatcggTTAAcaatca (SEQ ID 292)
F$ROX1	Repressor of hypoxic genes	F$ROX1.01	Heme-dependent transcriptional repressor of hypoxic genes	-704	-692	-	ttgaTTGTtaacc (SEQ ID 293)
F$YMAT	Yeast mating factors	F$MATALPHA2.02	Homeodomain transcriptional repressor Matalpha2	-703	-691	-	cttgaTTGTtaac (SEQ ID 294)
F$MMAT	M-box interacting with Mat1-Mc	F$MAT1MC.01	HMG-BOX protein interacts with M-box site, cooperativity with HMG-Box STE11 protein	-702	-692	-	ttgATTGttaa (SEQ ID 295)
F$YHSF	Yeast heat shock factors	F$HSF1.01	Trimeric heat shock transcription factor	-678	-646	-	aacacctactgaatatGGAAaggagcattcaga (SEQ ID 296)
F$PHD1	Pseudohyphal determinant 1	F$PHD1.03	Transcription factor involved in regulation of filamentous growth	-635	-623	-	gcaGTGCatgcaa (SEQ ID 297)

F$MGCM	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**	**-628**	**-612**	**+**	**cactgCGGAagaattag** (SEQ ID 298)
F$CSRE	**Carbon source-responsive elements**	**F$CSRE.01**	**Carbon source-responsive element (yeast)**	**-626**	**-612**	**-**	**ctaattctTCCGcag** (SEQ ID 299)
F$YRSC	Yeast transcription factors remodeling chromatin structure	F$RSC3.01	Component of the RSC chromatin remodeling complex	-614	-594	+	tagccaatagCGCGtttcata (SEQ ID 300)
F$YMCB	Yeast Mlu I cell cycle box	F$STUAP.01	Aspergillus Stunted protein, (bHLH)-like structure, regulates multicellular complexity during asexual reproduction	-609	-597	-	gaaaCGCGctatt (SEQ ID 301)
F$YMCB	Yeast Mlu I cell cycle box	F$MCB.01	Mlu I cell cycle box, activates G1/S-specific transcription (yeast)	-608	-596	+	atagCGCGttca (SEQ ID 302)
F$DUIS	DAL upstream induction sequence	F$DAL82.01	Transcriptional activator for allantoin catabolic genes	-597	-589	+	cataTGCGc (SEQ ID 303)
F$PHD1	Pseudohyphal determinant 1	F$PHD1.02	Transcription factor involved in regulation of filamentous growth	-597	-585	+	cataTGCGctttt (SEQ ID 304)
F$RDNA	RDNA binding factor	F$REB1.02	rDNA enhancer binding protein 1, termination factor for RNA polymerase I and transcription factor for RNA polymerase II	-589	-577	+	cttTTACcccctc (SEQ ID 305)

F$YMIG	**Yeast GC-Box Proteins**	**F$MIG1.02**	**MIG1, zinc finger protein mediates glucose repression**	**-586**	**-568**	**-**	**ttgacaaaagaGGGGgtaa** (SEQ ID 306)
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	-586	-572	-	caaaagaGGGGgtaa (SEQ ID 307)
F$BZIP	Fungal basic leucine zipper family	F$YAP1.02	Yeast activator protein of the basic leucine zipper (bZIP) family	-585	-565	+	taccccctctttt GTCAagcg (SEQ ID 308)
F$TALE	Fungal TALE homeodomain class	F$TOS8.01	Homeodomain-containing transcription factor	-579	-567	+	ctcttttGTCAag (SEQ ID 309)
F$DUIS	DAL upstream induction sequence	F$DAL82.01	Transcriptional activator for allantoin catabolic genes	-567	-559	-	atttTGCGc (SEQ ID 310)
F$YMIG	**Yeast GC-Box Proteins**	**F$MIG1.01**	**MIG1, zinc finger protein mediates glucose repression**	**-553**	**-535**	**+**	**taagatttggtGGGGgtgt** (SEQ ID 311)
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type I	F$RAP1.06	RAP1 (TUF1), activator or repressor depending on context	-546	-524	-	gctaacggctcaCACCccacca (SEQ ID 312)
F$IRTF	Iron-responsive transcriptional activators	F$AFT2.01	Activator of Fe (iron) transcription 2, iron-regulated transcriptional activator	-543	-529	-	cggctcaCACCccca (SEQ ID 313)
O$VTBP	Vertebrate TATA binding protein factor	O$ATATA.01	Avian C-type LTR TATA box	-530	-514	-	ttgtactTCAGctaacg (SEQ ID 314)
F$RRPE	Ribosomal RNA processing element	F$STB3.01	Ribosomal RNA processing element (RRPE)-binding protein	-504	-488	-	tgcagtttTTTCaggga (SEQ ID 315)
F$MGCM	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**	**-442**	**-426**	**-**	**atatcAGGAaaaacata** (SEQ ID 316)

F$GATA	Fungal GATA binding factors	F$GZF3.01	GATA zinc finger protein Gzf3	-434	-420	+	tcctGATAtgcatca (SEQ ID 317)
F$PHD1	Pseudohyphal determinant 1	F$PHD1.01	Transcription factor involved in regulation of filamentous growth	-430	-418	+	gataTGCAtcaaa (SEQ ID 318)
F$YMAT	Yeast mating factors	F$MATA1.01	Homeodomain protein mating factor a1	-429	-417	-	ttttGATGcatat (SEQ ID 319)
F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$CHA4.01	Fungal zinc cluster transcription factor Cha4, single triplet	-408	-388	+	taaaacctgaatctCCGCtat (SEQ ID 320)
F$MGCM	Monomeric Gal4-class motifs	F$YRR1.01	Zinc cluster transcription factor, activates genes involved in multidrug resistance (PDR2)	-403	-387	-	aatagCGGAgattcagg (SEQ ID 321)
F$RDR1	Repressor of Drug Resistance 1	F$RDR1.01	Repressor of Drug Resistance 1 (transcriptional repressor involved in the control of multidrug resistance	-399	-389	-	tagCGGAgatt (SEQ ID 322)
F$RFXP	Regulatory factor X protein, homologous to mammalian RFX1-5	F$RFX1.02	RFX1 (CRT1), acts by recruiting Ssn6 and Tup1, general repressors to the promoters of damage-inducible genes	-366	-352	-	ttgtcacgaAAACgg (SEQ ID 323)
F$YMCB	Yeast Mlu I cell cycle box	F$SWI4.01	DNA binding component of the SBF(SCB binding factor) complex (Swi4p-Swi6p)	-364	-352	-	ttgtcaCGAAaac (SEQ ID 324)
F$BZIP	Fungal basic leucine zipper family	F$YAP1.02	Yeast activator protein of the basic leucine zipper (bZIP) family	-361	-345	-	tggaaattaatttGTCAcgaa (SEQ ID 325)

F$RRPE	Ribosomal RNA processing element	F$STB3.01	Ribosomal RNA processing element (RRPE)-binding protein	-359	-347	-	aattaattTGTCacgaa (SEQ ID 326)
F$TALE	Fungal TALE homeodomain class	F$CUP9.01	Homeodomain transcriptional repressor Cup9	-361	-341	-	ttaattTGTCacg (SEQ ID 327)
F$HOMD	Homeodomain-containing transcriptional regulators	F$YOX1.01	Yeast homeobox 1, homeodomain-containing transcriptional repressor	-358	-344	-	aaattAATTtgtcac (SEQ ID 328)
F$HOMD	Homeodomain-containing transcriptional regulators	F$YOX1.01	Yeast homeobox 1, homeodomain-containing transcriptional repressor	-357	-343	+	tgacaAATTaatttc (SEQ ID 329)
F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$TEA1.01	Ty1 enhancer activator, zinc cluster DNA-binding protein	-357	-337	+	tgacaaaTTAAtttccaacgg (SEQ ID 330)
F$MGCM	Monomeric Gal4-class motifs	F$YRR1.01	Zinc cluster transcription factor, activates genes involved in multidrug resistance (PDR2)	-352	-336	-	cccgtTGGAaattaatt (SEQ ID 331)
F$ASG1	Activator of stress genes	F$ASG1.01	Fungal zinc cluster transcription factor Asg1	-340	-324	-	tCCGGacaagaccccgt (SEQ ID 332)
F$MGCM	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**	**-337**	**-321**	-	**ttatcCGGAcaagaccc** (SEQ ID 333)
F$MGCM	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**	**-330**	**-320**	+	**ttgtcCGGAtaagagaa** (SEQ ID 334)

F$RDR1	Repressor of Drug Resistance 1	F$RDR1.01	Repressor of Drug Resistance 1 (transcriptional repressor involved in the control of multidrug resistance	-332	-316	+	gtcCGGAtaag (SEQ ID 335)
F$GATA	Fungal GATA binding factors	F$GATA.01	GATA binding factor (yeast)	-329	-315	+	tccgGATAagagaat (SEQ ID 336)
F$PRES	Pheromone response elements	F$STE12.01	Transcription factor activated by a MAP kinase signaling cascade, activates genes involved in mating or pseudohyphal/ invasive growth pathways	-315	-303	-	taatcaAACAaaa (SEQ ID 337)
F$GATA	Fungal GATA binding factors	F$GAT1.01	GATA-type Zn finger protein Gat1	-311	-297	-	aacggATAAtcaaac (SEQ ID 338)
F$MGCM	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**	**-310**	**-294**	**-**	**ccgaaCGGAtaatcaaa** (SEQ ID 339)
O$MTEN	Core promoter motif ten elements	O$DMTE.01	Drosophila motif ten element	-310	-290	-	ttatccgAACGgataatcaaa (SEQ ID 340)
F$YORE	Yeast oleate response elements	F$OAF1.01	Oleate-activated transcription factor, acts alone and as a heterodimer with Pip2p	-307	-283	-	cgtccatttaTCCGaacggataatc (SEQ ID 341)
F$MGCM	Monomeric Gal4-class motifs	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**	**-299**	**-289**	**+**	**ccgttCGGAtaaatgga** (SEQ ID 342)

F$YGAL	Yeast GAL4 factor	F$GAL4.01	GAL4 transcriptional activator in response to galactose induction	-301	-285	-	agcaggcgtccatttatCCGAacgg (SEQ ID 343)
F$CSRE	**Carbon source-responsive elements**	**F$SIP4.01**	**Zinc cluster transcriptional activator, binds to the carbon source-responsive element (CSRE) of gluconeogenic genes**	**-299**	**-285**	**-**	**tCCATttatccgaac** (SEQ ID 344)
F$RDR1	Repressor of Drug Resistance 1	F$RDR1.01	Repressor of Drug Resistance 1 (transcriptional repressor involved in the control of multidrug resistance	-301	-277	+	gttCGGAtaaa (SEQ ID 345)
F$YGAL	Yeast GAL4 factor	F$LAC9.01	LAC9 binding site, homologous to GAL4 of Saccharomyces cerevisiae	-299	-275	+	gttCGGAtaaatggacgcctgctcc (SEQ ID 346)
F$FBAS	Fungi branched amino acid biosynthesis	F$LEU3.02	LEU3, S. cerevisiae, zinc cluster protein	-275	-261	-	taaCCGGaaaaatatgg (SEQ ID 347)
F$CSRE	**Carbon source-responsive elements**	**F$CSRE.01**	**Carbon source-responsive element (yeast)**	**-276**	**-260**	**+**	**catattttTCCGgtt** (SEQ ID 348)
F$MGCM	**Monomeric Gal4-class motifs**	**F$RGT1.01**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**	**-275**	**-259**	**-**	**ataacCGGAaaaatatg** (SEQ ID 349)
F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$TEA1.01	Ty1 enhancer activator, zinc cluster DNA-binding protein	-269	-249	-	aggtgggGTAAtaaccggaaa (SEQ ID 350)

F$RDNA	RDNA binding factor	F$REB1.02	rDNA enhancer binding protein 1, termination factor for RNA polymerase I and transcription factor for RNA polymerase II	-262	-250	+	ttaTTACcccacc (SEQ ID 351)
F$YMCM	Yeast cell cycle and metabolic regulator	F$MCM1.02	Yeast factor MCM1 cooperating with MATalpha factors	-258	-250	-	cTTCCaggtggggtaat (SEQ ID 352)
F$YMIG	**Yeast GC-Box Proteins**	**F$MIG1.01**	**MIG1, zinc finger protein mediates glucose repression**	**-260**	**-244**	**-**	**cacttccaggtGGGGtaat** (SEQ ID 353)
F$YADR	**Yeast metabolic regulator**	**F$ADR1.01**	**Alcohol Dehydrogenase Regulator, carbon source-responsive zinc-finger transcription factor**	**-260**	**-242**	**+**	**taCCCCacc** (SEQ ID 354)
F$MGCM	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**	**-239**	**-223**	**-**	**atcccCGGAaaattctg** (SEQ ID 355)
F$YMIG	**Yeast GC-Box Proteins**	**F$MIG1.01**	**MIG1, zinc finger protein mediates glucose repression**	**-239**	**-221**	**+**	**cagaattttccGGGGatta** (SEQ ID 356)
F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$TEA1.01	Ty1 enhancer activator, zinc cluster DNA-binding protein	-232	-224	-	attatccGTAAtccccggaaa (SEQ ID 357)
F$ARPU	Regulator of pyrimidine and purine utilization pathway	F$PPR1.01	Pyrimidine pathway regulator 1	-231	-223	-	atccgtaatccCCGGaa (SEQ ID 358)
F$PDRE	Pleiotropic drug resistance responsive elements	F$PDRE.01	Pleiotropic drug resistance responsive element (yeast)	-232	-216	-	TCCCcggaa (SEQ ID 359)

F$ARPU	Regulator of pyrimidine and purine utilization pathway	F$PPR1.01	Pyrimidine pathway regulator 1	-231	-215	+	tccggggattaCGGAta (SEQ ID 360)
F$PDRE	Pleiotropic drug resistance responsive elements	F$PDRE.01	Pleiotropic drug resistance responsive element (yeast)	-230	-216	+	TCCGgggat (SEQ ID 361)
F$CYTO	Activator of cytochrome C	F$HAP1.01	HAP1, S. cerevisiae member of GAL family, regulates heme dependent cytochrome expression	-233	-213	+	ccggggatTACGgat (SEQ ID 362)
F$YQA1	Neurospora crassa QA1 gene activator	F$QA1F.01	qa-1F, required for quinic acid induction of transcription in the qa gene cluster	-228	-208	+	ggggattacggaTAATacggt (SEQ ID 363)
F$MGCM	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**	**-225**	**-209**	**+**	**gattaCGGAtaatacgg** (SEQ ID 364)
F$CYTO	Activator of cytochrome C	F$HAP1.01	HAP1, S. cerevisiae member of GAL family, regulates heme dependent cytochrome expression	-221	-207	+	acggataaTACGgtg (SEQ ID 365)
F$BZIP	Fungal basic leucine zipper family	F$CIN5.01	bZIP transcriptional factor of the yAP-1 family that mediates pleiotropic drug resistance and salt tolerance	-208	-188	+	tggtctggattaatTAATacg (SEQ ID 366)

F$BZIP	Fungal basic leucine zipper family	F$CIN5.01	bZIP transcriptional factor of the yAP-1 family that mediates pleiotropic drug resistance and salt tolerance	-203	-189	-	cttggcgtattaatTAATcca (SEQ ID 367)
F$HOMD	Homeodomain-containing transcriptional regulators	F$YOX1.02	Yeast homeobox 1, homeodomain-containing transcriptional repressor	-202	-188	-	gtattaATTAatcca (SEQ ID 368)
F$HOMD	Homeodomain-containing transcriptional regulators	F$YOX1.02	Yeast homeobox 1, homeodomain-containing transcriptional repressor	-203	-183	+	ggattaATTAatacg (SEQ ID 369)
F$YABF	Yeast ABF factors	F$ABF1.04	ARS (autonomously replicating sequence)-binding factor I	-202	-184	+	ggATTAattaatacgccaa (SEQ ID 370)
F$PHRR	pH responsive regulators	F$RIM101.01	Transcriptional repressor involved in response to pH and in cell wall construction	-192	-176	+	atacGCCAagtcttaca (SEQ ID 371)
F$PRES	Pheromone response elements	F$STE12.01	Transcription factor activated by a MAP kinase signaling cascade, activates genes involved in mating or pseudohyphal/ invasive growth pathways	-175	-163	-	gactgcAACAaaa (SEQ ID 372)
F$FKHD	Fungal fork head transcription factors	F$FKH2.01	Fork head transcription factor Fkh2	-148	-132	+	gcaataaTAAAcaagat (SEQ ID 373)
F$YCAT	Yeast CCAAT binding factors	F$HAP234.01	Yeast factor complex HAP2/3/5, homolog to vertebrate NF-Y/CP1/CBF	-124	-112	-	ctaatCCAAtaaa (SEQ ID 374)

F$YORE	Yeast oleate response elements	F$ORE.01	Oleate response element, binding motif of Oaf1 homodimers or Oaf1/Pip2 heterodimers	-120	-96	-	CGGGgtcaagctgcaactaatccaa (SEQ ID 375)
F$AAAU	A. nidulans activator of acetate utilization genes	F$FACBCB.01	FACB, activator of acetate utilization genes with a GAL4-type Zn(II)2Cys6 zinc binuclear cluster	-109	-93	+	GCAGcttgaccccgcca (SEQ ID 376)
F$YMIG	Yeast GC-Box Proteins	F$MIG3.01	Zinc finger transcriptional repressor MIG3	-104	-86	-	ctagctatggcGGGGtcaa (SEQ ID 377)
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type I	F$RAP1.06	RAP1 (TUF1), activator or repressor depending on context	-74	-52	-	tgcatcatctaaCACCcatagca (SEQ ID 378)
F$PHD1	Pseudohyphal determinant 1	F$PHD1.03	Transcription factor involved in regulation of filamentous growth	-60	-48	-	caaGTGCatcatc (SEQ ID 379)
O$VTBP	Vertebrate TATA binding protein factor	O$VTATA.01	Cellular and viral TATA box elements	-31	-15	+	gagtaTAAAagatcctt (SEQ ID 380)
F$MGCM	Monomeric Gal4-class motifs	F$LYS14.01	Transcriptional activator involved in regulation of genes of the lysine biosynthesis pathway	-17	-1	-	aagggtGGAAttttaag (SEQ ID 381)

Table 2: Affected TFBS of the pG1 promoter sequence in the deletion mutants pG1-Δ1 to Δ12. Sequence analysis was done using MatInspector from Genomatix. Glucose- and carbon- related TFBS which were selected for deletion are shown in bold and the corresponding ID (1-12) and deleted positions are stated in column 1 and 2.

Deletion	Position	Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information
1	**-785 to -777**	**F$YADR**	**Yeast metabolic regulator**	**F$ADR1.01**	**Alcohol Dehydrogenase Regulator, carbon source-responsive zinc-finger transcription factor**
	-628 to -612	F$PHD1	Pseudoh yphal determinant 1	F$PHD1.03	Transcription factor involved in regulation of filamentous growth
2		**F$MGCM**	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**
		F$CSRE	**Carbon source-responsive elements**	**F$CSRE.01**	**Carbon source-responsive element (yeast)**
	-586 to -568	F$RDNA	RDNA binding factor	F$REB1.02	rDNA enhancer binding protein 1, termination factor for RNA polymerase I and transcription factor for RNA polymerase II
3		**F$YMIG**	**Yeast GC-Box Proteins**	**F$MIG1.02**	**MIG1, zinc finger protein mediates glucose repression**
		F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response
		F$BZIP	Fungal basic leucine zipper family	F$YAP1.02	Yeast activator protein of the basic leucine zipper (bZIP) family
		F$TALE	Fungal TALE homeodomain class	F$TOS8.01	Homeodomain-containing transcription factor
4	**-553 to -535**	**F$YMIG**	**Yeast GC-Box Proteins**	**F$MIG1.01**	**MIG1, zinc finger protein mediates glucose repression**
		F$YRAP	Yeast activator of glycolyse genes / repressor of mating type I	F$RAP1.06	RAP1 (TUF1), activator or repressor depending on context
		F$IRTF	Iron-responsive transcriptional activators	F$AFT2.01	Activator of Fe (iron) transcription 2, iron-regulated transcriptional activator

5	**-442 to -426**	**F$MGCM**	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**
		F$GATA	Fungal GATA binding factors	F$GZF3.01	GATA zinc finger protein Gzf3
		F$PHD1	Pseudoh yphal determinant 1	F$PHD1.01	Transcription factor involved in regulation of filamentous growth
	-337 to -316	F$ASG1	Activator of stress genes	F$ASG1.01	Fungal zinc cluster transcription factor Asg1
6		**F$MGCM**	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**
		F$MGCM	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**
		F$RDR1	Repressor of Drug Resistance 1	F$RDR1.01	Repressor of Drug Resistance 1 (transcriptional repressor involved in the control of multidrug resistance
		F$GATA	Fungal GATA binding factors	F$GATA.01	GATA binding factor (yeast)
	-310 to -299	F$PRES	Pheromone response elements	F$STE12.01	Transcription factor activated by a MAP kinase signaling cascade, activates genes involved in mating or pseudohyphal/invasive growth pathways
		F$GATA	Fungal GATA binding factors	F$GAT1.01	GATA-type Zn finger protein Gat1
7		**F$MGCM**	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**
		O$MTEN	Core promoter motif ten elements	O$DMTE.01	Drosophila motif ten element
		F$YORE	Yeast oleate response elements	F$OAF1.01	Oleate-activated transcription factor, acts alone and as a heterodimer with Pip2p
		F$MGCM	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**
		F$YGAL	Yeast GAL4 factor	F$GAL4.01	GAL4 transcriptional activator in response to galactose induction
8	**-293 to -285**	**F$CSRE**	**Carbon source-responsive elements**	**F$SIP4.01**	**Zinc cluster transcriptional activator, binds to the carbon source-responsive element (CSRE) of gluconeogenic genes**
		F$RDR1	Repressor of Drug Resistance 1	F$RDR1.01	Repressor of Drug Resistance 1 (transcriptional repressor involved in the control of multidrug resistance
		F$YGAL	Yeast GAL4 factor	F$LAC9.01	LAC9 binding site, homologous to GAL4 of Saccharomyces cerevisiae

		F$FBAS	Fungi branched amino acid biosynthesis	F$LEU3.02	LEU3, S. cerevisiae, zinc cluster protein
9	**-275 to -261**	**F$CSRE**	**Carbon source-responsive elements**	**F$CSRE.01**	**Carbon source-responsive element (yeast)**
		F$MGCM	**Monomeric Gal4-class motifs**	**F$RGT1.01**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**
		F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$TEA1.01	Ty1 enhancer activator, zinc cluster DNA-binding protein
		F$RDNA	RDNA binding factor	F$REB1.02	rDNA enhancer binding protein 1, termination factor for RNA polymerase I and transcription factor for RNA polymerase II
		F$YMCM	Yeast cell cycle and metabolic regulator	F$MCM1.02	Yeast factor MCM1 cooperating with MATalpha factors
10	**-258 to -242**	**F$YMIG**	**Yeast GC-Box Proteins**	**F$MIG1.01**	**MIG1, zinc finger protein mediates glucose repression**
		F$YADR	**Yeast metabolic regulator**	**F$ADR1.01**	**Alcohol Dehydrogenase Regulator, carbon source-responsive zinc-finger transcription factor**
11	**-239 to -221**	**F$MGCM**	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**
		F$YMIG	**Yeast GC-Box Proteins**	**F$MIG1.01**	**MIG1, zinc finger protein mediates glucose repression**
		F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$TEA1.01	Ty1 enhancer activator, zinc cluster DNA-binding protein
		F$ARPU	Regulator of pyrimidine and purine utilization pathway	F$PPR1.01	Pyrimidine pathway regulator 1
		F$PDRE	Pleiotropic drug resistance responsive elements	F$PDRE.01	Pleiotropic drug resistance responsive element (yeast)
		F$ARPU	Regulator of pyrimidine and purine utilization pathway	F$PPR1.01	Pyrimidine pathway regulator 1
		F$PDRE	Pleiotropic drug resistance responsive elements	F$PDRE.01	Pleiotropic drug resistance responsive element (yeast)

		F$CYTO	Activator of cytochrome C	F$HAP1.01	HAP1, S. cerevisiae member of GAL family, regulates heme dependent cytochrome expression
		F$YQA1	Neurospora crassa QA1 gene activator	F$QA1F.01	qa-1F, required for quinic acid induction of transcription in the qa gene cluster
12	**-220 to -209**	**F$MGCM**	**Monomeric Gal4-class motifs**	**F$RGT1.02**	**Glucose-responsive transcription factor involved in regulation of glucose transporters**
		F$CYTO	Activator of cytochrome C	F$HAP1.01	HAP1, S. cerevisiae member of GAL family, regulates heme dependent cytochrome expression

Table 3: Positions and TFBS deletions of pG1 TFBS deletion variants

Targeted and affected TFBS in pG1 TFBS deletion variants (pG1-Δ1 to Δ12) are listed. Targeted carbon source-related TFBS are shown in bold. Detailed information for all TFBS and for the deleted TFBS is provided in Table 1 and Table 2, respectively.

pG1-Δ	Position	TFBS Deletions (TF Matrices)
1	**-785 to -777**	**F$ADR1.01**
2	**-628 to -612**	F$PHD1.03, **F$RGT1.02, F$CSRE.01**
3	**-586 to -568**	F$REB1.02, **F$MIG1.02**, F$MSN2.01, F$YAP1.02, F$TOS8.01
4	**-553 to -535**	**F$MIG1.01**, F$RAP1.06, F$AFT2.01
5	**-442 to -426**	**F$RGT1.02**, F$GZF3.01, F$PHD1.01
6	**-337 to -316**	F$ASG1.01, **F$RGT1.02, F$RGT1.02**, F$RDR1.01, F$GATA.01
7	**-310 to -299**	F$STE12.01, F$GAT1.01, **F$RGT1.02**, O$DMTE.01, F$OAF1.01
8	**-293 to -285**	F$OAF1.01, **F$RGT1.02**, F$GAL4.01, **F$SIP4.01**, F$RDR1.01, F$LAC9.01
9	**-275 to -261**	F$LEU3.02, **F$CSRE.01, F$RGT1.01**, F$TEA1.01
10	**-258 to -242**	F$REB1.02, F$MCM1.02, **F$MIG1.01**, F$ADR1.01
11	**-239 to -221**	**F$RGT1.02, F$MIG1.01**, F$TEA1.01, F$PPR1.01, F$PDRE.01, F$PPR1.01, F$PDRE.01
12	**-220 to -209**	F$HAP1.01, F$QA1F.01, **F$RGT1.02**, F$HAP1.01

Table 4: Primer sequences

#	Name	Product	Sequence (SEQ ID NO.)	T_M
1	pG1_fw	pG1	GATAGGGCCCCAAACATTTGCTCCCCCTAGTCTC (SEQ ID 382)	71
2	pG1_back	pG1/pG1-s	GATACCTGCAGGAAGGGTGGAATTTTAAGGATCTTTTAT (SEQ ID 383)	70
3	pG1-858_fw	pG1-s858	GATAGGGCCCGGAATCTGTATTGTTAGAAAGAACGAGAG (SEQ ID 384)	71
4	pG1-663_fw	pG1-s663	GATAGGGCCCCCATATTCAGTAGGTGTTTCTTGCAC (SEQ ID 385)	69
5	pG1-492_fw	pG1-s492	GATAGGGCCCCTGCAGATAGACTTCAAGATCTCAGG (SEQ ID 386)	69
6	pG1-371_fw	pG1-s371	GATAGGGCCCGACCCCGTTTTCGTGACAAATT (SEQ ID 387)	70
7	pG1-328_fw	pG1-s328	GATAGGGCCCCCGGATAAGAGAATTTTGTTTGATTAT (SEQ ID 388)	70
8	pG1-283_fw	pG1-s283	GATAGGGCCCGCCTGCTCCATATTTTTCCGG (SEQ ID 389)	71
9	pG1-211_fw	pG1-s211	GATAGGGCCCCGGTGGTCTGGATTAATTAATACG (SEQ ID 390)	68
10	pG1-66_fw	pG1-s66	GATAGGGCCCGTGTTAGATGATGCACTTGGATGC (SEQ ID 391)	68
11	pG1-Δ1_fw	pG1-Δ1	GAAAACAGCTTGAACTTTCAAAGGTTCTGTTGCTATACAC GAAC (SEQ ID 392)	69
12	pG1-Δ1_bw	pG1-Δ1	GTTCGTGTATAGCAACAGAACCTTTGAAAGTTCAAGCTG TTTTCACACGGCC (SEQ ID 393)	68
13	pG1-Δ2_fw	pG1-Δ2	GTAGGTGTTTCTTGCACTTTTGCATGCCAATAGCGCGTT TCATATGC (SEQ ID 394)	67
14	pG1-Δ2_bw	pG1-Δ2	GCATATGAAACGCGCTATTGGCATGCAAAAGTGCAAGAA ACACCTAC (SEQ ID 395)	68
15	pG1-Δ3_fw	pG1-Δ3	CGCGTTTCATATGCGCTTGCGCAAAATGCCTGTAAGATT TG (SEQ ID 396)	68
16	pG1-Δ3_bw	pG1-Δ3	CAAATCTTACAGGCATTTTGCGCAAGCGCATATGAAACG CG (SEQ ID 397)	65
17	pG1-Δ4_fw	pG1-Δ4	GTCAAGCGCAAAATGCCTGGAGCCGTTAGCTGAAGTAC AACAG (SEQ ID 398)	65
18	pG1-Δ4_bw	pG1-Δ4	CTGTTGTACTTCAGCTAACGGCTCCAGGCATTTTGCGCT TGAC (SEQ ID 399)	67
19	pG1-Δ5_fw	pG1-Δ5	GGGATTCCCACTATTTGGTATTCTGAGCATCAAAACTCTA ATCTAAAACCTGAATCTC (SEQ ID 400)	67
20	pG1-Δ5_bw	pG1-Δ5	GAGATTCAGGTTTTAGATTAGAGTTTTGATGCTCAGAATA CCAAATAGTGGGAATCCC (SEQ ID 401)	68
21	pG1-Δ6_fw	pG1-Δ6	GTTTTCGTGACAAATTAATTTCCAACGTTTTGTTTGATTAT CCGTTCGG (SEQ ID 402)	65

22	pG1-Δ6_bw	pG1-Δ6	CCGAACGGATAATCAAACAAAACGTTGGAAATTAATTTGT CACGAAAAC (SEQ ID 403)	68
23	pG1-Δ7_fw	pG1-Δ7	CCGGATAAGAGAATTTTGTTCGGATAAATGGACGCCTG (SEQ ID 404)	67
24	pG1-Δ7_bw	pG1-Δ7	CAGGCGTCCATTTATCCGAACAAAATTCTCTTATCCGGA CAAGACC (SEQ ID 405)	68
25	pG1-Δ8_fw	pG1-Δ8	GAATTTTGTTTGATTATCCGTTCGGCGCCTGCTCCATATT TTTCCG (SEQ ID 406)	70
26	pG1-Δ8_bw	pG1-Δ8	CGGAAAAATATGGAGCAGGCGCCGAACGGATAATCAAA CAAAATTC (SEQ ID 407)	67
27	pG1-Δ9_fw	pG1-Δ9	CGGATAAATGGACGCCTGCTCATTACCCCACCTGGAAGT GCC (SEQ ID 408)	68
28	pG1-Δ9_bw	pG1-Δ9	GGCACTTCCAGGTGGGGTAATGAGCAGGCGTCCATTTA TCCG (SEQ ID 409)	70
29	pG1-Δ10_fw	pG1-Δ10	GCCTGCTCCATATTTTTCCGGTTATCCCAGAATTTTCCG (SEQ ID 410)	53
30	pG1-Δ10_bw	pG1-Δ10	CGGAAAATTCTGGGATAACCGGAAAAATATGGAGCAGG C (SEQ ID 411)	69
31	pG1-Δ11_fw	pG1-Δ11	TATTACCCCACCTGGAAGTGCCCGGATAATACGGTGGTC TGGATTAAT (SEQ ID 412)	67
32	pG1-Δ11_bw	pG1-Δ11	ATTAATCCAGACCACCGTATTATCCGGGCACTTCCAGGT GGGGTAATA (SEQ ID 413)	68
33	pG1-Δ12_fw	pG1-Δ12	CCAGAATTTTCCGGGGATTATGGTCTGGATTAATTAATAC GCCAAGTC (SEQ ID 414)	68
34	pG1-Δ12_bw	pG1-Δ12	GACTTGGCGTATTAATTAATCCAGACCATAATCCCCGGA AAATTCTGG (SEQ ID 415)	65
35	pG1-ΔTAT14_fw	pG1-ΔT14	CAAAACTCTAATCTAAAACCTGAATCTCCGCGATGACCC CGTTTTCGTGAC (SEQ ID 416)	67
36	pG1-ΔTAT14_bw	pG1-ΔT14	GTCACGAAAACGGGGTCATCGCGGAGATTCAGGTTTTA GATTAGAGTTTTG (SEQ ID 417)	69
37	pG1-TAT18_fw	pG1-T18	CCTGAATCTCCGCTTTTTTTTTTTTTTTTTTTGATGACCCCG (SEQ ID 418)	70
38	pG1-TAT18_bw	pG1-T18	CGGGGTCATCAAAAAAAAAAAAAAAAAAGCGGAGATTCA GG (SEQ ID 419)	70
39	pG1-TAT20_fw	pG1-T20	CCTGAATCTCCGCTTTTTTTTTTTTTTTTTTTTGATGACCC CG (SEQ ID 420)	70
40	pG1-TAT20_bw	pG1-T20	CGGGGTCATCAAAAAAAAAAAAAAAAAAAAGCGGAGATT CAGG (SEQ ID 421)	70
41	pG1-TAT22_fw	pG1-T22	CCTGAATCTCCGCTTTTTTTTTTTTTTTTTTTTTTTGATGAC CCCG (SEQ ID 422)	70

42	pG1-TAT22_bw	pG1-T22	CGGGGTCATCAAAAAAAAAAAAAAAAAAAAAGCGGAGATTCAGG (SEQ ID 423)	70
43	pG1-d-472_fw	pG1-d1240/-d1427	GATACTGCAGCTCAGGGATTCCCACTATTTGGTATTC (SEQ ID 424)	68
44	pG1-d-188_bw	pG1-d1240	GATAGATCTCGTATTAATTAATCCAGACCACCG (SEQ ID 425)	64
45	pG1-d-1_bw	pG1-d1427	GATAGATCTAAGGGTGGAATTTTAAGGATCTTTTAT (SEQ ID 426)	64

Table 5: Fed batch cultivation of pG1 (herein referred to as pG1 #8) and pG1-x variants (herein also referred to as pG1-variants) expressing eGFP Relative eGFP fluorescence is shown for the batch end and for the fed batch end. The time points were set to 0 at the batch end. A clone expressing eGFP under control of pG1 (#8) was compared to clones expressing under control of a pG1 deletion (pG1-Δ2), a TAT14 mutation (pG1-T16), and a duplication (pG1-D1240) variant. The biomass concentrations (YDM) in the batch and fed batch were as expected.

	Batch End						**Fed Batch End**					
Clone	**t [h]**	**YDM [g/L]**	**relative eGFP fluorescence**			**%**	**t [h]**	**YDM [g/L]**	**relative eGFP fluorescence**			**%**
pG1 #8	-5.3	9.8	44	+/-	1	**100**	19.5	118.6	**2005**	+/-	36	**100**
pG1-Δ2 #3	-4.6	11.0	51	+/-	1	**116**	19.5	110.6	**1819**	+/-	43	**91**
pG1-T16 #3	-3.0	14.2	70	+/-	1	**160**	19.5	113.1	**2383**	+/-	24	**119**
pG1-D1240 #3	-3.0	14.9	62	+/-	1	**141**	19.5	113.3	**2948**	+/-	33	**147**

Table 6: Promoter strength compared to pG1 and promoter induction ratio of pG1 variants from a comparative deep-well screening. The expression strength of the pG1-x variants (induced) is related to the eGFP expression level obtained with the original pG1 promoter The induction ratio is calculated from the GFP level in the induced and repressed state.

	pG1 (P_{GTH1})	pG1-Δ8	pG1-Δ9	pG1-T16	pG1-T18	pG1-T20	pG1-D1240	pG1-D1427
Repression	6.1	5.8	9.4	5.4	6.7	5.3	5.3	5.5
Induction	15.3	11.0	21.4	17.0	20.8	16.2	21.6	22.9
Expression level	1.00	0.72	1.40	1.11	1.36	1.06	1.41	1.49
Induction ratio	2.52	1.89	2.27	3.12	3.10	3.03	4.05	4.18

CLAIMS

1. An isolated and/or artificial pG1-x promoter, which is a functional variant of the carbon source regulatable pG1 promoter of *Pichia pastoris* identified by SEQ ID 1, where the pG1-x promoter consists of or comprises at least a part of SEQ ID 1, the part of SEQ ID 1 having a length of at least 293 bp, the pG1-x promoter comprising the following promoter regions

a) at least two copies of a core regulatory promoter region; and

b) a non-core regulatory promoter region, which is any region within the pG1-x promoter sequence other than the core regulatory region;

wherein the core regulator promoter region has a sequence identity which is at least 80% to SEQ ID 2 and SEQ ID 3, the non-core regulatory promoter region has a sequence identity of at least 50% with a corresponding region in SEQ ID 1; and further

wherein the pG1-x promoter is characterized by an increased promoter strength and/or induction ratio as compared to the pG1 promoter, wherein

- the promoter strength is at least 1.1-fold increased in the induced state as compared to the pG1 promoter, and/or

- the induction ratio is at least 1.1-fold increased as compared to the pG1 promoter.

2. The pG1-x promoter of claim 1, wherein SEQ ID 2 and/or SEQ ID 3 comprises one or more transcription factor binding sites (TFBS).

3. The pG1-x promoter of claim 1 or 2, wherein the core regulatory region comprises the nucleotide sequence SEQ ID 4, or a functional variant of SEQ ID 4 with at least 80% sequence identity, which comprises one or more TFBS.

4. The pG1-x promoter of any of claims 1 to 3, wherein the core regulatory region is incorporated into a main regulatory region represented by SEQ ID 5, or a functional variant of SEQ ID 5 with at least 80% sequence identity, which comprises one or more TFBS.

5. The pG1-x promoter of any of claims 1 to 4, which comprises one or more TFBS for any of the transcription factors selected from the group consisting of Rgt1, Cat8-1 and Cat8-2.

6. The pG1-x promoter of claim 3, further comprising a deletion of one or more nucleotides between the nucleotide sequences SEQ ID 2 and SEQ ID 3.

7. The pG1-x promoter of any of claims 1 to 6, which comprises at least one or at least two T motifs identified by any of SEQ ID 12-29.

8. The isolated and/or artificial pG1-x promoter of any of claims 1 to 7,comprising a nucleotide sequence selected from a group consisting of any of

a) SEQ ID 37–44, preferably any of SEQ ID 45-76;

b) SEQ ID 77–80, preferably any of SEQ ID 81-112;

c) SEQ ID 113-114, preferably any of SEQ ID 115-130;

d) SEQ ID 131-132, preferably any of SEQ ID 133-148; or

e) SEQ ID 185-186, preferably any of SEQ ID 187-202;

or a nucleotide sequence which has at least 80% sequence identity to any of the foregoing.

9. The pG1-x promoter of claim 8, which comprises the nucleotide sequence selected from the group consisting of any of SEQ ID 37–44, preferably any of SEQ ID 45-76.

10. The pG1-x promoter of any of claims 1 to 9, which is operably linked to a nucleotide sequence encoding a protein of interest (POI), which pG1-x promoter is not natively associated with the nucleotide sequence encoding the POI.

11. The pG1-x promoter of claim 10, which further comprises a nucleotide sequence encoding a signal peptide enabling the secretion of the POI, preferably wherein nucleotide sequence encoding the signal peptide is located adjacent to the 5'-end of the nucleotide sequence encoding the POI.

12. An expression construct comprising the pG1-x promoter of claim 10 or 11, preferably an autonomously replicating vector or plasmid, or a vector or plasmid which integrates into the chromosomal DNA of a host cell.

13. A recombinant host cell which comprises the expression construct of claim 12, preferably a eukaryotic cell, more preferably a yeast or filamentous fungal cell, more preferably a yeast cell of the Saccharomyces or Pichia genus.

14. A method of producing a protein of interest (POI) by culturing a recombinant host cell line of claim 13, comprising the steps of

a) cultivating the cell line under conditions to express said POI, and

b) recovering the POI.

15. The method of claim 14, wherein the cultivation comprises

a) a first step using a basal carbon source repressing the pG1-x promoter, preferably wherein the basal carbon source is selected from the group consisting of glucose, glycerol, ethanol, a mixture thereof, and complex nutrient material, followed by

b) a second step using no or a limited amount of a supplemental carbon source de-repressing the pG1-x promoter to induce production of the POI, preferably wherein the supplemental carbon source is a hexose such as glucose, fructose, galactose or mannose, a disaccharide, such as saccharose, an alcohol, such as glycerol or ethanol, or a mixture of any of the foregoing.

16. The method of claim 15, wherein the second step b) employs a feed medium that provides for the supplemental carbon source in a growth limiting amount to keep the specific growth rate within the range of 0.001 h^{-1} to 0.2 h^{-1}, preferably 0.005 h^{-1} to 0.15 h^{-1}.

17. The method of claims 15 or 16, wherein the cultivation is performed in a bioreactor starting with a batch phase as the first step, followed by a fed-batch phase or a continuous cultivation phase as the second step, wherein the batch phase is performed for 20 to 36h; and the cultivation in the fed-batch phase is performed for any of 15 to 80h, 15 to 70h, 15 to 60h, 15 to 50h, 15 to 45h, 15 to 40h, around 15 to 35h, 15 to 30h, 15 to 35h, 15 to 25h, or 15 to 20h; preferably 20 to 40h.

18. The method of any of claims 14 to 17, wherein the recombinant host cell is yeast of any of the Saccharomyces genus or Pichia genus or *Komagataella* genus, preferably Pichia pastoris or Komagataella pastoris.

19. The method of any of claims 14 to 18, wherein the pG1-x promoter is any of SEQ ID 37–44, preferably any of SEQ ID 45-76, preferably wherein the pG1-x promoter comprises SEQ ID 39, more preferably SEQ ID 49.

20. The method of any of claims 14 to 19, wherein the POI is produced at a transcription rate of at least 15% as compared to the native pGAP promoter of the cell.

21. The method of any of claims 14 to 20, wherein the POI is a heterologous protein, preferably selected from therapeutic proteins, including antibodies or fragments thereof, enzymes and peptides, protein antibiotics, toxin fusion proteins, carbohydrate - protein conjugates, structural proteins, regulatory proteins, vaccines and vaccine like proteins or particles, process enzymes, growth factors, hormones and cytokines, or a metabolite of a POI.