Antonia Volpini de Maestri, BSc

Quantitative determination of transcript levels, regulated by *CTA*1 promoter variants

Diploma Thesis

in fulfillment of the requirements

for the academic degree of

Diplom-Ingenieurin

Master´s degree program - Biotechnology

submitted to

Graz University of Technology

Supervisors

Assoc. Prof. Dipl.-Ing. Dr.techn. Harald Pichler

&

Ao.Univ.-Prof. Mag.rer.nat. Dr.rer.nat. Anton Glieder

Institute of Molecular Biotechnology

Faculty of Technical Chemistry, Chemical and Process Engineering, Biotechnology

Graz, June 2019

Supervisors

Assoc.Prof. Dipl.-Ing. Dr.techn. Pichler Harald

&

Ao.Univ.-Prof.Mag.rer.nat. Dr.rer.nat. Anton Glieder

EIDESSTAATLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzen Quellen wörtlich und inhaltlich entnommen Stellen als solche kenntlich gemacht habe.

Ort, Datum

Unterschrift

Englische Fassung:

STATUROY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Locality, Date

Signiture

DANKSAGUNG

Da sich mein Studium langsam dem Ende nähert, möchte ich mich vorweg bei einigen Menschen bedanken, ohne deren Hilfe ich meine Ziele nie erreicht hätte.

Zu allererst möchte ich mich bei meinen Eltern bedanken, die mich nicht nur finanziell und ideell unterstütz haben, sondern mir auch stets mit guten Ratschlägen zur Seite gestanden sind. Ohne ihre großartige Hilfe hätte ich es nie geschafft, meine Träume zu verwirklichen.

Des Weiteren möchte ich mich bei meinem Professor Ao.Univ.-Prof. Mag.rer.nat. Dr.rer.nat. Anton Glieder bedanken, der mir nicht nur die Möglichkeit für diese Diplomarbeit gegeben hat, sondern auch zu meiner fachlichen Weiterentwicklung beigetragen hat. Danke Toni, dass du mir die Chance gegeben hast, über mich hinaus zu wachsen und dass du all die Zeit stets für Unterhaltungen offen bzw. fachliche Diskussionen bereit warst. Durch dein produktives Arbeitsklima habe ich sowohl Planung und Organisation, als auch selbständiges Arbeiten gelernt.

Außerdem möchte ich mich bei meinem Betreuer Assoc.Prof. Dipl.-Ing. Dr.techn. Harald Pichler bedanken, der mir stets mit seinem fachlichen Input weiterhelfen konnte. Danke Harald, dass du dir immer für mich die Zeit genommen hast und mich bei allen organisatorischen Punkten so gut unterstützt hast.

Ein großes Dankeschön gilt außerdem Anna-Maria Hatzl, MSc, die mich täglich mit ihrem fachlichen Wissen unterstützt hat und mir emotional bei allen Höhen und Tiefen zur Seite gestanden ist. Liebe Anna, danke für alles. Es war eine großartige Zeit.

Mein weiterer Danke gilt auch meinen Arbeitskollegen Katharina Ebner, BSc und Johannes Bitter, BSc. Danke euch beiden für die tolle Zusammenarbeit und den fachlichen Austausch. Danke Jo, dass du so ein toller Arbeitspartner am Beginn meiner Diplomarbeit warst und mir das Projekt von Grund auf erklärt hast. Danke Kathi, dass du mir stets fachlich zur Seite gestanden bist und auch beim Verfassen meiner Diplomarbeit immer gute Tipps parat hattest. Ohne eure Hilfe wäre das Projekt nicht so ein toller Erfolg geworden.

Ein großes Dankeschön gilt vor allem auch der ganzen Arbeitsgruppe 308, ehemals 208. Ich möchte mich nicht nur für jegliche Art der Hilfestellung und Unterstützung im Labor bedanken,

sondern auch für diese aufregende und lustige Zeit. Danke euch allen, dass ihr mich so warm und familiär in eure Gruppe aufgenommen habt.

Zu guter Letzt möchte ich meinem Freund und Lebenspartner Philipp Egger danken, der immer für mich da war und ohne dessen Hilfe ich das alles hier nie geschafft hätte, sowie all meinen Freunden und Studienkollegen für diese schöne Zeit.

ABSTRACT

The methylotrophic yeast *Komagataella phaffii* (*Pichia pastoris*) is a widely used host organism for the production of biopharmaceuticals and industrial enzymes. The simplicity of genetic manipulation and the capability to produce heterologous proteins at high titers by cultures with high cell densities, along with the ability to perform typical eukaryotic posttranslational modifications, make *P. pastoris* to an extremely popular and successful expression system. The versatile and efficient use of this yeast mainly arises from the strong and tightly controllable promoters originating from the MUT pathway, which can be induced with methanol. Nevertheless, methanol still represents a fire and health hazard, especially when it comes to large scale production for industrial use. Alternative efficient regulatory principles need to be found. Therefore, new processes and tools that allow high level expression without the necessity of induction with methanol are highly desirable. The promoter of the peroxisomal catalase gene, P_{CTA1}, offers such opportunities, but quantification of transcript levels, induced by de-repression or induction of this promoter, were so far unknown and therefore analyzed in this study.

Keywords:
Pichia pastoris, MUT pathway, promoter, catalase, quantification of transcript levels, methanol induction, de-repression

KURZFASSUNG

Die methylotrophe Hefe *Komagataella phaffii* (*Pichia pastoris*) ist ein weit verbreiteter Wirtsorganismus zur Herstellung von Biopharmazeutika und industriellen Enzymen. Die Einfachheit der genetischen Manipulation und die Fähigkeit, heterologe Proteine mit hohem Titer in Kulturen mit hohen Zelldichten zu produzieren, sowie die Fähigkeit, typische eukaryotische posttranslationale Modifikationen durchzuführen, machen *P. pastoris* zu einem äußerst populären und erfolgreichen Expressionssystem. Die vielseitige und effiziente Verwendung dieser Hefe beruht hauptsächlich auf den starken und streng kontrollierbaren Promotoren, die aus dem MUT-Weg stammen und mit Methanol induziert werden können. Trotz alledem, stellt Methanol eine Brand- und Gesundheitsgefahr dar, insbesondere wenn es sich um eine großtechnische Produktion für den industriellen Einsatz handelt. Alternative effiziente Regulierungsprinzipien müssen gefunden werden. Daher sind neue Verfahren und Werkzeuge, die eine Expression auf hohem Niveau ermöglichen, ohne dass eine Induktion mit Methanol erforderlich ist, äußerst wünschenswert. Der Promotor des peroxisomalen Katalase-Gens, P_{CTA1}, bietet solche Möglichkeiten, aber die Quantifizierung der Transkriptmengen, die durch Derepression oder Induktion dieses Promotors induziert werden, waren bisher unbekannt und wurden deshalb in dieser Studie untersucht.

Schlagwörter:
Pichia pastoris, MUT Stoffwechselweg, Promotor, Katalase, Quantifizierung von Transkriptlevels, Methanolinduktion, De-repression

TABLE OF CONTENTS

ABBREVIATIONS

AOX1	Alcohol oxidase gene
bp	base pairs
CalB	*Candida antarctica* lipase B
cDNA	complementary DNA
CTA1	Catalase gene
D_{600}	(optical) density of cells measured at 600 nm
ddH_2O	Double distilled water (“Fresenius” water)
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EtBr	Ethidium bromide
EtOH	Ethanol
FAD	Flavin adenine dinucleotide
gDNA	genomic DNA
GFP	Green fluorescent protein
GOI	Gene of interest
h	hours
HKG	Housekeeping gene
M	Molar mass (g/mol)
MeOH	Methanol
min	minute(s)

miRNA	micro RNA
mRNA	messenger RNA
MUT	Methanol utilization pathway
ncRNA	non-coding RNA
ONC	Overnight culture
ori	origin of replication
PCR	Polymerase chain reaction
PPB	Potassium phosphate buffer
qPCR	quantitative real-time PCR
RNA	Ribonucleic acid
RNase	Ribonuclease
rRNA	ribosomal RNA
RT	Room temperature
s	second(s)
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
ssDNA	single-stranded DNA
tRNA	transfer RNA
WT	Wild type
YNB	Yeast nitrogen base

1 INTRODUCTION

1.1 Synthetic biology and the versatile use of yeasts

Synthetic biology mainly focuses on the generation of more practical applications by assembling several biological components in new ways to create and develop novel systems. Nowadays, the ability to control the flow of genetic information, to interfere with gene regulation and optimize proteins and posttranslational modifications, is of huge interest. Protein level and subsequent levels of metabolites and other cellular properties underlie gene regulation and thus, improvements depend on these complex mechanisms. As cellular information flow starts with transcription, scientists are highly interested in evolving and enhancing those regulatory tools in order to control it. However, synthetic biology faces many challenges to find and design appropriate genetic sequences that can be used to control protein expression and subsequently, optimize cellular functions (Engstrom & Pfleger, 2017). Transcriptional regulation is a key method to investigate promoters and their consequences for protein expression. Understanding of these transcriptional regulatory networks could be used to design various production strains for specific needs and thereby, presents interesting basic approaches for synthetic biology and metabolic engineering (Vogl & Glieder, 2013).

In modern biotechnology, yeasts are highly desirable organisms due to their remarkable potential to serve as expression systems. They combine the ease of genetic manipulation with the capability to perform posttranslational modifications, making them attractive production hosts for various applications. Everything started with *Saccharomyces cerevisiae* (*S. cerevisiae*), which has been used for more than thousands of years for brewing beer and baking bread. For decades, information about this yeast has been collected making *S. cerevisiae* the best characterized eukaryotic system today (Gellissen & Hollenberg, 1997). As a result of the tremendous possibilities, many recombinant protein production systems are built on *S. cerevisiae,* although this yeast species has its limitations. However, there are alternative yeasts exhibiting more beneficial properties and are successfully used to produce live-saving biopharmaceutical drugs and large quantities of recombinant proteins. The broad range of available alternative yeasts for protein production includes the popular group of methylotrophic yeasts like *Pichia pastoris* (*P. pastoris*), *Hansenula polymorpha* (*H. polymorpha*), *Candida boidinii* (*C. boidinii*) and *Blastobotrys* (*Arxula*) *adeninivorans* (*B.* (*A.*) *adeninivorans*) (G. P. L. Cereghino & Cregg, 1999; Gellissen & Hollenberg, 1997; Malak, Baronian, & Kunze, 2016).

As eukaryotic production system, yeasts are capable of performing many eukaryotic-specific posttranslational modifications like folding, proteolytic processing, N- and O-glycosylation or disulfide bond formation, which makes them suitable for industrial use and research (G. P. L. Cereghino & Cregg, 1999). These traits have made yeasts essential organisms used to reproduce and reengineer complex molecular properties of *P. pastoris* and eukaryotes in general (G. P. L. Cereghino & Cregg, 1999).

1.2 Expression host *Pichia pastoris*

Over the past decades, *Komagataella phaffii*, also commonly known under its former name *Pichia pastoris*, has become a widely reported and extensively used expression system for the production of various biopharmaceuticals and industrial enzymes (Ahmad, Hirz, Pichler, & Schwab, 2014; Kurtzman, 2009). As expression host, this yeast offers many advantages over other organisms as it can, for example, produce milligram to multi-gram quantities of foreign proteins, both intracellular and extracellular, and can easily be scaled up for bioreactor application to not only reach higher protein titer, but also to better control influencing parameters like the oxygen supply, carbon source feed or pH (Macauley-Patrick, Fazenda, McNeil, & Harvey, 2005). Compared to *S. cerevisiae*, *P. pastoris* belongs to the group of Crabtree negative yeasts, meaning that ethanol production only occurs at low levels in the presence of oxygen. This characteristic facilitates growth to high cell densities and high protein product yields (Hartner & Glieder, 2006). Other higher eukaryotic hosts like mammalian cells, require complex cultivation conditions and growth media to produce foreign proteins, whereas *P. pastoris* combines the simplicity of molecular genetic manipulation with easy and inexpensive cultivation media (J. L. Cereghino & Cregg, 2000). Another huge benefit is its capability to perform higher eukaryotic protein processing and other posttranslational modifications like phosphorylation, glycosylation and disulfide bridge formation (Macauley-Patrick et al., 2005). As protein secretion in *P. pastoris* only occurs for a minority of the endogenous proteins, the purification process of heterologous produced proteins gets simplified by directing them outside the cell, which also offers a great way to avoid toxicity from intracellularly accumulated cell material. (J. L. Cereghino & Cregg, 2000).

Considering all these facts, the yeast provides other advantageous properties for producing recombinant proteins. It has a tightly regulated transcription system allowing the separation of cell growth and protein production phase. This system is based on the strong and tightly

regulated promoters of the methanol utilization (MUT) pathway; without their induction biomass accumulates but heterologous gene expression is completely repressed (J. L. Cereghino & Cregg, 2000). Based on these facts, recombinant protein production is one of the major applications of *P. pastoris* and various promoter elements derived from genes involved in the MUT pathway are excellent regulators for this task since they are tightly regulated and can be induced with methanol (Hartner & Glieder, 2006).

Overall, the *P. pastoris* expression system has gained high importance for the production of foreign proteins. Although the system already offers great features, there is still potential to further develop and improve certain marker/host strain combinations or to resolve problems associated with protein secretion and glycosylation. Furthermore, it is crucial to find and develop new and enhanced regulatory sequences (promoters) that could lead to an increased transcriptional activity and thereby, to higher product yields while maintaining good protein stability (J. L. Cereghino & Cregg, 2000).

1.3 Methanol Utilization Pathway

Methylotrophic yeasts like *P. pastoris* can use methanol as their sole carbon and energy source and belong to the different genera of *Candida, Pichia, Komagataella, Torulopsis* and *Ogataea (*formerly *Hansenula)* (Kurtzman, 2009). They all share a specific methanol utilization pathway, also called MUT pathway, which involves various unique enzymes and promoters (Macauley-Patrick et al., 2005). The enzymes alcohol oxidase (AOX) and dihydroxyacetone synthase (DAS), are essential in the methanol metabolism and are only produced at high levels when the cells are grown on methanol and the absence of repressible carbon sources such as e.g., glucose, glycerol, or ethanol (J. L. Cereghino & Cregg, 2000).

In specialized cell organelles, the peroxisomes, the initial reactions of the MUT pathway are compartmentalized (Macauley-Patrick et al., 2005; Rußmayer et al., 2015). By shifting the cells to methanol as sole carbon source, the peroxisomes are strongly induced and start to produce and sequester certain enzymes. After induction with methanol, these specialized microbodies significantly increase and can occupy up to 80% of the cytoplasmic space. The first step of methanol utilization is the oxidation of methanol to formaldehyde and hydrogen peroxide catalyzed by AOX. As hydrogen peroxide is toxic, it has to be neutralized to water and molecular oxygen which is done by catalase (CAA) (Hartner & Glieder, 2006). More recent annotation, based on three letter codes used for *S. cerevisiae* described this same gene as *CTA*1 gene (Valli et al., 2016). Both enzymes are sequestered within the peroxisomes (J. L. Cereghino

& Cregg, 2000). Not only CAT and AOX are released there, also the key enzyme DAS is sequestered in the peroxisomes, which makes them indispensable to the MUT pathway (Macauley-Patrick et al., 2005). Even though formaldehyde is important for growth, it can be fatal for the cells at higher concentrations (Koutz et al., 1989). Due to this fact, the remaining formaldehyde is either assimilated in the cell metabolism through the condensation with xylulose 5-phosphate, Xu_5P, catalyzed by DAS (J. L. Cereghino & Cregg, 2000) or oxidized by two dehydrogenase reactions in the dissimilation pathway (Hartner & Glieder, 2006). The products of the DAS catalyzed reaction, dihydroxyacetone and glyceraldehyde 3-phosphate, are released from the peroxisome to enter the cytoplasmic pathway. Xu_5P is regenerated in this pathway together with one net molecule of glyceraldehyde 3-phosphate for every three cycles (J. L. Cereghino & Cregg, 2000). *Figure 1.1* describes the most important reaction steps of the methanol utilization pathway in the yeast *P. pastoris*.

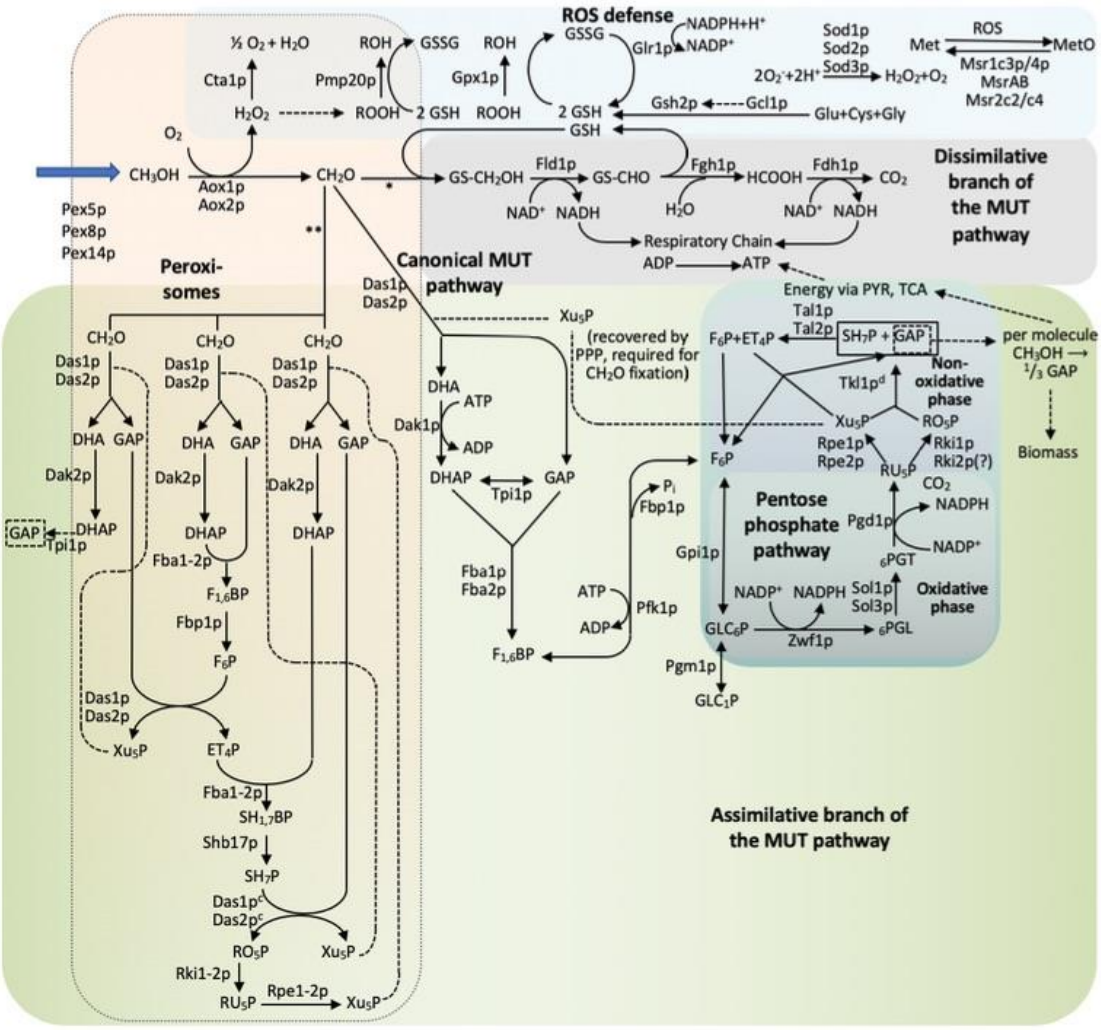

Figure 1.1 Methanol utilization pathway in *P. pastoris* (based on the drawings of Vogl, et al. (2016) and Rußmayer, et al (2015), provided by Johannes Bitter).

The main pathways involved in the methanol metabolism are visualized. Abbreviations of the main enzymes and metabolites involved in the methanol metabolism are listed: AOX, alcohol oxidase; CAT, catalase; FLD: formaldehyde dehydrogenase; FGH, formyl glutathione hydrolase; FDH, formate dehydrogenase; DAS, dihydroxyacetone synthase; DHA, dihydroxyacetone; DAK, dihydroxyacetone kinase; DHAP, dihydroxyacetone phosphate; TPI, triosephosphate isomerase; FBA, fructose 1,6-bisphosphate aldolase; GAP, glyceraldehyde-3-phosphate; $F_{1,6}BP$, fructose-1,6-bisphosphate; FBP, fructose-1,6-biphosphatase; Pi, inorganic phosphate; F_6P: fructose-6-phosphate; GPI, glucose-6-phosphate isomerase; GLC_6P, glucose-6-phosphate; PGM, phosphoglucomutase; GLC_1P, glucose-1-phosphate; Xu_5P, xylose-5-phosphate; SHB17, seduheptulose-1,7-bisphosphatase; GSH, glutathione; GSSG, oxidized glutathione self-dimer; MetO, methionine sulfoxide; $SH_{1,7}BP$, seduheptulose-1,7-bisphospate; PYR, pyruvate; TCA, tricarboxylic acid cycle; SH_7P, sedoheptulose-7-phosphate; RKI, ribose-5-phosphate keto-isomerase; RO_5P, ribose-5-phosphate; RU_5P, ribulose-5-phosphate; RPE, ribulose-phosphate 3-epimerase; ET_4P, erythrose-4-phsohpate; $_6PGL$, 6-phosphogluconolactone; $_6PGT$, 6-phosphogluconate. *, The reaction of CH_2O is proceeded nonenzymatically. Initial reactions of methanol assimilation, localized in the peroxisomes, are based on an alcohol oxidase (AOX), which converts methanol to formaldehyde and hydrogen peroxide, as well as a special transketolase, dihydroxyacetone synthase (DAS), which forms a C-C bond with the C1 molecule formaldehyde. The toxic hydrogen peroxide is neutralized through the conversion by a catalase (CAT) (Rußmayer et al., 2015; Vogl et al., 2016).

1.3.1 State-of-the-art expression system employing P_{AOX1} or P_{GAP}

The reaction pattern to oxidize methanol to formaldehyde distinguishes methylotrophic yeasts from bacterial organisms. In bacteria, this reaction is catalyzed by the action of a dehydrogenase linked to the electron transport chain, whereas yeasts and fungi possess an flavin-dependent oxidase, using oxygen as electron acceptor. The so-called alcohol oxidase not only metabolizes methanol, but also oxidizes other lower primary aliphatic alcohols. This flavoprotein acts as an octamer consisting of identical subunits, each containing a noncovalently bound flavin adenine dinucleotide cofactor. To compensate the low oxygen affinity of AOX, the cells produce large amounts of the enzyme (Cregg, Madden, Barringer, Thill, & Stillman, 1989).

The *P. pastoris* genome contains two different AOX genes, *AOX1* and *AOX2*, both encoding the respective enzymes with alcohol oxidase activity. *AOX1* and *AOX2* are homologous with more than 90% identity for protein-coding regions at both, DNA and predicted amino acid sequence level (H. Zhang et al., 2009). Outside the protein-coding sequences, no homologies were found (Cregg et al., 1989). *AOX1* is under control of the *AOX1* promoter, P_{AOX1}, and the expressed alcohol oxidase 1 can account up to 30% of the total soluble protein in methanol-grown cells (Krainer et al., 2012). It is not detectable in cells grown in presence of glucose (Cregg et al., 1989). Expression of *AOX1* is tightly regulated at the level of transcription, as for cells only grown on methanol around 5% of poly(A)$^+$ RNA is from *AOX1* (J. L. Cereghino & Cregg, 2000). Compared with the exceeding strength of P_{AOX1}, the second alcohol oxidase *AOX2* underlies the control of a much weaker promoter, P_{AOX2}, and hence only comprises up to 15% of the total cellular alcohol oxidase activity (Krainer et al., 2012). Overall, *AOX1* expression leads to the majority of AOX activity in methanol-grown cells, due to the differences in the respective promoter sequences located 5´ of the protein-coding regions. For both AOX genes, regulation of expression in response to environmental conditions is similar but the quantity of mRNA differs distinctly. This probably reflects the consequence of different transcription initiation rates of *AOX1* and *AOX2*, as fusion studies revealed identical proteins and nearly identical mRNAs (except for 5´ non-translated regions) for both genes (Cregg et al., 1989). Until now, it is unclear why a second *AOX* gene exists in *P. pastoris*, although there seems to be no physiological reason. Even though the circumstances are not yet understood, there might be a reason why *P. pastoris* has conserved the sequences required for the high specificity of *AOX2* (Cregg et al., 1989). In other yeasts, the different roles and kinetics were characterized more systematically (Nakagawa et al., 2002, 2001, 2005).

Currently, there are three different *P. pastoris* phenotypes available: 1. Mut^S (methanol utilization slow), where *AOX1* gene is knocked out and thus, growth on methanol is slow; 2. Mut^+ (methanol utilization plus), where both *AOX* genes are active and intact and 3. Mut^- (methanol utilization minus), where both *AOX* genes are deleted and growth on methanol as sole carbon source is not possible (Krainer et al., 2012). So far, Mut^+ and Mut^S are the most frequently used phenotypes (Macauley-Patrick et al., 2005). For these two phenotypes, expression of the genes involved in MUT pathway is repressed when glucose or glycerol is used as a sole carbon and energy source. Protein expression can be induced by the addition of methanol without one of these carbon sources present, or when either of the two AOX promoters are employed (Krainer et al., 2012). Generally, expression of the MUT pathway and the respective enzymes derived from this metabolism, is repressed by glucose, glycerol and ethanol and strongly induced by methanol (Hartner & Glieder, 2006; Vogl et al., 2016). Upon de-repression state, the mRNA level of *AOX1* accounts for around 1-2% of the induced one (Hartner & Glieder, 2006).

Typical methanol induced processes, employing P_{AOX1} or P_{DAS2}, build on a two-step cultivation process. Biomass can be produced by growth to high cell densities under promoter repressing conditions, whereas protein production only gets activated by the induction with methanol. Cell growth and protein production are two separated processes and thereby, potentially detrimental impacts caused by the produced recombinant protein on the cells are reduced. For favorable, non-toxic proteins, separation of growth and production phase is not necessary, thus constitutive expression system can be used in order to minimize cultivation efforts and duration. For example, the glyceraldehyde 3-phosphate dehydrogenase gene, *GAP*, is under the control of the GAP promoter (P_{GAP}), which facilitates constitutive expression and has a similar strength as the P_{AOX1} . Although P_{GAP} seems to be constitutively expressed, studies showed that certain carbon sources such as glycerol, oleic acid or methanol, affect the expression strength (Liang, Zou, Lin, Zhang, & Ye, 2013; Vogl & Glieder, 2013). The results proved that the mRNA levels of the *GAP* gene were highest in glucose-grown cells and decreased in the presence of the carbon sources named above. Furthermore, studies described the impact of oxygen on P_{GAP}, as under hypoxic conditions the specific productivity of certain recombinant produced proteins increased three- to six-fold. Additionally, these studies revealed that under hypoxic conditions, *P. pastoris* started to produce ethanol resulting in a change to the oxido-fermentative metabolism. However, studies revealed a promoter with significantly stronger activities than the classic constitutive promoters P_{GAP} and P_{TEF1} under certain circumstances. Based on the results, this so called P_{GCW14} promoter seems to offer promising potential as an additional

constitutive promoter for the expression of heterologous genes in *P. pastoris* (Liang et al., 2013). Anyway, the fact that constitutive expression can have negative (depending on the heterologous protein even cytotoxic) side effects and lower final product titers are the reason why such promoters have not been as widely used as the P_{AOX1} (Macauley-Patrick et al., 2005; Vogl & Glieder, 2013).

1.3.2 Catalase and the use of its promoter P_{CTA} for gene expression

In general, catalases are important enzymes, especially for alcohol oxidase mediated methanol oxidation, as they break down hydrogen peroxide into water and oxygen (Hartner & Glieder, 2006). As H_2O_2 is toxic and a byproduct of many metabolic reactions such as the oxidation of fatty acids in the peroxisomes, the detoxification function of catalases is highly important for the cells in general (Rußmayer et al., 2015). Furthermore, the physiological role of catalases is crucial as the lack of catalase activity may cause genetic diseases as well as neurological disorders (Horiguchi et al., 2001). Based on these facts, the function of catalases in the peroxisomal oxidative metabolism is of fundamental importance (Nakagawa et al., 2010).

Yeasts autonomously produce catalases when cells are grown under cultivation conditions that require the detoxification of harmful metabolites like hydrogen peroxide. Unusual carbon and nitrogen sources such as fatty acids, alcohols, D-amino acids or primary amines lead to their specific induction (Rußmayer et al., 2015). The exact localization of these enzymes has been discussed for several years. However, their initial localization in the peroxisomes was proven by means of different biochemical and cytochemical studies. The yeast *S. cerevisiae* is known to have additional catalase activity in the mitochondria and thus, the peroxisomal protein (catalase A) can be targeted to both, peroxisomes and the mitochondria. Nevertheless, catalase targeting to mitochondria seems to be a frequent capacity in both, fermentative (*Saccharomyces*) and respiratory (*Pichia*, *Hansenula*) yeasts (Koleva, Petrova, Hristozova, & Kujumdzieva, 2008). Furthermore, a second catalase gene in baker´s yeast, the *CTT*1 gene, is responsible for the expression of additional cytosolic catalase (Subramaniyan, Alugoju, SJ, Veerabhadrappa, & Dyavaiah, 2019).

Although catalases are usually induced during β-oxidation of fatty acids, in methylotrophic yeasts like *P. pastoris*, *C. boidinii* or *H. polymorpha*, synthesis of catalases is also strongly induced through the presence of methanol, as its metabolism causes the production of hydrogen peroxide. As this toxic byproduct has detrimental impacts on the cells, catalases are produced upon glucose depletion, where lipid bodies and other storage compounds get metabolized. This

situation also leads to a preparation for alternative carbon sources such as methanol, leading to the same toxic side product H_2O_2. Thereby, their production precedes the expression of *AOX* and other proteins involved in the MUT pathway. Studies showed that the expression of catalases, as most genes which are responsible for carbon source utilization, is primarily regulated at transcriptional level. However, there also might be an alternative regulation at posttranslational level such as the transport into the peroxisomes (Horiguchi et al., 2001; Sakai, Yurimoto, Matsuo, & Kato, 1998). Overall, peroxisomal proteins are synthesized on free polyribosomes and targeted into peroxisomes after translation is completed. Transport of the finished proteins is mediated by *cis*-acting peroxisomal targeting signals (PTSs) together with the respective receptors (Horiguchi et al., 2001). In this diploma thesis, catalase promoter variants were tested and analyzed using the methylotrophic yeast *P. pastoris*.

Previous studies showed, that expression of the promoter of the single catalase gene of *P. pastoris* P_{CTA1} is already induced upon carbon source depletion. Therefore, this promoter sequence presents a feasible starting point for the engineering of sequence variants, with the goal to enable strong methanol-free expression in *P. pastoris*. In case of success, the implementation of these sequences could eventually represent a valuable alternative to the P_{AOX1} and so minimize the necessity of methanol. In this diploma thesis, *CTA1* promoter variants, generated by other colleagues in the lab, were tested under varying cultivation conditions. In order to understand their mode of regulation, samples have been taken during the entire cultivation processes and mRNA levels of the transcripts of the target gene have been analyzed via quantitative real-time PCR (qPCR).

1.4 Transcriptional regulation in eukaryotes

Methylotrophic yeasts such as *P. pastoris*, *H. polymorpha*, *P. methanolica* and *C. boidinii* are widely used expression systems when it comes to heterologous protein production due to many different advantages, as already mentioned before. The production of a specific protein requires expression of the respective target gene which includes transcription, translation, protein folding, certain posttranslational modifications, if necessary, and finally, correct intra- or extracellular targeting by the host organism. In both, prokaryotic and eukaryotic organisms, initial transcription of the gene of interest (GOI) is often problematic and presents challenges for which solutions must be found. As a consequence, strong and regulatable promoters are essential for an efficient protein production (Vogl & Glieder, 2013). Inside cells, the genetic information determines the development of complex organisms. This information includes all protein-coding sequences of genes as well as non-coding regulatory elements that control

where, when and to what extend genes will be expressed (Haberle & Stark, 2018). Starting with transcription, the DNA sequence gets copied into the appropriate RNA transcript catalyzed by an enzyme called RNA polymerase II (Pol II) (Lubliner et al., 2015). This enzyme transcribes all protein-coding and several non-coding sequences (Haberle & Stark, 2018). Typically, transcription process starts at a defined position known as the transcription start site (TSS) at the 5´end of a gene. The TSS is arranged within the core promoter, a short sequence comprising ~50 bp up- and ~50 bp downstream of the transcription start site, to which the transcription machinery (e.g. Poly II and associated transcription factors (TFs)) binds. Core promoters mediate transcription initiation from precise positions at defined levels, but generally possess low fundamental activity and thus, can be further activated or suppressed by enhancers or chromatin modification, respectively. Basically, the term promoter encompasses the core promoter and an upstream proximal promoter, which together independently drive transcription. Enhancers are regulatory elements that bind TFs, regulatory proteins, in the presence of transcriptional cofactors leading to an increased transcription process (Haberle & Stark, 2018). For *P. pastoris*, the activator Mxr1 is constitutively expressed among carbon source regulation, thereby playing a key role in the de-repression and activation of P_{AOX1}. On the other hand, Nrg1 repressor competes for Mxr1 binding elements, thus representing a crucial part in the inhibition process of P_{AOX1} (Liang et al., 2013) .

In general, core promoters are significantly involved in the assembly of the pre-initiation complex (PIC), which consists of Pol II and various TFs (Haberle & Stark, 2018). Generally, there are several different core promoter motifs with defined positions relative to a single TSS, as for example the commonly known TATA-box motif, located ~30 bp upstream of a single dominant TSS, to which the PIC is recruited (Haberle & Stark, 2018; Lubliner et al., 2015). In lower eukaryotes like yeasts, the TATA-box is found to be the only conserved motif in the core promoter (Portela, Vogl, Ebner, Oliveira, & Glieder, 2018). However, for eukaryotes, it has to be distinguished between promoters containing a TATA-box and TATA-less promoters. It has been widely reported that in yeasts, TATA-containing genes indicate high plasticity (higher or lower expression levels) and show induction under stress compared to genes lacking a TATA-box. In *S. cerevisiae*, approximately 17% of the total promoters constitute of a TATA-containing core promoter (Yella & Bansal, 2017). The TATA-box binding protein (TBP) recognizes the TATA-box and binds there. TBP is part of the transcription factor IID (TFIID) complex, a TF which conveys Pol II recruitment and PIC assembly and due to that, might define a certain TSS at a fixed downstream position. An alternative core promoter motif represents the initiator motif (Inr). Inr directly overlaps the TSS and appears more frequently than the TATA-

box. In contrast with the TATA-box consensus sequence TATAAA, which is conserved from yeast to human, Inr motif differs between species. Promoters without a TATA-box often consist of the Inr motif and an additional one, the downstream promoter element (DPE) located downstream of the TSS. Overall, the TATA-box, Inr and DPE are the most abundant core promoter motifs (Haberle & Stark, 2018). In higher eukaryotes, synthetic core promoters have been designed based on frequently occurring motifs such as the TATA box, Inr (initiator), DPE (downstream core promoter element) and MTE (motif ten element). In lower eukaryotes like yeasts, the only conserved motif in the core promoter seems to be the TATA box (Portela et al., 2018). *Figure 1.2* shows the schematic representation of a eukaryotic promoter containing a TATA-box and 5´upstream-specific configuration of *cis*-motif.

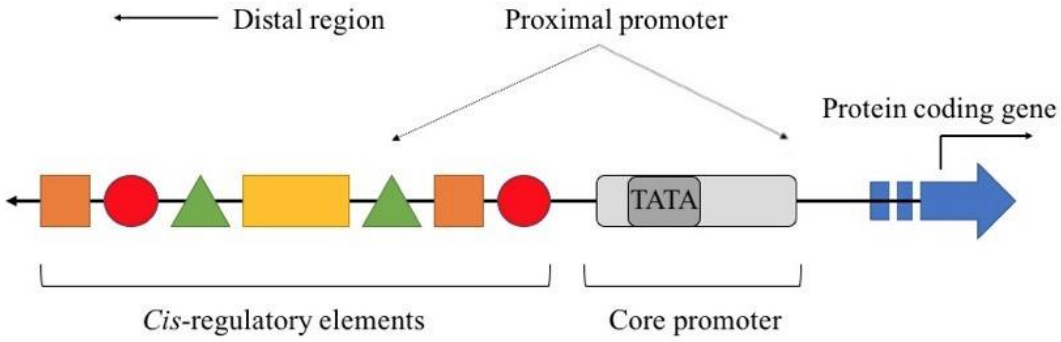

Figure 1.2 Schematic representation of a eukaryotic promoter containing a TATA-box and 5´upstream-specific configuration of cis-motif.
A typical eukaryotic promoter comprises of two distinct regions: a core promoter element and upstream enhancer elements, the *cis*-regulatory elements. The direction of transcription as well as TSS are defined by the core element, while the *cis*-regulatory elements help determine promoter strength or transcriptional frequency. The core element itself mainly functions as transcription initiator at defined TSS to promote basal transcription and typically composed of less than 80 nucleotides (Blazeck, Alper, & Nagamune, 2013).

Promoters are DNA sequences that regulate gene transcription by providing specific DNA-binding sites to which transcription factors and the respective cellular transcription machinery can bind, thus enabling transcription initiation. The interactions of promoter and TFs favor the recruitment of the transcription machinery and thereby, leading to transcription of an open reading frame (ORF). Prokaryotic transcription is initiated by the σ-factor of RNA polymerase, which is responsible for the promoter recognition. In contrast, eukaryotic transcription initiation is way more complex as the process requires binding of DNA sequence-specific TFs within the promoter element. Additionally, interactions with transcriptional coactivators are essential in order to localize the cellular transcription machinery. In order to make the assembly of Pol II into PIC at the core promoter region possible, up to 30 protein-based elements encompassing the five general TFs are necessary. This procedure initiates transcription as the promoter DNA sequence is unwound around the PIC. Thereafter, it comes to a conformational change of PIC into an open complex allows for scanning for a proper TSS and thus, initiating an active elongation (Blazeck et al., 2013).

Generally, eukaryotic TFs are *trans*-acting elements that bind to *cis*-regulatory elements, 5´upstream-specific configurations, acting as either activators or repressors (Vogl & Glieder, 2013). The *cis*-regulatory elements localize the *trans*-acting elements to the core promoter. Within the enhancer region, transcriptional activators or repressors affect the transcriptional process through binding on specific transcription factor binding sites (TFBSs). However, so called upstream activation sequences lead to an increased transcription rate, while upstream repressive sequences diminish transcription frequency. Overall, the mode of promoter regulation and rate of transcription depends on TF-mediated interactions in the 5´upstream enhancer element (Blazeck et al., 2013). Promoter engineering can help finding optimal transcriptional regulation in order to optimize heterologous protein production. Thereby, exchanging different natural promoters with modified and controllable ones, may enable fine-tuned gene expression for synthetic biology applications such as metabolic engineering and lead to an overall understanding of their mode of regulation (Blazeck et al., 2013; Vogl & Glieder, 2013).

1.5 Transcriptome analysis

Over the last decade, there have been enormous innovations and developments regarding genome and transcriptome analyses. In the past, gene expression studies were based on cross-

species hybridization on microarrays or limited to small-scale quantitative PCR analyses. Next-generation sequencing and the enhancements of various analytical tools made whole-genome or whole-transcriptome investigations possible, even for non-model organisms (Wolf, 2013). Today, more and more scientists focus on analyses of the transcriptome rather than on the genome, as only 1-2% of the genes are coding and 80-90% of the transcribed genes are not translated into proteins. They are known to function in the regulation of gene expression and epigenetic regulation mechanisms, as these complex processes require numerous regulating steps starting with gene transcription and transcript processing, translation into protein and posttranslational modifications. A single gene can be transcribed into several products based on the presence of many different transcription start sites (TSSs). Furthermore, alternative splicing and polyadenylation of the primary RNA can lead to the production of various transcript forms based on the same gene. Each mature transcript codes for a different protein due to multiple different TSSs. Eventually, non-translated RNA molecules, such as tRNAs, rRNAs, snRNAs, etc., have structurally and catalytically important functions in the translation processes. Understanding the regulation patterns of gene expression and transcript processing as well as identification of transcription factor binding sites provides detailed information about the transcriptome and how it functions (Anamika, Verma, Jere, & Desai, 2016).

The term transcriptomics encompasses various techniques used to study an organism´s transcriptome, the sum of all RNA molecules in one cell or a population of cells. The total amount of different RNA molecules produced is also referred as transcript (Lowe, Shirley, Bleackley, Dolan, & Shafee, 2017). The transcriptome comprises of a variety of different RNA molecules, like mRNA, miRNA, ncRNA, rRNA or tRNA. Depending on the type of RNA molecule, each of them has a diverse function in the physiological response. To get a deeper understanding of the functional genome, it is crucial to investigate in the regulation of these molecules and try to understand how they interact with each other (Anamika et al., 2016).

In each organism, the DNA sequences, long-term storage of information, are expressed through transcription. The transient produced transcript, the mRNA, plays an essential role as molecule in information network, while noncoding RNAs execute other different important functions. Overall, the transcriptome provides a temporal profile of the total transcripts present in a cell under certain growth and cultivation conditions (Lowe et al., 2017). Various technologies like sequencing-based methods or hybridizations are used to achieve transcriptome profiling. Hybridization methods such as microarrays, based on binding of fluorescently labeled fragments to complementary probes either on a solid surface or in solution, are limited by low

sensitivity and specificity as well as low resolution. Later on, newly developed approaches based on Sanger-sequencing became available, but also suffer from limitations. With the development of powerful sequencing technology, called Next Generation Sequencing (NGS), used to sequence millions of nucleotide fragments in parallel, RNA Sequencing (RNA-Seq) has become one of the methods of choice to study the transcriptome. RNA-Seq provides many advantages over hybridization techniques as it offers a genome-wide coverage of transcripts while maintaining high sensitivity and specificity; no prior knowledge is needed to investigate targets; detection of low-abundance transcripts is possible as well as detection of novel transcripts (Anamika et al., 2016).

Furthermore, quantitative transcript analysis can be performed by using qPCR, a technique which collects data throughout the entire PCR process. This method combines target amplification after reverse transcription and detection in one single step by the use of different fluorescent chemistries. Overall, qPCR was the method of choice in this diploma thesis as this technique provides many advantages such as the production of quantitative data in a broad dynamic range, high sensitivity, the least biased results and the ability to even detect single copies of specific transcripts – all within relatively short analysis times (Wong & Medrano, 2005).

1.5.1 Quantitative real-time PCR

Quantitative real-time PCR (qPCR) is considered to be one of the most precise and reliable techniques when it comes to the study of genetic regulation patterns and the understanding of their mode of expression. Over the last years, qPCR has become a widely used method for quantifying target nucleic acids because of its many different advantages, such as speed of analysis and real-time detection of reaction process (Kozera & Rapacz, 2013). This quantification technique offers tremendous sensitivity and high sequence-specificity over a broad, dynamic range with low post-amplification processing (Wong & Medrano, 2005). It provides enormous potential for both, quantitative and analytical approaches and allows the direct detection of qPCR products during the exponential growth phase and thereby, enabling detection and amplification in one single step (Giulietti et al., 2001). However, a comprehensive understanding of the principles the method is based on is of prime importance. Although this technique is a reliable quantification method, qPCR suffers from problems associated with amplification of unspecific products or efficiencies, the formation of primer-dimers or hetero-duplex formation. Accordingly, , one has to consider several individual steps before starting an experiment in order to obtain accurate and significant quantification results. As RNA samples

get quantified, it is crucial to ensure their high purity and integrity and thus, only utilize those templates that proved to be of highest quality, DNA-free and undegraded for qPCR experiments, according "*A-Z of quantitative PCR, Chapter 3: Quantification strategies in real-time PCR*" by S.A. Bustin.

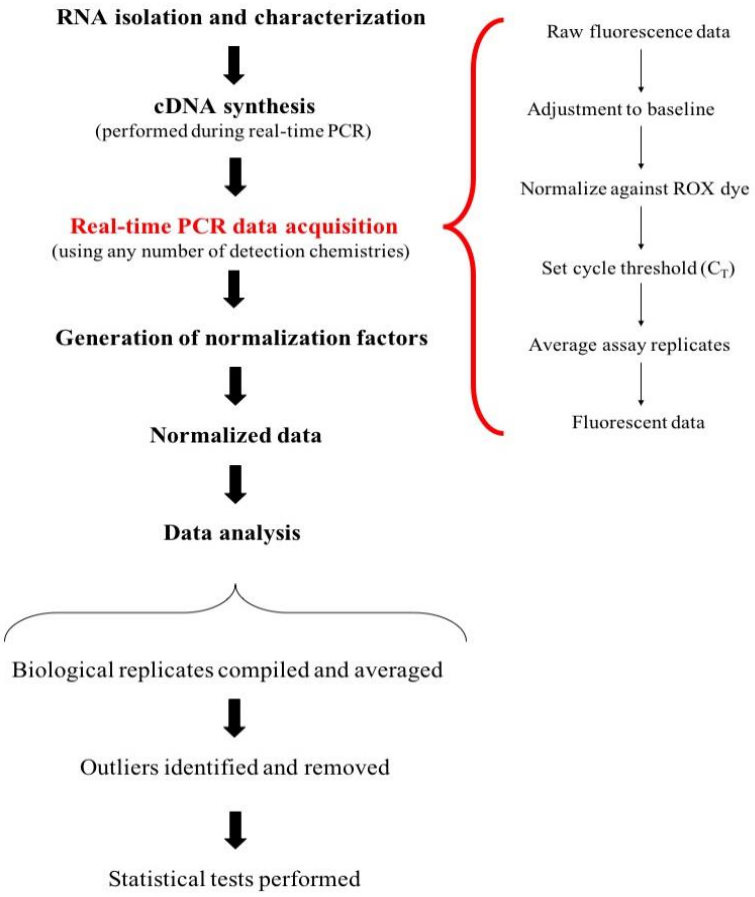

Figure 1.3 Schematic representation of a typical workflow for determination of gene expression through a real-time PCR experiment.
Samples from the desired target strains must be taken in order to isolate the appropriate RNAs. After isolation, the RNA samples must be characterized for integrity and quality. In a one-step qPCR analysis method, the target RNAs can be directly used as templates for the reaction and reverse transcribed into cDNA during the assay. For a two-step reaction, cDNA gets first synthesized, which is then used as qPCR template. The steps highlighted through the red bracket are performed during the qPCR process on the real-time PCR machine. For normalization, proper housekeeping genes (HKGs) must be chosen and a normalization factor must be calculated for each individual sample (Wong & Medrano, 2005).

Generally, the amount of target gene transcripts can be analyzed in either an absolute or relative fashion. By using the absolute quantification method, the template copy number is determined based on a standard curve prepared with dilution series of known concentrations (Kozera & Rapacz, 2013). As scientists are often more interested in the expression change of a specific

target than knowing the actual copy number of a gene, relative quantification is commonly used to determine gene expression analysis. However, in a relative quantification experiment, the produced qPCR data of the specific target gene are normalized on endogenous control genes as internal reference (Gong et al., 2016). Over the years, several different mathematical models have been established in order to calculate the relative changes in gene expression of a specific target with respect to an internal reference gene. The most popular approach in qPCR is called $\Delta\Delta C_T$ method, which requires optimal and identical PCR efficiencies of target and endogenous control gene (Teste, Duquenne, François, & Parrou, 2009). The results are expressed as fold-change of the expression levels; like for example one wants to compare changes in the expression of a certain gene over a given time period in treated and untreated samples. For this hypothetical case, an untreated calibrator (reference) sample and an endogenous control gene would be necessary to normalize input amounts. Via real-time PCR, levels of both target and endogenous control genes of all samples would be determined. The data output, levels of target normalized to levels of endogenous control, would be expressed as differences in the fold-change according the "*Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR*" (Applied Biosystems, Inc., Foster City, CA, Unites States).

In a qPCR experiment, the logarithmic amplification of genetic material proceeds in three repetitive reactions in their respective varying temperature programs, namely matrix denaturation, primer hybridization and elongation. Theoretically, the starting input material gets duplicated twice in each cycle assuming that 100% PCR efficiency can be reached. However, the basal reaction processes are preceded by reverse transcription of the target RNAs into their complementary single-stranded DNA copies (cDNA) catalyzed through the enzyme reverse transcriptase (RT). The primary phases of qPCR proceed quite slowly as a result of the relatively small initial target quantity. The faster the process continues into the exponential phase, the more template can be amplified and fluorescence emission begins to exceed the background level. This cycle threshold (C_T) represents the start of the logarithmic phase according "*A-Z of quantitative PCR, Chapter 3: Quantification strategies in real-time PCR*" by S.A. Bustin (Kozera & Rapacz, 2013). In a standard end-point PCR quantitation, it may happen that the reaction is no longer producing templates at an exponential rate caused by inhibitors of the polymerase reaction, reagent limitation or accumulation of certain pyrophosphate molecules, and thereby leading to uneven reaction products. However, an end-point quantitation may result in unreliable amplification products compared to real-time PCR as this technique measures its products as they occur in the exponential range of the reaction. Only if

amplification is measured in this range, it is possible to determine the initial amount of template gene (Ginzinger, 2002). For *P. pastoris*, qPCR was first described by Abad, et al. (2010).

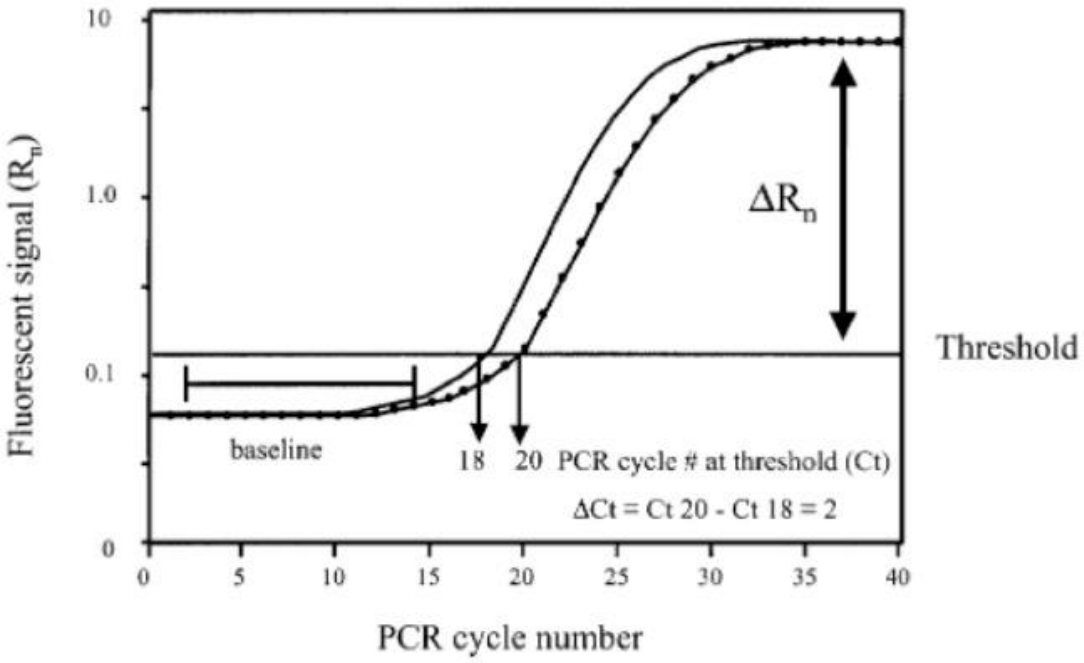

Figure 1.4 Theoretical overview of a typical amplification plot used in qPCR experiments (Ginzinger, 2002).
During real-time PCR assays, a cycle threshold (C_T) is defined as the number of cycles required for the fluorescence signal to exceed the background level. This point is used to compare all samples and calculated as a function of the amount of background signal. The calculated C_T values directly represent the quantity of starting template used to calculate gene expression levels. As can be seen, the PCR cycle number is plotted against the fluorescence signal detected during the quantification process. The baseline represents the PCR cycles in which fluorescence signal is accumulating but not yet detectable. The C_T value of each sample is defined as the number of PCR cycles that were required to detect a fluorescent signal above the threshold. In a relative quantification method, these values are then used to calculate the relative abundance of the template.

1.5.1.1 Detection chemistries

Over the years, different types of detection chemistries have been developed such as SYBR Green I, Molecular Beacons, Scorpions and *Taq*Man probes. These detection methods originate from the principle of fluorescence resonance energy transfer, or shortly FRET (Giulietti et al., 2001). FRET describes an energy transfer between a donor and an acceptor and depends on their close proximity that energy can be transferred from the excited fluorophore (donor) to the other (acceptor) resulting in detectable fluorescence emission (Sekar & Periasamy, 2003). By

using one of these methods, the increase of fluorescence emission can be detected during the entire quantification reaction in "real-time" and directly corresponds to the amount of target amplification. The qPCR instrument continuously collects and sends data to a computer software program that calculates ΔRn using the following equation:

$$\Delta Rn = Rn^{+} - Rn^{-}$$

Rn^{-} represents the fluorescence emission of the baseline, where Rn^{+} stands for the fluorescence emission of the product at each time point. As the software program receives data, amplification plots are generated based on the fluorescence data collected during the process. The calculated ΔRn values are plotted against the cycle numbers, although primary values at the initial stages of the process do not cross the baseline. Considering the variability of the baseline, an arbitrary threshold is chosen to be used for the calculations of C_T values by determining the point at which the fluorescence signal passes the threshold. As C_T indicates the cycle number at a certain point, it decreases with increasing input amount and can be used for a quantitative measurement of the target input (Giulietti et al., 2001). Some of the most commonly used detection chemistries are described below.

- **Hydrolysis or *Taq*Man Probes:**

  With this type of detection chemistry, a probe in combination with forward and reverse primer is used. These three oligonucleotides are all specific for the target sample and able to bind there. Based on the probe technology, this assay method utilizes the specific function of the *Taq* polymerase, an enzyme with 5´→ 3´-nuclease activity. The efficiency of the hydrolysis reaction mainly depends on the 5´→ 3´-activity of the *Taq* polymerase. The oligonucleotide probe used in this method has dyes attached at each end – at the 5´end a fluorescent reporter dye and at the 3´end a quencher dye. Hydrolysis of the probe by the polymerases 5´→ 3´-exonuclease activity causes the separation of reporter and quencher leading to a fluorescence signal which can be detected. This signal corresponds to the quantity of target input as an increase in the fluorescence emission only occurs when the probe is annealed to the target. *Taq*Man probes are commonly applied for diagnostic analysis like virus quantification, gene detection and cytokine quantification. (Giulietti et al., 2001).

- **SYBR Green I:**

  In contrast to the other detection methods mentioned, the principle of SYBR Green I detection relies on the ability to bind double-stranded DNA molecules and thereby,

leading to target fluorescence signal. In its unbound form, the dye has an undetectable fluorescence, but as soon as SYBR Green I is intercalated into the target molecule, the dye emits a fluorescence signal which can be detected. Compared with other types of detection chemistries, SYBR Green I is applicable with all kind of primers for any target, therefore being less expensive than probes. However, as the dye binds to all double-stranded molecules including primer dimers and nonspecific PCR products, specificity of the amplification process is decreased. By using SYBR Green I intercalating dyes, real template products and artificial ones cannot be differentiated in contrast to Molecular Beacons or *Taq*Man probes. To overcome this problem, melting curve analyses are added to evaluate amplification specificity. By establishing melting curves, the software program is able to adjust the fluorescence signal above the primer dimers´ melting temperature but below that of the product (Ginzinger, 2002; Giulietti et al., 2001).

- **Molecular Beacons:**

  Molecular Beacons are probes composed of single-stranded DNA molecules formed in stem-and-loop structures. This method is based on the binding of a specific probe to the complementary target nucleic acid. Each DNA molecule is labelled with a fluorophore at one end and a quencher at the other end. As long as the hairpin-structure persists, the fluorescence signal is quenched based on the close proximity of the fluorophore and the quencher. Once probe and target hybridize with each other, the hairpin-structure loses its conformation resulting in a linear annealed molecule and thereby, increasing the fluorescence emission as fluorophore and quencher are separated. This type of chemistry is often used for quantification of pathogens, mutation detection, virus replication and gender detection in embryos (Giulietti et al., 2001; Tyagi & Kramer, 1995).

- **Scorpions:**

  Based on the combination of an acceptor fluorophore and a hybridization probe (donor fluorophore), scorpions consist of a specific probe designed to stick in a hairpin-loop configuration by complementary stem sequences on the 5´ and 3´ end of the probe. As long as no hybridization reaction between the target nucleic acid and acceptor and donor occurs, fluorescence emission is blocked. The formation of a head-to-tail conformation based on annealing between target and probes leads to a fluorescence signal as acceptor

and donor are in close proximity. Compared to other type of chemistries such as *Taq*Man probes or Molecular Beacons, Scorpions have a faster reaction time resulting in much stronger fluorescence signal and thereby, allowing highly specific PCR amplification. However, this detection form represents a relatively new fluorescence chemistry (Giulietti et al., 2001; Thelwell, Brown, Millington, Solinas, Booth, & Brown, 2000).

1.5.1.2 Housekeeping genes

Gene expression analysis can be performed via qPCR in either an absolute or relative way. Although these techniques allow rapid, sensitive and highly specific quantification of mRNA targets, errors can occur due to varying input amounts of the starting material between samples. Accordingly, to this, it is especially important to correct these type of errors, otherwise they may influence the accuracy of the experiment and lead to misinterpretation of the quantification results. To do so, internal reference genes, also called housekeeping genes (HKGs), are used to normalize the target RNA inputs. In terms of qPCR, internal reference genes are also known as endogenous controls. Before starting a qPCR experiment, the selection of an adequate internal reference gene is of prime importance in order to obtain reliable results. To determine gene expression of a specific target, mRNA samples under various treatments such as cultivation conditions, stress and growth phase, get quantified via qPCR. To normalize target input amounts, the ideal endogenous control gene should be unaffected by growth conditions and experimental treatment and expressed at constant levels. The expression of HKGs in the cell is anticipated to be stable, thus they are commonly used for normalization (Løvdal & Lillo, 2009). Although researches empirically assume that the expression of HKGs is not regulated and independent among experimental conditions, studies indicate that these genes seem to be regulated at least to some extent. This suggests that there probably is not a universal reference gene which is expressed at constant levels, regardless of the external conditions (Teste et al., 2009). However, expression levels of the internal control gene and specific target should be comparable, otherwise the calculations of the qPCR data are not meaningful (Gong et al., 2016).

Generally, it is crucial to consider the fact that not every endogenous control is suitable for normalizing gene expression and therefore, the selection must correlate with the respective organism (Løvdal & Lillo, 2009). In addition, studies proved that expression of HKGs often varies considerably among different experimental treatments and thus, identification of proper internal reference genes and validation of their stability under different conditions is a prerequisite for qPCR experiments (Gong et al., 2016). Nevertheless, the identification and

validation of more than one HKG is often not performed due to time reasons, so accurate normalization and correct quantification cannot be fully guaranteed (Gong et al., 2016).

In this diploma thesis, two different HKGs, *ACT1* and *ARG4*, were tested and validated for normalization of gene expression. As described by Rebnegger et al. (2014), *ACT1* has been successfully used as a HKG to normalize expression levels for *P pastoris* cultures grown on glucose-limited chemostat cultures (Rebnegger et al., 2014). Furthermore, the HKG *ARG4* has been reported by Abad, et al. (2010) to be a stable and reliable endogenous control for *P. pastoris* expression strains using DNA as starting material (Abad et al., 2010; Krainer et al., 2012). Many studies have performed qPCR to determine expression levels of certain target genes and report the use of various different HKGs. In spite of everything, it is recommended to test at least two different reference genes as the use of only one may lead to errors. The accuracy of a qPCR experiment strongly depends on several different factors. To ensure the reliability of the amplification results, it is crucial to choose a HKG that shows a similar threshold cycle as the gene of interest and is able to compensate for any variation of genetic material to the same extent (Kozera & Rapacz, 2013; Sinha et al., 2015).

2 OBJECTIVES

Since the necessity of methanol induction represents a hazard, alternative expression mechanisms need to be found. Based on this fact, we intended to find a promoter with a tight regulatory mechanism that does not require methanol as inducer but still achieves high product yields. As catalase expression precedes the expression of alcohol oxidase upon carbon source depletion, the peroxisomal promoter derived from this gene offers tremendous possibilities based on its de-repressed regulatory mechanisms, thus represents a potential alternative to methanol required induction. Therefore, previous master students have generated more than a hundred of different promoter variants originating from a 500 bp fragment of the parental P_{CTA1}, which showed to be fully functional and sufficient for *CTA*1 promoter activity. For deeper investigations, selected P_{CTA1} variants in comparison to the parental P_{CTA1} and the benchmark promoter P_{DF} were characterized on a transcriptional level. To get a better understanding of their mode of regulation, the expression levels of the target protein were determined at different stages of cultivation by the use of qPCR. Moreover, the target *P. pastoris* strains were cultivated with two different carbon sources, namely glucose and glycerol, to examine a possible influence of the respective carbon source and cultivation strategy on promoter regulation and recombinant protein production. Previous unpublished studies in our lab indicated, that glucose might cause stronger repression of the *CTA*1 promoter than glycerol. The major goal of this study was to get access to quantitative transcript data in order to identify possible correlations between transcript data and protein product analysis. Such data can be used to design improved future cultivation procedures, while reducing the risk of failure in scale up experiments.

3 MATERIAL AND METHODS

3.1 Devices and equipment

All devices and equipment used in the course of this thesis are listed in *table 3.1*. Pipette tips for normal use, Eppendorf tubes, Greiner tubes or other smaller vessels are not included there.

Table 3.1 All devices and equipment used during this diploma thesis are listed below.

Instrument	Company
Centrifuge	
Eppendorf Centrifuge 5415R	Eppendorf AG, Hamburg, Germany
Eppendorf Centrifuge 5810R	Eppendorf AG, Hamburg, Germany
Pipettes	
Denville Xl 3000i single-channel pipette 0.1-2 µL	Denville Scientific Inc., Metuchen, NJ, United States
Denville Xl 3000i single-channel pipette 2-20 µL	Denville Scientific Inc., Metuchen, NJ, United States
Denville Xl 3000i single-channel pipette 20-200 µL	Denville Scientific Inc., Metuchen, NJ, United States
Denville Xl 3000i single-channel pipette 100-1000 µL	Denville Scientific Inc., Metuchen, NJ, United States
Biohit Proline® multi-channel electronic pipettor 5-100 µL	Biohit Oyj, Helsinki, Finnland
Biohit Proline® multi-channel electronic pipettor 50-1200 µL	Biohit Oyj, Helsinki, Finnland
Filter tips	
Greiner Bio-One™ Filter Tips for Eppendorf FT 1000 (E)	Greiner Bio-One GmbH, Frickenhausen, Germany
Greiner Bio-One™ Filter Tips for Eppendorf FT 100 (E)	Greiner Bio-One GmbH, Frickenhausen, Germany
Greiner Bio-One™ Filter Tips for Eppendorf FT 20 (E)	Greiner Bio-One GmbH, Frickenhausen, Germany
Greiner Bio-One™ Filter Tips for Eppendorf FT 10 (E)	Greiner Bio-One GmbH, Frickenhausen, Germany
Shaker	
HT Infors Multitron Shaker	Infors AG, Bottmingen, Switzerland
HT Infors Orbitron shaker	Infors AG, Bottmingen, Switzerland
HT Infors RS306 shaker	Infors AG, Bottmingen, Switzerland
Plate Reader	

SynergyMx Plate Reader	Biotek Inc., Winooski, United States
Microtiter Plates	
96 well PS Microplater sterile	Greiner Bio-One GmbH, Frickenhausen, Germany
96 well PS Microplater unsterile	Greiner Bio-One GmbH, Frickenhausen, Germany
MicroAmp® Optical 96-Well Reaction Plate	Applied Biosystems, Inc., Foster City, CA, Unites States
Membranes	
Applied Biosystems™ MicroAmp™ Optical Adhesive Film	Applied Biosystems, Inc., Foster City, CA, Unites States
NuPAGE™ 4-12% Bis-Tris Protein Gels, 1.0 mm, 15-well	Invitrogen by Thermo Fisher Scientific, Waltham, MA
PCR cycler	
Applied Biosystems™ 7500 Real-Time PCR System	Applied Biosystems, Inc., Foster City, CA, Unites States
Applied Biosystems 2720 Thermal Cycler	Applied Biosystems, Inc., Foster City, CA, Unites States
Other Devices	
Eppendorf Thermomixer Comfort	Eppendorf AG, Hamburg, Germany
Certoclav LVEL 12L	CertoClav GmbH, Traun, Austria
Vortex Genie 2	Scientific Industries Inc., Bohemia, NY, United States
Hamilton® Polyplast lab pH electrode	Sigma-Aldrich GmbH, Vienna, Austria
Eppendorf BioPhotometer plus	Eppendorf AG, Hamburg, Germany
MT PG12001-S DeltaRange Balance Mettler	Toledo Inc., Greifensee, Switzerland
Heidolph MR 2002 Mangentic Stirrer	Heidolph Instruments, Schwabach, Germany
inoLab pH 720 pH-Meter	WTW GmbH, Weilheim, Germany
Thermo Scientific Nanodrop 2000c	Thermo Fisher Scientific Inc., Waltham, MA, United States
Whatman® quantitative filter paper, ashless, Grade 43	Sigma-Aldrich GmbH, Vienna, Austria
Whatman® Puradisc 13 syringe filters	Sigma-Aldrich GmbH, Vienna, Austria
PreSens SFR System	PreSens - Precision Sensing GmbH, Regensburg, Germany
Nalgene® baffled shake flask 250 mL	Sigma-Aldrich GmbH, Vienna, Austria

3.2 Strains, plasmids and variants description

3.2.1 Strains

The target plasmids used in this study were designed and constructed by Katharina Ebner and Johannes Bitter based on a preliminary study of Mohamed Hussein. All experiments were executed using the expression strain *K. phaffii* BSY11d*KU70*. The promoter variants used were triple mutants originating from the $P_{CTA1/500}$ promoter, the 500 bp sequence upstream of the *Komagataella phaffii* catalase gene, as established by Vogl, et al. (2016) and in the following called *PpCTA*1 promoter (or in short *CTA*1 promoter P_{CTA1}). These variants differed from the parental promoter through three 10-bp long sequence exchanges, A and B: A sequence – ATCCTTTTAG and B sequence – GATAACCGTG. In a first series, single exchanges were performed, followed by combinatorial designs and mutagenesis. In preliminary screenings, variants containing the mutations 14B and 26B (sequence exchange 130 bp-140 bp and 250 bp-260 bp from the 5´end of the promoter, respectively) showed to be the most promising candidates in the de-repressed state, which is why they were used for further mutation combinations.

The plasmids encoding the target protein CalB were set under the control of novel synthetic P_{CTA1} promoter variants and designed based on the vector pPp_*AOX1*-pUC-ZeoR-NotI-TT_*AOX1* (pKEB#1). The vector backbone contained an up- and downstream *AOX*1 homologous region for double-crossover insertion at the *AOX*1 locus resulting in a Mut^S *Pichia pastoris* strain. Moreover, the plasmids encoded a pUC ori element for autonomous replication in *E. coli*, a Zeocin resistance cassette for antibiotic selection and a NotI restriction site for linearization. Selected variants (P_{CTA1}1A+14B+26B, P_{CTA1}14B+26B+48B) and the control promoters P_{CTA1}-WT and P_{DF} were amplified to attach an overhang to the CalB gene associated α-signal sequence resulting in pPp_*AOX1*-pUC-ZeoR-SpeI-P_{CTA1}_variant-α_CalB-TT_*AOX1* plasmids. P_{DF} is a promoter derived from the upstream region of *Ogataea polymorpha* (*H. polymorpha*) FMD gene, and showed to be an exceptional strong de-repressed and methanol inducible promoter for protein expression by *P. pastoris*. *Table 3.2* shows all expression constructs used in the course of this project.

Table 3.2 List of all expression constructs used during this diploma thesis.
Each construct encoded CalB as target protein. In variant 19, the first 10 bp (position 1-10) from the 5´end of the promoter were exchanged to sequence A, as well as the bases at position 131-140 and 251-260 to sequence B. In variant 26, the bases from position 131-140, 251-260 and 471-480 from the 5´end of the promoter were exchanged to sequence B. The parental P_{CTA1} promoter was used for comparison of expression data, NC as control (lacking of any promoter) and P_{DF} served as benchmark representing a strong de-repressed promoter.

Acronym of the variant	Expression construct
19	pKEB#1-PCTA500-1A+14B+26B-αMF-CalB
26	pKEB#1-PCTA500-14B+26B+48B-αMF-CalB
NC	pKEB#1-αMF-CalB
WT	pKEB#1-PCTA500_WT-αMF-CalB
P_{DF}	pKEB#1-P_{DF}-αMF-CalB
pKEB#1_CalB_23a	Positive control

3.2.2 Plasmids

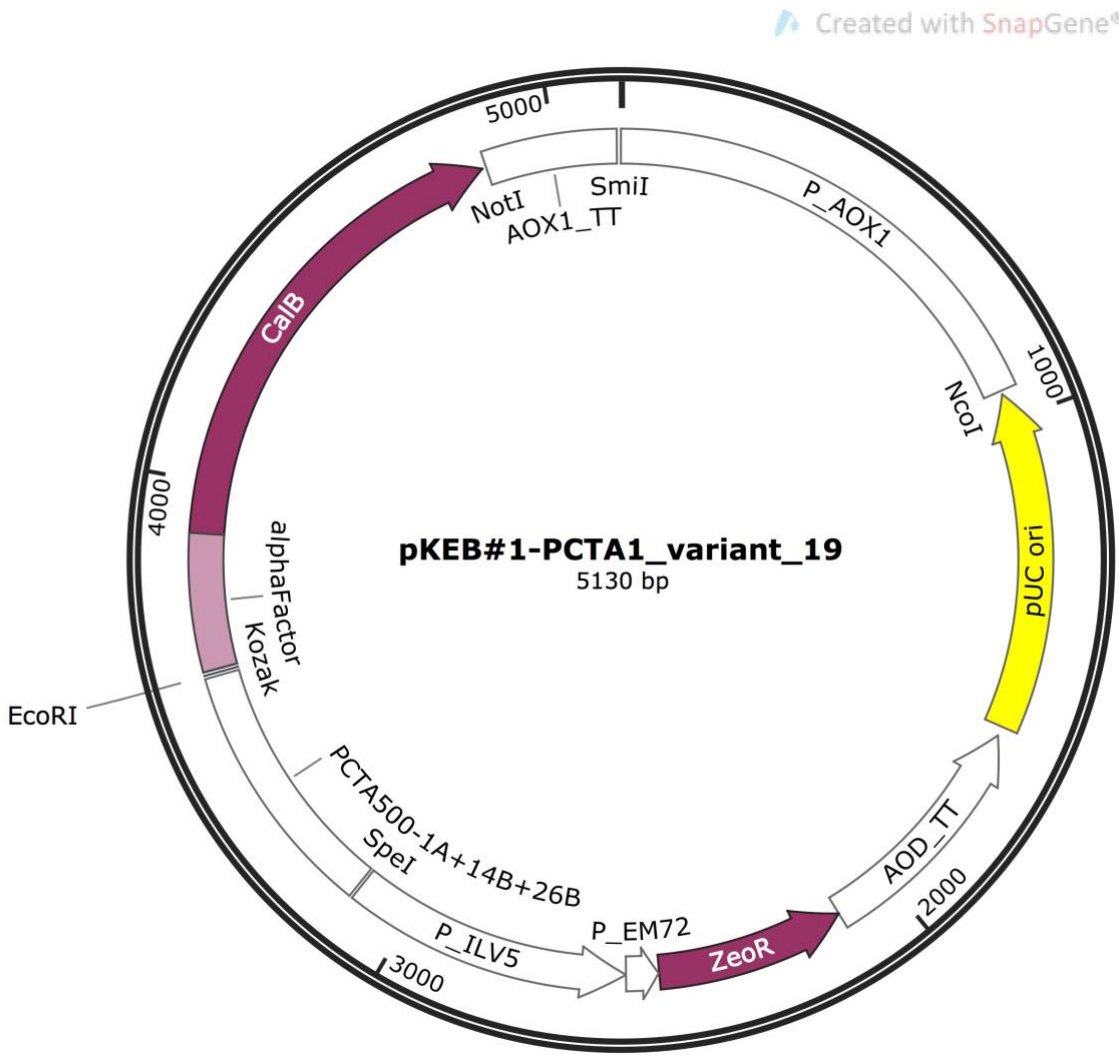

Figure 3.1 Expression construct of P_{CTA1} promoter variant 19.
The plasmid construct encodes the target gene CalB under the control of the corresponding synthetic P_{CTA1} promoter. P_{AOX1}: 5´*AOX*1 promoter; *AOX*1_TT: *AOX*1 transcription termination region; pUC ori: pUC replication origin for *E. coli*; ZeoR: Zeocin resistance gene; alpha factor: α-signal sequence for protein secretion.

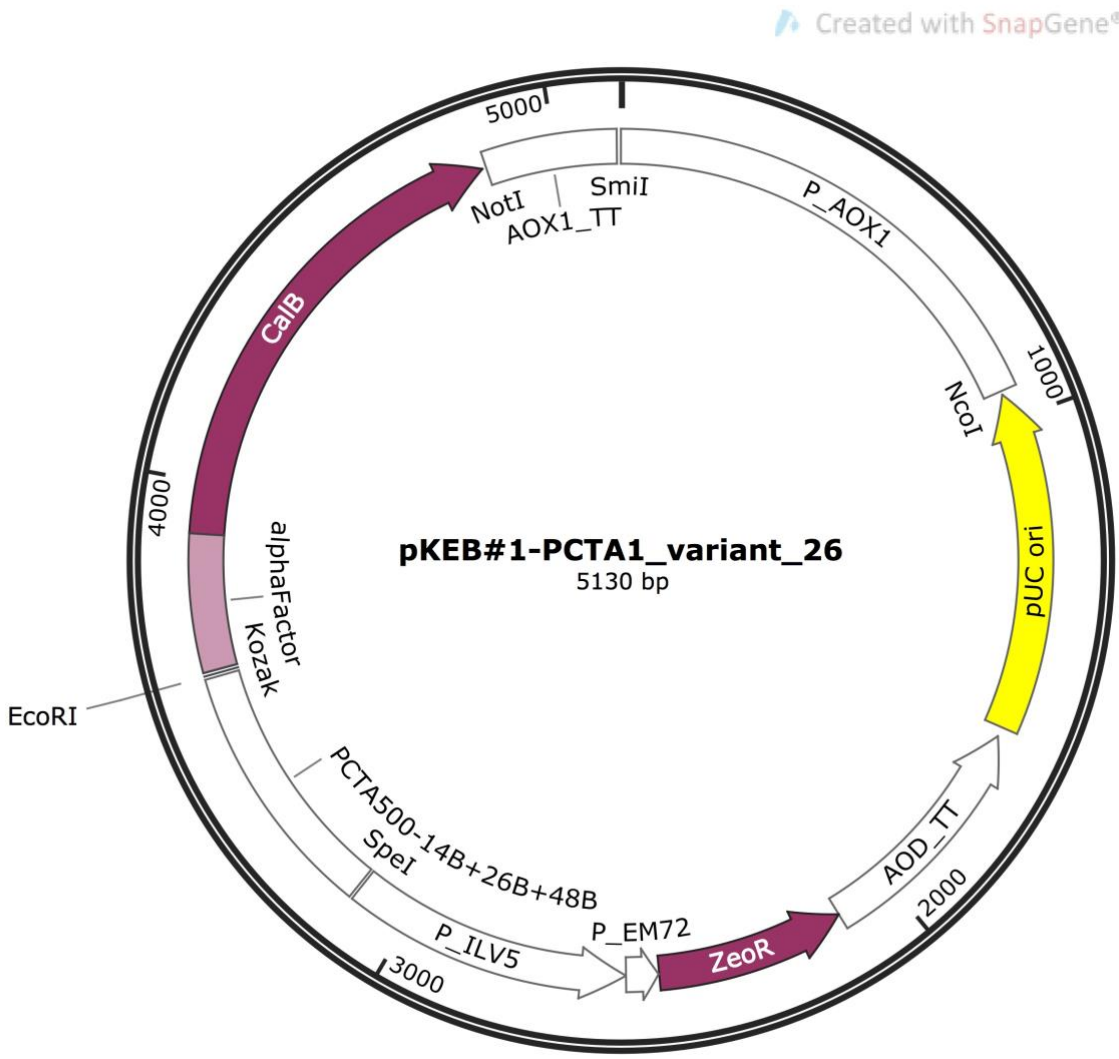

Figure 3.2 Expression construct of P_{CTA1} promoter variant 26.
The plasmid construct encodes the target gene CalB under the control of the corresponding synthetic P_{CTA1} promoter. P_{AOX1}: 5´*AOX*1 promoter; *AOX*1_TT: *AOX*1 transcription termination region; pUC ori: pUC replication origin for *E. coli*; ZeoR: Zeocin resistance gene; alpha factor: α-signal sequence for protein secretion.

3.3 Primers and enzymes

All primers and enzymes used in this study are mentioned in *table 3.3* and *3.4*. Enzymes contained in kits are not listed there.

Table 3.3 List of all primers used in the course of this project.

Primer name	Sequence 5´→ 3´	Internal number
AOD_TT_rev	ttctgcagctaaggtaatcagatccaagtttcc	P16246
ZeoR_fwd	atggctaaactcacctctgctgttccagtc	P14896
CalB_1fw	tgatggcatttgctcctgacta	P17459
CalB_1rev	gtgtcaatccaaccagcgtttc	P17460
ACT1_qPCR_1_fw	cgttttgtccctgtacgcttc	-
ACT1_qPCR_1_rev	aatctctaccggccaagtcg	-
ARG4_qPCR_fw_3	tggctggtttcctcatgtctatt	-
ARG4_qPCR_rev_3	ggtagaaactacaccggatgct	-
ARG4_qPCR_fw_4	ttgatgccgaacgaatgaagaatg	-
ARG4_qPCR_rev_4	ttcaactcctcggcttgtctg	-
ACT1_1fw	ggtttctccttaccacacgctatt	-
ACT1_1rev	ctccttgatgtcacggacgattt	-
ACT1_2fw	tgctcttgactttgaccaggaa	-
ACT1_2rev	ccaagtacagatgggtggaaca	-
ACT1_4fw	tatgccggtttctccttaccac	-
ACT1_4rev	gatgtcacggacgatttctctctc	-

Table 3.4 All enzymes which were used during this project and the corresponding suppliers are stated below.

Enzymes	Type of enzyme	Company
Q5® High-Fidelity DNA Polymerase	Polymerase	New England Biolabs, Ipswich, MA

3.4 Media, buffer and chemicals

The different media compositions used in this study are listed in *table 3.5*, buffer and stock solution preparations are mentioned below. In *table 3.6*, all used chemicals for this diploma thesis and their corresponding suppliers are described.

3.4.1 Media

Table 3.5 All media prepared in the course of this diploma thesis are described below.
For each medium, the amount of ingredients corresponds to the preparation of one liter. All ingredients were dissolved in deionized water. Stock solutions were sterile filtered, buffers and carbon sources dissolved in ddH_2O were autoclaved. Media composing of different ingredients were autoclaved prior to adding the stock solutions and/or buffers. In case of an antibiotic selection, Zeocin was added at a final concentration of 100 µg/mL.

Medium	Ingredients	Amount per liter
BMD0.5%	Glucose monohydrate	5.5 g
	10X YNB stock	100 mL
	1M PPi buffer pH 6	200 mL
	500X Biotin stock	2 mL
BMG0.5%	50% Glycerol stock	10 mL
	10X YNB stock	100 mL
	1M PPi buffer pH 6	200 mL
	500X Biotin stock	2 mL
BMM10	10X YNB stock	100 mL
	1M PPi buffer pH 6	200 mL
	500X Biotin stock	2 mL
	Methanol conc.	50 mL
YPD	Yeast extract	10 g
	Peptone	20 g
	10X Dextrose	100 mL
	Agar (only for agar plates)	15

3.4.2 Buffer and stock solutions

10X YNB (Yeast Nitrogen Base):

33.5 g YNB w/aa were filled up to 250 mL deionized water and stirred until completely dissolved. Subsequently, the solution was sterile filtered twice, through a 0.45 and 0.22 µm filter. The stock solution was stored at 4 °C.

50% (v/v) Glycerol:

50 mL of 100% Glycerol were dissolved in 50 mL deionized water and stored at 4 °C.

500X Biotin:

20 mg Biotin were dissolved in 100 mL ddH_2O, sterile filtered through a 0.45 and a 0.22 µm filter and stored at 4 °C.

YPD (Yeast Extract Peptone Dextrose Medium):

10 g Yeast extract and 20 g Peptone were dissolved and mixed in 900 mL deionized water and autoclaved. After cooling down of the medium below 50 °C, 100 mL of 10X Dextrose were added. The medium was stored at 4 °C.

YPD Agar Plates:

Before autoclaving, YPD medium was supplemented with 15 g Agar per liter. For antibiotic selection, a Zeocin™ stock solution (100 mg/mL) was prepared and added to a final concentration of 25 µg/mL after autoclaving and cooling down of the medium to room temperature. The agar plates were stored at 4 °C.

1 M PPi buffer, pH 6:

46.4 g K_2HPO_4 and 100 g KH_2PO_4 were dissolved in 950 mL deionized water and mixed by stirring. The pH was adjusted to pH 6 under continuous stirring by the addition of conc. KOH. After the pH was set, the solution was filled up with deionized water to 1000 mL and autoclaved. The buffer was stored at 4 °C.

300 mM Tris-HCl buffer, pH 7:

3.63 g Tris were dissolved in 90 mL deionized water and mixed by stirring. The pH was adjusted to pH 7 under continuous stirring by the addition of 1M HCl. After the pH was set, the solution was filled up with deionized water to 100 mL. The buffer was stored at 4 °C.

4 mM *p*-Nitrophenyl butyrate (*p*NPB):

42 µL *p*-Nitrophenyl butyrate were mixed with 458 µL DMSO to a final concentration of 1%. This stock solution was frozen at -20 °C under exclusion of light.

10X MOPS buffer 0.2 M, pH 7:

82 g MOPS, 13.6 g sodium acetate and 40 mL 0.5 M EDTA pH 8 were dissolved in 1800 mL deionized water and mixed by stirring. The pH was adjusted to pH 7 under continuous stirring by slow and stepwise addition of NaOH. After the pH was set, the solution was filled up with deionized water to 2000 mL. The buffer was autoclaved and stored at room temperature under exclusion of light.

0.5 M EDTA, pH 8:

46.525 g Na_2-EDTA.$2H_2O$ were dissolved in 200 mL deionized water and mixed by stirring. The pH was adjusted to pH 8 under continuous stirring by the addition of NaOH (to dissolve EDTA.Na_2.$2H_2O$ completely, approximately 20 g of NaOH pellet were necessary). After the pH was set, the solution was filled up with deionized water to 250 mL. The buffer was stored at room temperature.

5% SDS solution:

50 g SDS were dissolved in 950 mL deionized water and mixed by stirring. The pH was adjusted to pH 8 under continuous stirring by the addition of conc. NaOH. After the pH was set, the solution was filled up with deionized water to 1000 mL. The solution was stored at room temperature.

20X MOPS buffer, pH 7.7:

209.2 g MOPS, 121.2 g Tris, 20 g SDS and 6.0 g Na_2-EDTA.$2H_2O$ were dissolved in 950 mL deionized water and mixed by stirring. Subsequently, the buffer was filled up with deionized water to 1000 mL. The solution was stored at 4 °C until use. For electrophoresis, the buffer was diluted to 1X with water resulting in the appropriate pH of 7.7. No additional acid or base was necessary to adjust the pH.

1M sorbitol:

18.2 g Sorbitol were dissolved in 100 mL deionized water and sterilized by filtration. The solution was stored at room temperature.

DEPC (Diethylpyrocarbonate) water:

1 mL DEPC at 0.1% end concentration was dissolved in 1 L deionized water. The solution was autoclaved and stored at 4 °C.

3.4.3 Chemicals

Table 3.6 All chemicals used during this thesis and their corresponding suppliers are listed below.

Chemicals	Company
α-D(+)-glucose monohydrate	Carl Roth GmbH, Karlsruhe, Germany
Aqua bidest. "Fresenius"	Fresenius Kabi Austria GmbH, Graz, Austria
Acetic acid (100%)	Carl Roth GmbH, Karlsruhe, Germany
Bacto™ peptone	Becton Dickinson and Company, Sparks, MD, USA
Bacto™ yeast extract	Becton Dickinson and Company, Sparks, MD, USA
Bacto™ Agar	Becton Dickinson and Company, Sparks, MD, USA
D-Biotin	Fluka Chemia AG, Basel, Switzerland
Difco™ Yeast Nitrogen Base w/o Amino Acids	Becton Dickinson and Company, Sparks, MD, USA
Dithiothreitol (DTT)	Carl Roth GmbH, Karlsruhe, Germany
DMSO	Karl Roth GmbH & Co., Karlsruhe, Germany
4-Nitrophenyl butyrate (*p*NPB)	Sigma-Aldrich GmbH, Vienna, Austria
Ethanol abs.	Merck KGaA, Darmstadt, Germany
Ethidium bromide (≥ 98%)	Carl Roth GmbH, Karlsruhe, Germany
Glycerol (≥ 98%)	Carl Roth GmbH, Karlsruhe, Germany
HCl (37%)	Merck KGaA, Darmstadt, Germany
K_2SO_4	Carl Roth GmbH, Karlsruhe, Germany
KOH	Carl Roth GmbH, Karlsruhe, Germany
KH_2PO_4	Carl Roth GmbH, Karlsruhe, Germany
$K_2HPO_4.3H_2O$	Carl Roth GmbH, Karlsruhe, Germany
KCl	Merck KGaA, Darmstadt, Germany
Methanol	Carl Roth GmbH, Karlsruhe, Germany
LB-medium (Luria/Miller)	Carl Roth GmbH, Karlsruhe, Germany

NaCl	Carl Roth GmbH, Karlsruhe, Germany
NaOH conc.	Carl Roth GmbH, Karlsruhe, Germany
Zeocin™	InvivoGen-Eubio, Vienna, Austria
TRIS	Carl Roth GmbH, Karlsruhe, Germany
Na_2-EDTA.$2H_2O$	Carl Roth GmbH, Karlsruhe, Germany
Bio-Safe™ Coomassie Stain	Bio-Rad Laboratories Inc., Hercules, CA, USA
Bio-Rad Protein Assay Dye Reagent Concentrate	Bio-Rad Laboratories Inc., Hercules, CA, USA
MOPS (≥ 99.5%)	Carl Roth GmbH, Karlsruhe, Germany
Glycerol FeedBeads	Adolf Kühner AG, Basel, Switzerland
SDS (≥ 99%)	Carl Roth GmbH, Karlsruhe, Germany
Bovine Serum Albumin (BSA)	Sigma-Aldrich GmbH, Vienna, Austria
DEPC water	Invitrogen by Thermo Fisher Scientific, Waltham, MA

3.5 Kits and protocols

3.5.1 Kits

Table 3.7 All kits and their corresponding suppliers used in this study are listed below.

Kit name	Company
SV Total RNA Isolation System	Promega GmbH, Mannheim, Germany
Luna® Universal One-Step qPCR Kit	New England Biolabs, Ipswich, MA
RQ1 RNase-free DNase	Promega GmbH, Mannheim, Germany
Quick Start™ Bradford Protein Assay	Bio-Rad Laboratories Inc., Hercules, CA, USA

3.5.2 Cultivation and sampling protocol

Cultivations of pre- and main cultures of *P. pastoris* strains were performed in sterilized 250 mL baffled wide-necked shake flasks covered with cotton sheets and rubber bands. Each *P. pastoris* strain was cultivated twice, in BMD and BMG medium. In this diploma thesis, two cultivation rounds were executed. For the first round, each target plasmid was prepared in biological duplicates for each medium, in the second round we decided to use biological

triplicates for each medium. For both rounds, sterile controls for BMD and BMG were prepared and cultivated.

To gather single colonies for the preparation of the pre-cultures, target *P. pastoris* strains were streaked out on YPD-agar plates supplemented with Zeocin™ as antibiotic and incubated at 30 °C for 48 h. After incubation, two different ONCs per variant were prepared in 250 mL baffled wide-necked shake flasks covered with cotton sheets and rubber bands: one containing 50 mL BMD, the other 50 mL BMG medium. A single colony of each *P. pastoris* strain carrying an expression cassette was transferred into each medium and incubated at 28 °C for around 18 h with vigorous shaking (100 rpm).

After approximately 18 h of incubation, we inoculated the main cultures containing 50 mL final volume of BMD or BMG for each variant with a starting D_{600} of 0.2. The main cultures were incubated at 28 °C with shaking (100 rpm) and two shake flasks with integrated O_2 sensors were connected with the PreSens online monitoring system to examine the growth behavior of the cells (Fischer, Hatzl, Weninger, Schmid, & Glieder, 2019). The shake flasks with the integrated O_2 sensors were placed onto the shake flask reader units (SFR) to transfer the measured data to the computer. A crucial point here was to use one shake flask from each medium containing the same variant as representatives for all the other flasks. All working steps were performed according Fischer, et. al (2019).

After around 24 h of cultivation, carbon sources were depleted as biomass was produced. BMD cultures had to be induced with BMM at 0.125% end concentration, whereas BMG cultures were fed with three glycerol feed discs. Exact induction times and feed with glycerol can be found in *table 3.8* and *3.9.* Around 48 h after inoculation, BMD cultures were pulsed again with 100% methanol to 0.125% final concentration. As the glycerol feed discs continuously released carbon source, no further feed was necessary here. Induction and feed of the cultures was executed based on the data provided by the PreSens system (Fischer et al., 2019).

To gain more information about the regulation of the modified promoter variants and to check if protein activity correlates with the amount of transcript present, samples were taken during the entire cultivation process from each shake flask in order to perform expression level determinations. To do so, the cultivation was stopped and the monitoring process was paused. The flasks were put out and the cell densities in each culture were measured under sterile conditions in an RNase-free environment. By a simple inference calculation, the required amount of culture was calculated to obtain 10 D_{600} units of cells. After sampling of the

appropriate amount of culture, the flasks were placed back in the shaker The cultures were shaken for at least five min before switching on the monitoring system again, just to allow a gentle adjustment to the environment. In *table 3.8* and *3.9* all sampling points during the entire processes are listed.

Table 3.8 Sampling points during first round of shake flask cultivation.
For precise sampling points, the PreSens online monitoring system was used providing information about biomass, O_2 saturation (% O_2), oxygen uptake rate (OUR) and temperature. After carbon source depletion, BMD shake flasks were induced with BMM and then further pulsed with 100% methanol, BMG shake flask were fed with glycerol feed discs.

Growth phase	Sampling point	Acronym
Lag-phase	4h after inoculation	t1
Log-phase	16 h after inoculation	t2
Begin stationary phase	30 h after inoculation and 1st induction for around 3 h / 3 h feed w/ glycerol	t3
Middle stationary phase	42 h after inoculation and 1st induction for around 24 h / 24 h feed w/ glycerol	t4
End stationary phase	65 h after inoculation and 2nd induction for around 22 h / 47 h feed w/ glycerol	t5

Table 3.9 Sampling points during second round of shake flask cultivation.
For precise sampling points, the PreSens online monitoring system was used providing information about biomass, O_2 saturation (% O_2), oxygen uptake rate (OUR) and temperature. After carbon source depletion, BMD shake flasks were induced with BMM and then further pulsed with 100% methanol, BMG shake flask were fed with glycerol feed discs.

Growth phase	Sampling point	Acronym
Lag-phase	5h after inoculation	t1
Log-phase	16 h after inoculation	t2
End-Log phase	20 h after inoculation	t3
Begin stationary phase	22.5 h after inoculation and 1st induction for around 2 h / 2 h feed w/ glycerol	t4
Middle stationary phase	44 h after inoculation and 1st induction for around 21.5 h / 21.5 h feed w/ glycerol	t5
Middle-end stationary phase	46 h after inoculation and 2nd induction for around 2 h / 23.5 h feed w/ glycerol	t6
End stationary phase	68 h after inoculation and 2nd induction for around 23.5 h / 49.5 h feed w/ glycerol	t7

At all sampling points, 10 D_{600} units of each shake flask were taken and centrifuged at 4 °C for 5 min and 4000 rpm. Finally, the supernatants were transferred into new tubes. The pellets containing the cells were frozen immediately after collection and stored at -80 °C until use. Similarly, the supernatants were kept at 4 °C. The whole sampling process was performed completely RNase-free: gloves, RNase-free tubes, filter tips and the workbench was cleaned with a 5% (w/v) SDS solution.

3.5.3 RNA isolation and quality assessment

3.5.3.1 RNA isolation process

The frozen cell pellets were disrupted, purified and RNA was isolated. As RNA isolation is an extremely sensitive process, all working steps were executed in a ribonuclease-free environment. In order to achieve such an environment, gloves were worn all the time and the workbench and equipment were continuously sterilize with 5% SDS solution (to inactivate ribonucleases). Furthermore, sterile disposable plasticware was used for the entire isolation and experimental procedure (RNase-free tubes and filter tips for the pipettes). Besides, all purification and initial washing steps were performed on ice. The glass beads used for cell lysis were aliquoted in several glass vessels, covered with DEPC water and autoclaved. After this, the beads were dried at 100 °C for at least 48 h.

To rapidly isolate high-quality RNA, substantially free of genomic DNA contamination, the SV Total RNA Isolation System from Promega (Promega GmbH, Mannheim, Germany) was used. The isolation process was performed according to the manufacturer´s protocol, but single steps had to be improved for our specific needs. Each cell pellet was resuspended in 100 µL of the following solution, which was not offered by the kit (only ß-mercaptoethanol was included in the kit and recommended to use for cell lysis at 0.1% final concentration): 1 M Sorbitol, 0.1 M EDTA pH 8 and ß-mercaptoethanol at 0.1% final concentration. For cell lysis, 175 µL RNA Lysis Buffer were added to the resuspended cells and homogenized by gentle pipetting. The solution was then transferred into new tubes containing 100 µL glass beads and vortexed for 20 min (speed 7, Vortex Genie 2). Then, the lysed cells were centrifuged for one minute, full speed at RT, and the supernatants were transferred into new reaction tubes. After the addition of 350 µL RNA Dilution Buffer and mixing by inversion, the cells were centrifuged for 10 min maximum speed at RT. The cleared lysate solutions were transferred into fresh tubes by pipetting without disturbing the pelleted debris. 200 µL of absolute ethanol were added to the cleared lysates and mixed by pipetting. The mixtures were then transferred onto the Spin Column Assemblies and centrifuged for one minute full speed. The washing step with RNA Wash Solution was carried out according to the manufacturer´s protocol. The DNase incubation mix, prepared as described in the protocol, was applied onto each Spin Basket and incubated for around 20 min at 20-25 °C. After incubation, 200 µL DNase Stop Solution was added to the Spin Baskets and centrifuged for one min maximum speed at RT. The washing steps with RNA Wash Solution were carried out according to the manufacturer´s protocol. The Spin Baskets were then transferred from the Collection Tubes to the Elution tubes and left open at RT for 20-

30 min for the residual ethanol to evaporate. Finally, 40 µL Nuclease-free water were added onto each Spin Basket´s membrane, incubated for 5 min at RT with the lid open and then centrifuged for 2 min full speed at RT. The Spin Baskets were discarded, the Elution tubes containing the purified RNA were capped and RNA concentration was determined spectrophotometrically at 260 nm. The samples were stored at -80 °C until further use.

3.5.3.2 RNA integrity and quality assessment

The integrity of the RNA samples was validated by applying them on 1% agarose gels. For this purpose, 2 g agarose were dissolved in 200 mL RNA running buffer composing of 20 mL 10X MOPS buffer pH 7, 180 mL deionized water and 20 µL EtBr. 7 µL of purified RNA per sample was mixed with 7 µL of RNA Loading Dye (2X) from New England Biolabs (New England Biolabs, Ipswich, MA) and heated at 65 °C for 10 min. After heat incubation, the samples were placed on ice until use. For each gel, 2 µL ssRNA Ladder from New England Biolabs (New England Biolabs, Ipswich, MA) were used as molecular marker and mixed with 8 µL RNA Loading Dye (2X) from New England Biolabs (New England Biolabs, Ipswich, MA) and heated at 90 °C for two min. After heat incubation, the ladder was placed on ice before use. Additionally, 10 µL GeneRuler 1 kb DNA Ladder from Thermo Fisher Scientific (Thermo Fisher Scientific Inc., Waltham, MA) were used as molecular marker per gel. For each RNA gel, 200 mL 10X MOPS buffer pH 7 and 1800 mL deionized water were mixed and used as RNA running buffer. The gel was run for 90 min at 100 V and 400 mA. All working steps were performed completely RNase-free as mentioned before – chamber, comb and electrophoresis tank were sterilized with 5% SDS solution.

In order to asses purity of the isolated RNA and to exclude the possibility of genomic DNA contaminations, PCR controls were carried out. Each reaction was prepared on ice by mixing 5 µL of 5X Q5® Reaction Buffer from New England Biolabs (New England Biolabs, Ipswich, MA), 0.5 µL of 10 mM dNTPs, 1.25 µL of each 10 µM forward and reverse primer (P14896, P16246), 0.25 µL of Q5® High-Fidelity DNA Polymerase from New England Biolabs (New England Biolabs, Ipswich, MA) with 300 ng RNA template. The reaction mix was filled up with ddH_2O to a final volume of 25 µL. As positive control, pKEB#1_CalB plasmid was used and amplified. Cycling conditions and temperature profile of each reaction was performed as the following: 1st stage: initial denaturation at 98 °C for 30 s; 2nd stage for 30 cycles: denaturation at 98 °C for 10 s, annealing at 58 °C for 30 s, extension at 72 °C for 30 s; 3rd stage: final extension at 72 °C for 7 min. After amplification, the PCR reactions were mixed with 4.2 µL DNA Gel Loading Dye (6X) from Thermo Fisher Scientific (Thermo Fisher Scientific Inc.,

Waltham, MA) and applied onto 1% agarose gels. As molecular marker, 10 µL GeneRuler 1 kb DNA Ladder from Thermo Fisher Scientific (Thermo Fisher Scientific Inc., Waltham, MA) were used per gel. The gels ran at 120 V and 400 mA for 50 min.

3.5.3.3 Digestion of genomic DNA contamination

As qPCR is a highly accurate and sensitive method, it was extremely important to work with high-quality RNAs, otherwise the results were not meaningful. If PCR showed gDNA contamination of the samples, they were further purified by digestion of gDNA. For this purpose, the RQ1 RNase-free DNase Kit from Promega (Promega GmbH, Mannheim, Germany) was used. All working steps were performed completely RNase-free and according to the manufacture's protocol. Per DNase treatment of RNA sample, 350 ng RNA were mixed with RQ1 RNase-free DNase 10X Reaction Buffer and 1 µL RQ1 RNase-free DNase (1 u/µg RNA). The reaction mix was filled up with Nuclease-free water to a final volume of 10 µL and incubated at 37 °C for 10 min. After incubation, 1 µL RQ1 DNase Stop Solution was added to terminate the reaction. To inactivate the DNase, the reaction was incubated for further 10 min at 65 °C. The digested RNA samples were used for expression level determination via qPCR without any further treatment.

3.5.4 Screening and protein assays

Amount and activity of the produced protein was determined by various protein assays, all depicted in the following paragraphs. As CalB is a secreted protein, the supernatants, obtained by centrifugation of cell cultures, contained the protein of interest. Hence, no further downstream processing was necessary before determination of enzyme activities as a read out to be correlated to expression efficiency.

3.5.4.1 Esterase activity determination

Esterase activity was determined according to Zhang *et al.* (2003). For each sample, 20 µL of the supernatant were used and transferred into a sterile optical 96-well microtiter plate in technical quadruplets. For one 96-well plate, 200 µL 4 mM *p*-nitrophenyl butyrate (*p*NPB) were mixed with 20 mL 300 mM Tris-HCl buffer pH 7. 180 µL of this assay solution were used per well. Immediately after adding the solution, color development was followed over 5 min at 405 nm at RT using a SynergyMx Plate Reader. The activity of the target protein was assayed by

measuring the formation of *p*-nitrophenol (*p*NP) over time at the respective wavelength (N. Zhang et al., 2003).

One unit of protein activity was defined as the amount of enzyme required to release 1 µmol *p*-nitrophenol per min (Σ = 17700M^{-1} cm^{-1}) (Su, Huang, Han, Zheng, & Lin, 2010). The volumetric activity of each sample was calculated using the formula stated below.

$$Volumetric\ activity\left[\frac{U}{mL}\right] = \frac{v_{max} * V_{total} * DF}{V_{sample} * \varepsilon * d}$$

3.5.4.2 Bradford determination

To determine total protein content, Quick Start™ Bradford Protein Assay instructions from Bio-Rad (Bio-Rad Laboratories Inc., Hercules, CA, USA) were used. Four dilutions of BSA protein standard from Sigma-Aldrich (Sigma-Aldrich GmbH, Vienna, Austria) were prepared as representatives of the protein solutions to be tested. As the linear range of this microtiter plate assay ranges from 0.05 mg/mL to approximately 0.5 mg/mL, a calibration curve was required in order to calculate total protein content of the samples. BSA protein standards were prepared as followed: 0.5, 0.25, 0.125 and 0.0625 mg/mL dilutions.

5 µL supernatant per sample and 5 µL of each protein standard were mixed with 200 µL diluted dye reagent solution. To prepare the solution, 4 parts deionized water were mixed with 1 part Protein Assay Dye Reagent Concentrate from Bio-Rad (Bio-Rad Laboratories Inc., Hercules, CA, USA) and filtered through a Whatman filter to remove particulates. For one optical 96-well microtiter plate, 4 mL dye reagent were mixed with 16 mL deionized water and filtered. Each sample reaction was prepared in technical quadruplets, whereas protein standards were prepared in technical triplicates. Once samples and assay solutions were applied, the 96-well plates were incubated at RT for at least 5 min. After incubation, absorbance was measured at 595 nm by using a SynergyMx Plate Reader.

For data analysis, a standard curve was created by plotting the measured absorbances at 595 nm (y-axis) against the concentrations of the standards in µg/mL (x-axis). The concentration x of each sample was calculated using the standard curve with the following formula:

$$y = k * x + d$$

3.5.4.3 Analytical polyacrylamide gel electrophoresis

For the analysis by SDS-PAGE, the samples were prepared by mixing 8 µL supernatant with 4 µL Invitrogen™ Novex™ NuPAGE™ LDS Sample Buffer (4X) from Thermo Fisher Scientific (Thermo Fisher Scientific Inc., Waltham, MA). Prior to application on the NuPAGE™ 4-12% Bis-Tris Protein Gels from Invitrogen (Invitrogen by Thermo Fisher Scientific, Waltham, MA), the samples were incubated at 70 °C for 10 min to denature proteins. Before applying the samples onto the gels, they were centrifuged for one min maximum speed at RT. As molecular marker, 4 µL PageRuler™ Prestained Protein Ladder by Thermo Fisher Scientific (Thermo Fisher Scientific Inc., Waltham, MA) were used. Electrophoresis was executed at 200 V and 400 mA for 50 min. As running buffer, 50 mL 20X MOPS were mixed with 950 mL deionized water.

3.5.5 Protocols for qPCR

3.5.5.1 General procedure to perform qPCR

As purity and integrity assessments confirmed high quality RNAs, the isolated samples could be used for qPCR to determine expression level. For each qPCR reaction, template RNA at a final concentration of 0.2 ng/µL (0.4 ng absolute RNA) was used.

To determine expression levels of the target strains, qPCR was performed and executed according the Luna® Universal One-Step qPCR Kit from NEB (New England Biolabs, Ipswich, MA). This kit provided an optimized protocol for dye-based real-time quantitation of target nucleic acids via the SYBR® Green I method. As qPCR is an extremely sensitive and highly accurate detection method, pipetting was executed very carefully in order to avoid any contamination and to ensure accurate quantification results and consistency of pipetting volumes. Besides, all reactions were run in technical duplicates or triplicates to reduce outlier traces like plate issues, edge effects, bubble formation or other problems. To evaluate amplification specificity of all samples, melt curve analyzes were added at the end of each quantification run. Additionally, it was especially important to perform all working steps in a ribonuclease-free environment: continuous sterilization of the workbench and equipment with 5% SDS, sterile disposable plasticware (including RNase-free tubes and filter tips), RNase-free MicroAmp™ Optical 96-Well Reaction Plates and all working steps were carried out wearing gloves and on ice.

The total volume for the appropriate amount of reactions was determined as described in *table 3.10* (adding 10% overage). Assay solutions containing all components except the nucleic acids were combined and mixed gently by pipetting up and down. By using a multichannel pipette, the assay solutions were aliquoted into the 96-well reaction plate. Subsequent, the RNA templates were added and the plate was sealed with MicroAmp™ Optical Adhesive Films. Briefly, the reaction plate was centrifuged for one minute at 3000 rpm at RT.

Table 3.10 Reaction setup to perform qPCR experiments.
Each reaction was prepared by mixing 10 µL Luna Universal One-Step Reaction Mix (2X), 1 µL Luna WarmStart® RT Enzyme Mix (20X), 0.8 µL of each 10 µM forward and reverse primer with 0.4 ng absolute RNA and filled up with Nuclease-free water to a final volume of 20 µL. Due to the large amount of samples to be tested, diploma mixes were prepared containing each component except the template RNA. All working steps were performed on ice and completely RNase-free. Reactions were gently mixed through pipetting.

Component	20 µL Reaction	Final concentration
Luna Universal One-Step Reaction Mix (2X)	10 µL	1X
Luna WarmStart® RT Enzyme Mix (20X)	1 µL	1X
Forward primer (10 µM)	0.8 µL	0.4 µM
Reverse primer (10 µM)	0.8 µL	0.4 µM
Template RNA	2 µL	0.2 ng/µL (0.4 ng absolute RNA)
Nuclease-free Water	up to 20 µL	

After centrifugation, the plate was placed into the qPCR instrument and the program was started according the appropriate detection setups as stated in *table 3.11*.

Table 3.11 Instrumental detection setup for the qPCR machine.
First step was the reverse transcription of the template RNA into the appropriate cDNA via Luna WarmStart Reverse Transcriptase. After reverse transcription, the target was quantified similar to a standard PCR procedure. As amplification preceded, the fluorescence accumulation of SYBR® Green dye was captured by the instrument after each cycle. At the end of

quantification, melt curve analyzes were added to evaluate amplification specificity. All results were analyzed using the 7500 Software v2.0.6 from Applied Biosystems.

Cycle step	Temperature	Time	Number of cycles
Reverse Transcription	55 °C	10 min	1
Initial Denaturation	95 °C	1 min	1
Denaturation	95 °C	10 s	
Extension	60 °C	60 s (+ plate read)	45
Melt curve	60-95 °C	60 min	1

3.5.5.2 Validation experiment and primer design

In this diploma thesis, various primers had to be designed to perform expression level determination via qPCR. Considering the facts that poorly annealed primers or those resulting in primer-dimer formation may significantly impact the reliability and quality of the qPCR results, it was crucial to design them properly under equal conditions. For this purpose, "PrimerQuest Tool", "OligoAnalyzer Tool" and "Primer-BLAST" were used. All primers were designed to have the same amplicon length of around 141 bp as well as the same melting temperature of 64 °C for both, the target and reference gene (HKG). Additionally, it was important to reduce the risk of 3´self-complementary structures and to aim for a GC content of 40-60%. Moreover, the primers should not bind to any intron sequences. As primer design is a critical step, several sets have been tested in order to obtain optimal primer pairs for qPCR.

In the course of a validation experiment, different HKGs were tested to analyze if amplification of both genes (target and HKG) result in comparable efficiencies. To determine if the method of choice, $\Delta\Delta C_T$ method, was valid or not, a validation experiment was executed. In order to obtain such a valid $\Delta\Delta C_T$ calculation, the amplification efficiency of the target and reference gene must have been approximately equal. Therefore, the PCR efficiency of both amplification reactions was defined. The amplification efficiency of each reaction was calculated using the following formula:

$$Exponential\ amplification = 10^{\left(-\frac{1}{slope}\right)}$$

$$PCR\ efficiency = \left[10^{\left(-\frac{1}{slope}\right)}\right] - 1$$

To calculate PCR efficiency of a qPCR experiment, dilution series from the specific target were prepared to look at how ΔC_T ($C_{T_{target}} - C_{T_{reference}}$) varied with template dilution. For the first validation experiment using *ARG4* mRNA as reference, dilution series ranging from 6.4 ng to 0.025 ng absolute RNA (final concentration: 3.2–0.0125 ng/µL) with a dilution factor of 4 were prepared and mRNA isolated from variant 26 at timepoint 4 (t4_BMG) was used as target. For the second validation experiment using *ACT1* mRNA as reference, dilution series ranging from 1.6 ng to 0.00625 ng absolute RNA (final concentration: 0.8–0.003125 ng/µL) with a dilution factor of 4 were prepared and mRNA isolated from the *P. pastoris* strain employing the parental P_{CTA1} promoter at timepoint 3 (t3_BMD) was used as target. The exact sampling timepoints for both cultivation rounds are listed In *table 3.8* and *3.9*.

For both validation experiments, dilution series of the target for the appropriate range were prepared as well as master mixes containing all components except the template (*table 3.10*). Per run, 4 master mixes were tested, each of them containing different primer pairs – one contained the primers for the target, whereas each of the others contained different primers for the corresponding reference gene. The master mixes were prepared and aliquoted into 96-well reaction plates in technical triplicates and gently mixed with each target dilution.

The primer sets for the target (CalB) and reference gene (*ARG4* or *ACT1*) with the most similar PCR efficiencies were chosen for the implementation of qPCR. Based on these results, the first qPCR run was performed using the primer pair CalB_1 (fw. and rev.) for the target gene and *ARG4*_qPCR_fw_3 and *ARG4*_qPCR_rev_3 as primer pair for the reference gene (*table 3.3*). The second qPCR was executed using the same primers for the target gene and *ACT1*_1fw and *ACT1*_1rev as primers for the reference gene (*table 3.3*).

To calculate the fold change x (RQ) of each sample, the following formulas were used in the appropriate order as described below:

1. $$\Delta C_T = C_{T\,target} - C_{T\,reference}$$

3. $$\Delta\Delta C_T = \Delta C_{T\,test\ sample} - \Delta C_{T\,calibrator}$$

4. Standard deviation of the $\Delta\Delta C_T$ value: $s = (s_1^2 + s_2^2)^{1/2}$,

where $s_1 = standard\ deviation\ of\ C_{T\,target}\ values$,

where $s_2 = standard\ deviation\ of\ C_{T\,reference}\ values$

5. $$RQ\ (fold\ change\ x) = 2^{-\Delta\Delta C_T}$$

6. $$RQ_{max} = 2^{-(\Delta\Delta C_T) - s\Delta\Delta C_T}$$

7. $$RQ_{min} = 2^{-(\Delta\Delta C_T) + s\Delta\Delta C_T}$$

4 RESULTS AND DISCUSSION

The methylotrophic yeast *P. pastoris* has become a widely used host organism for the production of recombinant proteins based on the strong and tightly regulated promoters of the MUT pathway. In particular, the P_{AOX1} was often applied for heterologous gene expression due to its strong inducibility with methanol, as well as simple availability as part of the basic *Pichia* expression kits. As oxidation of methanol causes the production of H_2O_2, catalases are produced upon glucose depletion in order to be prepared for the toxic by-product (Hartner & Glieder, 2006). As methanol represents a fire and health hazard, especially for large-scale production in industry, alternative induction mechanisms have to be found. Furthermore, consumption of methanol represents technical drawbacks as it provokes high heat evolution as well as an enhanced oxygen demand during fed batch phase (Prielhofer et al., 2013).

The aim of this project was to characterize new promoter variants with a tight regulatory mechanism that does not require methanol as inducer but still achieves high product yields. The de-repressible promoter P_{CTA1} of the *P. pastoris* catalase represented a potential alternative to regulatory sequences like the P_{AOX1}, which depend on methanol as induction agent in addition to prior de-repression. This is an attractive alternative to other methanol free strategies such as transcription factor overexpression or engineered platform strains with completely altered transcription factor patterns (Vogl et al., 2018). Strategies based on engineered promoters enable a versatile range of regulatory profiles and strength and thus can easily be adapted to the specific needs of individual target proteins which need to be produced.

In this study, *P. pastoris* strains secreting CalB under control of novel synthetic P_{CTA1} variants, were physiologically characterized to demonstrate their functionality. These two promoter variants named 19 and 26 originate from the endogenous *P. pastoris* catalase promoter (P_{CTA1}) and are triple-mutants as described in *section 3.2.* To get a deeper understanding of their mode of regulation and level of expression, target strains were cultivated under different experimental conditions. As many papers describe, the standard procedure for recombinant protein production in *P. pastoris* is based on the strong and tightly controllable promoters of the MUT pathway, for which expression is repressed at excess of glucose or glycerol, de-repressed upon carbon source depletion and initiated upon induction with methanol (V. Looser et al., 2014; Verena Looser et al., 2017). In order to understand how the P_{CTA1} variants behave under certain cultivation conditions, the respective *P. pastoris* strains were cultivated in minimal media with

two different sole carbon sources, dextrose and glycerol, BMD and BMG, respectively. The goal of the experiment was to obtain information about when the promoter is switched on and if protein activity could be detected without using methanol as an inducer by taking samples during the entire cultivation process. By working with the PreSens online monitoring system, growth behavior of the cells could be followed during the entire cultivation, which helped in chosen sampling points. To compare the performance of 19 and 26 and to see if they provoke any changes (e.g. improved promoter strength, enhanced regulation, more transcript production, higher protein yields, etc.), the transformants expressing reporter proteins regulated by the parental promoter P_{CTA1}, were also cultivated, as well as strains employing the P_{DF} promoter, which was used as benchmark for strength. The negative control (NC) was made with a CalB expression cassette lacking any promoter. In the BMD cultures, cells were grown on glucose to produce biomass until carbon-source depletion. We took samples here to see if the promoter is already switched on in the growth phase, during consumption of carbon-source, and subsequently started induction with methanol based on the data provided by an online monitoring system. By taking samples after glucose depletion and shortly after/during induction with methanol, conclusions can be drawn about how the promoter is regulated and if this effect causes changes in the transcript production. Furthermore, cells were grown in medium containing glycerol as sole carbon source. As described by Looser et al., excess of glycerol in cultivation represses target protein production, whereas glycerol limitation leads to product formation (V. Looser et al., 2014; Verena Looser et al., 2017). Theoretically, the P_{CTA1} and its variants should be repressed during excess of glycerol and induced in the de-repression state, when levels of glycerol are kept limited. Based on this, after initial depletion of glycerol excess, glycerol feed discs were added to the BMG cultures to keep carbon source level constant but low.

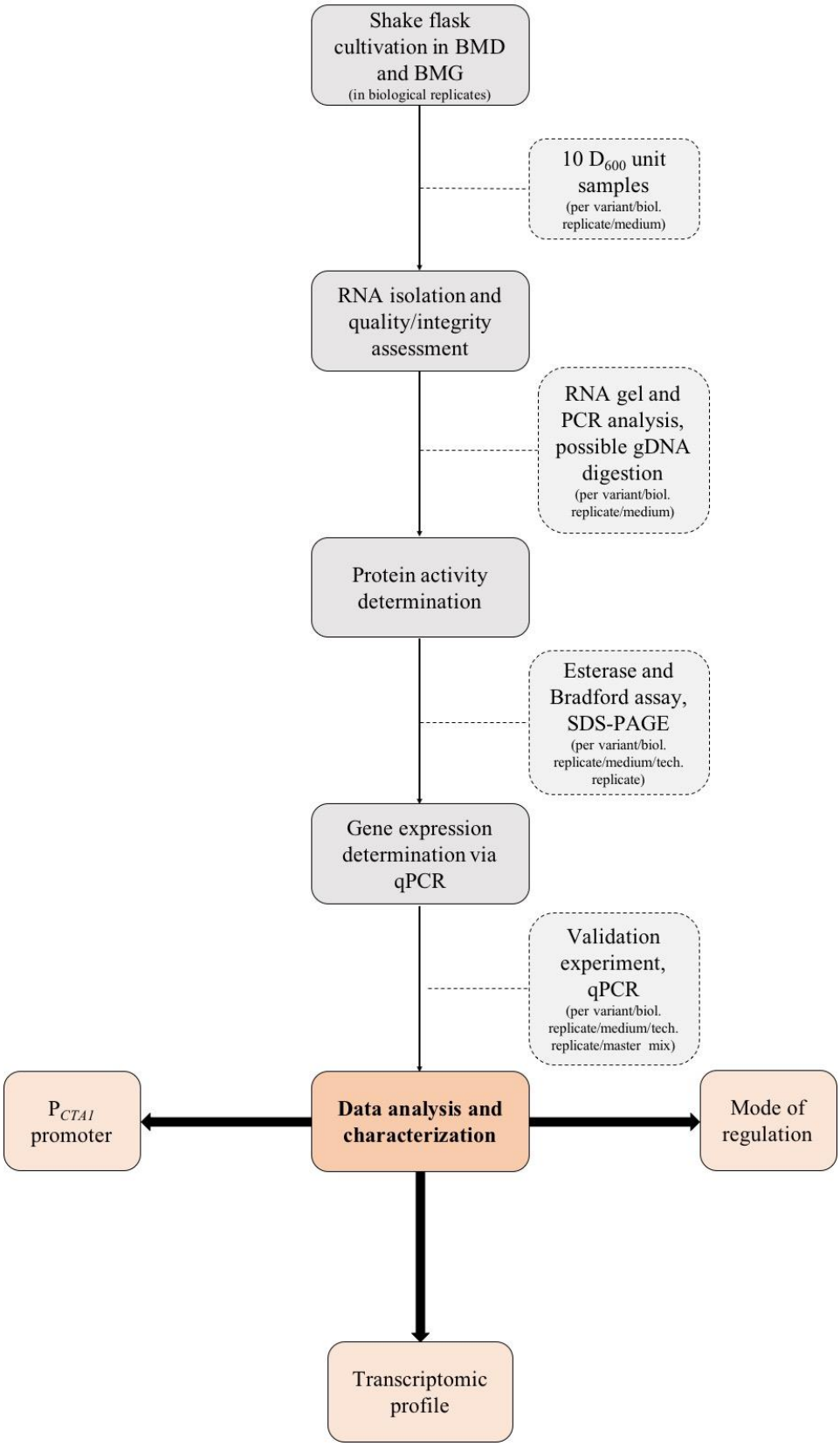

Figure 4.1 Schematic representation of the entire experimental procedure to determine gene expression levels in target *P. pastoris* strains.

This procedure has been performed twice during this diploma thesis. Starting with shake flask cultivation, cells were harvested according the data obtained by the PreSens online monitoring system. The supernatants were applied for protein activity measurements, whereas the pellets were further

processed and used for RNA isolation. Once purity and integrity assessment confirmed high quality, the isolated and purified RNA samples were analyzed via qPCR to determine expression levels of the target gene. To normalize target RNA input amounts, two different HKGs were validated and tested: *ARG4* was used for the first round of cultivation, *ACT1* for the second round. Through numerous preliminary experiments, individual steps in this procedure have been optimized and improved in order to ensure reproducible results. Gene, gene expression analysis can be used to understand the promoters mode of regulation and whether the induction with methanol or a de-repressed cultivation mode leads to higher protein yields. Moreover, these results used to determine the transcriptomic profile may contribute to a deeper understanding of the basal transcription machinery and how certain mutations in the core promoter affect expression itself. Overall, the knowledge obtained by characterization of P_{CTA1} variants can provide insight into their regulation patterns and may facilitate the finding of possible potential candidates for a methanol-free induction.

4.1 Shake flask cultivation and regulated gene expression of P_{CTA1} promoter variants

All in all, two cultivation rounds were carried out, which in principle were planned according to the same requirements: The respective *P. pastoris* strains were cultivated in two different media (BMD and BMG) and monitored via the PreSens online system. Based on the data provided by this system, the cultures were induced/fed with methanol/glycerol, respectively, and samples were taken and processed (i.e. RNA isolation and quality/integrity assessment). Additionally, quantity and activity of the secreted protein was determined (i.e. Esterase and Bradford assay, SDS-PAGE). Finally, the variants were analyzed and characterized by the use of qPCR. *Figure 4.1* shows the schematic representation of the entire experimental procedure used during this diploma thesis.

In the first cultivation round, variants 19 and 26 were cultivated together with strains employing the parental P_{CTA1} (WT), the P_{DF} and the vector without promoter (NC) as controls. All variants were cultivated and analyzed twice, in BMD and BMG, in biological duplicates. We took samples during the entire cultivation process according to the data provided by the PreSens online monitoring system. As indicated in *figure 4.2-A*, samples were taken at different stages of cell growth in the lag, log and stationary phase. After around 27 h of cultivation, heterologous protein production (CalB) was induced by the addition of methanol and glycerol feed discs for BMD and BMG cultures, respectively. As glycerol feed discs release a constant amount, no further addition of feed discs was required. In contrast, BMD cultures had to be pulsed again to keep methanol at limited levels after around 42 h of cultivation.

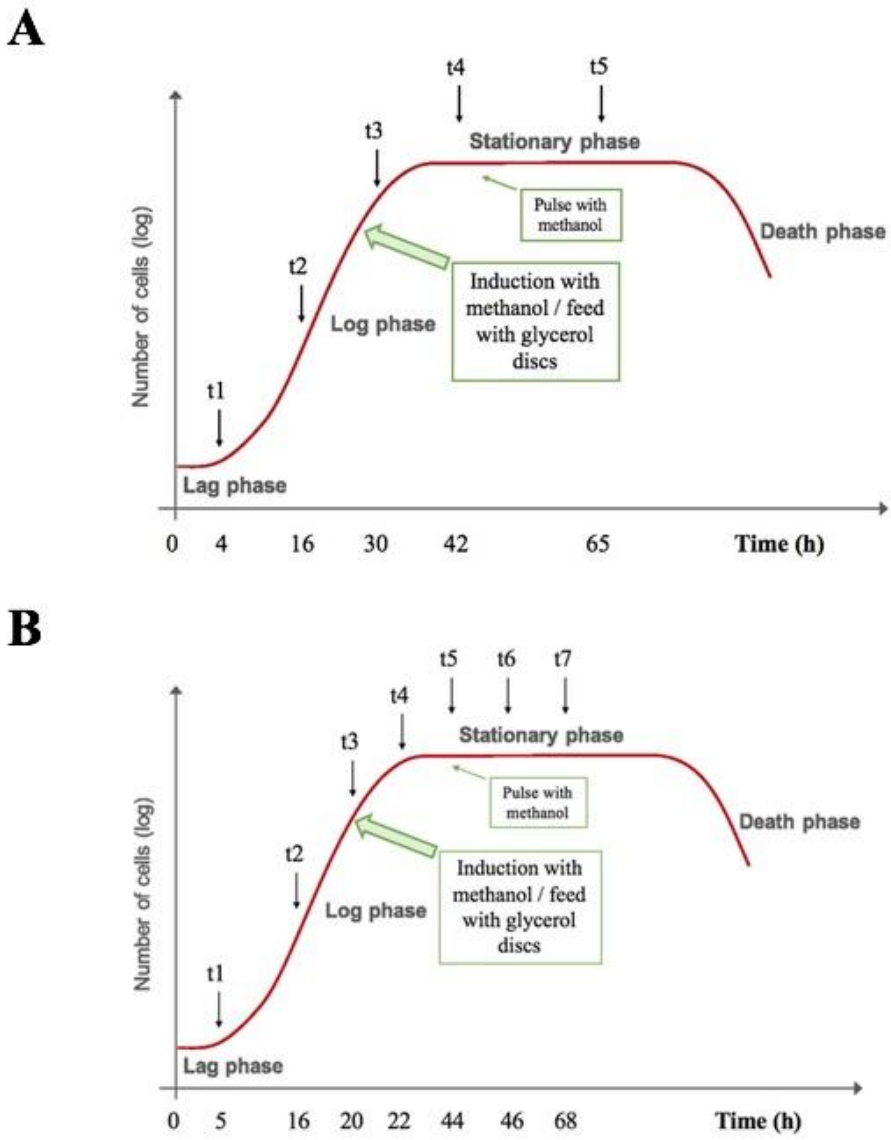

Figure 4.2 Sampling strategy during first (*A*) and second (*B*) round of cultivation.
This diagram gives an overview of the sampling strategy during the entire cultivation process. Based on the data obtained by the PreSens online monitoring system, cultures were induced/fed with methanol/glycerol feed discs, respectively, after around 27 (*A*) and 20 (*B*) h of cultivation when carbon-source was depleted. Cells were harvested at different timepoints: t1-t5 for *A* and t1-t7 for *B*. Furthermore, BMD cultures were pulsed with methanol after 42 (*A*) and 44 h (*B*). Based on the results of numerous preliminary experiments, a minimum amount of cells was harvested in order to achieve high RNA yields after isolation. As a result, 10 D_{600} unit samples were taken and further processed.

The second round of cultivation was prepared according to the same prerequisites used to execute the first round of cultivation. Target *P. pastoris* strains, in this case variant 19, the parental P_{CTA1} and the P_{DF}, were cultivated twice, in BMD and BMG, in biological triplicates. According to the PreSens online monitoring system, 10 D_{600} unit samples were taken during

the entire cultivation process under sterile and RNase-free conditions. Sampling was performed as mentioned above. As can be seen in *figure 4.2-B*, target gene production was started after around 20 h of cultivation with the addition of methanol and glycerol feed discs in BMD and BMG cultures, respectively. Moreover, cultures containing BMD medium were pulsed again after around 44 h of cultivation to keep methanol at limited levels.

Overall, both cultivation processes were performed over around 5 days (65 h of cultivation for the first, 68 h for the second cultivation process). Usually, 4 days before starting with the inoculation of the main cultures, the target *P. pastoris* strains were streaked out to obtain single colonies on the respective agar plates. The night before inoculation, ONCs were prepared for each variant in both, BMD and BMG media and incubated them as described in *section 3.5.2*. After approximately 18 h of incubation, depending on the varying cell densities, the main cultures were inoculated with an initial D_{600} of 0.2 and the monitoring with the PreSens online system was started. As described in *figure 4.2*, cells were harvested at different timepoints.

4.1.1 Overview of cell densities measured in the course of both cultivation processes

As mentioned above, 10 D_{600} units were taken from each shake flask during the entire cultivation processes to analyze and compare gene expression levels of the target protein. However, handling, sampling and reprocessing of all samples were performed under the same requirements. As can be seen from *figure 4.3*, similar levels of cell densities were achieved at the respective timepoints for both cultivation processes. Generally, it was difficult to ensure the exact same cultivation conditions for both processes as constant conditions could not be fully given. External factors like temperature and humidity of the room, time of sampling and duration of cultivation interruption, as cell density measurements and harvesting of the cells caused breaks in the cultivation processes, could have affected biomass production. Anyway, through comparison of both cultivation processes (*figure 4.3*), it appears that cellular growth behavior proceeded similarly.

Figure 4.3 shows the amount of cells present at the respective timepoints of sampling in BMD and BMG during both cultivation processes. All in all, similar levels of cell densities were measured for both cultivation processes (see appendix, *figure 9.6*)

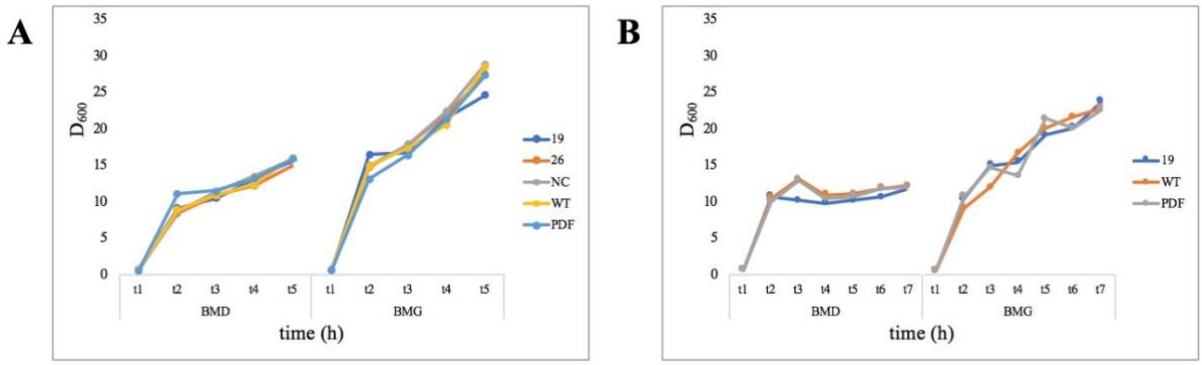

Figure 4.3 Comparison of the amount of cells present at the respective timepoints during cultivation round 1 (A) and 2 (B) between BMD and BMG cultures.
In both graphs, the cell densities (y-axis) are plotted against the timepoints (x-axis) at which the cells were harvested. Samples taken before t3 (*A*) and t4 (*B*) were cultivated at excess of glucose or glycerol. All samples that were subsequently taken, were cultured under induced conditions as methanol or glycerol feed discs were added to the cultures.

As can be obtained from *figure 4.3*, the amount of cells increased proportionally with time, which in turn reflects the different stages of growth (lag, log and stationary phase). Samples taken before t3 (*A*) and t4 (*B*) were cultivated at excess of glucose or glycerol, wherefore the cells were mainly producing biomass. Since we started protein production by induction and feed with methanol and glycerol feed discs (at t3 for *A* and t4 for *B*), respectively, the main focus of the cells was on protein production. In BMD cultures, especially at the beginning of cultivation, there was a sharp increase in biomass formation, which in turn tended to diminish and falter over time. In contrast, a constant increase of biomass formation could be observed for BMG cultures up to the very last moment of sampling. As expected, cells cultivated in BMG produced biomass throughout the entire cultivation process, as they were consistently supplied with glycerol. On the other hand, for cells cultured in BMD, a clearer switch from biomass formation to protein production was observed, as the cell densities obtained after induction hardly differed from before due to slow growth of Mut^S strains on methanol. This in turn reflects the fact that cells were mainly engaged in heterologous gene expression and to keep number of cells stable.

As can be seen in *figure 4.3*, similar levels of cell densities could be measured for both cultivation processes, although exact same external conditions could not be guaranteed. Factors such as handling, sampling, cultivation interruptions, temperature and humidity of the room, differed from the first cultivation process, thereby variations in the cultivation patterns were expected. Overall, comparable cultivation conditions were created, not only for both cultivation

processes, but also for each shake flask in the corresponding process, as only minor deviations were obtained within the cultivation rounds themselves.

4.2 RNA isolation process and quality assessment of total RNAs

During both cultivation processes, cells were harvested according to different time points as can be obtained from *figure 4.2* and further processed them. The cell pellets were disrupted, purified and RNA was isolated. Furthermore, the quality and integrity of the isolated samples were tested by applying them onto RNA gels. Additionally, PCR controls were performed to check for potential gDNA contaminations left in the samples, as described in *section 3.5.3.2*.

To address rapid isolation of high-quality RNA we worked according the SV Total RNA Isolation System from Promega (Promega GmbH, Mannheim, Germany). This kit provides a rapid and safe protocol to isolate RNA from plant tissues, yeasts and bacteria and offers purification without the necessity of phenol/chloroform extraction. The entire procedure involves effective cell disruption, in this project performed with the use of RNase-free glass beads, denaturation of nucleoprotein complexes, inactivation of endogenous ribonucleases and removal of contaminating DNA and proteins by DNase I digestion. To inactivate endogenous RNases present in the cell extracts, the kit offers the disruptive and protective properties of guanidine thiocyanate (GTC) and ß-mercaptoethanol. Finally, the purified RNA samples were eluted in nuclease-free water and spectrophotometrically measured at 260 nm to determine their concentrations, where 1 absorbance unit (A_{260}) equals 40 µg of single-stranded RNA per mL. The purity of the total RNAs obtained can also be estimated by spectrophotometry from the relative absorbances at 230, 260 and 280 nm according the "*SV Total RNA Isolation System*" (Promega GmbH, Mannheim, Germany). The ratio of A_{260}/A_{230} and A_{260}/A_{280} strongly depends on the individual starting materials as well as the performed procedure. In general, cultivated cells should result in a ratio of around 2.0 for both, A_{260}/A_{230} and A_{260}/A_{280} (See Kobs, 1998). By working according the Promega "*SV Total RNA Isolation System*", the total RNAs usually exhibit an A_{260}/A_{230} ratio of 1.8-2.2 and A_{260}/A_{280} ratio of 1.7-2.1.

Generally, it was of prime importance to work in a ribonuclease-free environment to inactivate potential RNases. Overall, the isolation process was performed according to the manufacturer´s protocol as described in *section 3.5.3.1,* but single steps had to be improved and optimized for our specific needs.

4.2.1 Improvements of single steps of the "SV Total RNA Isolation System" from Promega

One of the biggest problems to overcome was to find a suitable amount of cells for isolation to achieve high RNA yields. In addition, a simple and reproducible way had to be established in order to lyse the cells, as the kit´s protocol suggested the use of 20-gauge needles to shear the genomic DNAs. Unfortunately, this process would be extremely time-consuming and difficult to carry out due to the high number of samples to be processed. The working instructions offered by the kit recommended to use $1x10^5$-$5x10^6$ cells per 175 µL RNA Lysis Buffer, but the exact amount of cells equaling 1 D_{600} unit is hard to estimate. For the yeast *S. cerevisiae*, the kit suggests to use approximately $4x10^7$ cells. Generally, if the lysate contains more RNA, the capacity of the Spin Basket is exceeded and thereby, RNA could be lost during washing steps as excess causes clogging of the membranes and poor purification. On the other hand, an insufficient amount of starting material in turn increases the risks of a failed process as isolation involves several washing and purification steps, thus losses of total RNAs are very likely, according the manufacture's protocol.

However, many scientists claim that it is hard to predict the exact amount of cells that equal 1 D_{600} unit of cell density as this measurement mainly depends on the spectrophotometer used. As described by Klose et al., 1 D_{600} unit of *S. cerevisiae* (BY4741) grown in SCglc 2% Bacto agar media at 30 °C equals $2.7x10^7$ cells per mL (Klose et al., 2012). Compared to this, 1 D_{600} unit of *P. pastoris* (SMD1163) cultivated in BMMY medium (0.5 % methanol) corresponds to $5x10^7$ cells per mL (Asada et al., 2011). In the course of preliminary experiments, several parameters were tested such as varying D_{600} units of cells for RNA isolation and different amounts of glass beads for cell lysis. In a test series, target *P. pastoris* strains were streaked out on the appropriate YPD agar plates supplemented with Zeocin™ (25 µg/mL final concentration) as antibiotic and incubated for around 48 h at 28 °C. After incubation, sterile spatulae were used to scrape off the colonies from the plates and to resuspend them in 100 µL of 1 M Sorbitol, 0.5 M EDTA pH 8 and 0.1% ß-mercaptoethanol (similar as described in *section 3.5.3.1).* The cell densities of the resuspended colonies were measured and prepared samples with different D_{600} units of cells ranging from 1 to 14, approximately. Moreover, varying amounts of glass beads for cell lysis were tested by preparing reaction tubes containing 50 and 100 µL of glass beads. To obtain approximately 50 and 100 µL of glass beads, reaction tubes containing the respective amounts of water were used for comparison.

The results of the test series showed that loses of RNA during the isolation and purification process are too high when processing D_{600} units of less than 7. In order to obtain reliable and high RNA yields, at least 7 D_{600} units or more must be used. Besides, this procedure is an extremely sensitive and complex process depending on many different factors that could influence a successful isolation. As a fact of this, it is better to purify excess of RNA and deal with potential clogging than processing less amounts of starting material and risk a failed isolation. By using an excess of starting material, it is essential to use more glass beads, as the higher the number of cells to be processed, the more glass beads are required for cell rupture. Otherwise, the entire endogenous RNAs cannot be completely released.

Overall, highest RNA yields were obtained by using 100 µL of glass beads in combination with high D_{600} units of cells. Based on these results, 10 D_{600} unit samples were taken during the entire cultivation process and cell disruption was executed with an amount of glass beads corresponding to 100 µL.

4.2.2 RNA integrity and quality assessment of total RNAs obtained

To determine the integrity of the purified RNAs, 7 µL of the samples were applied, depending on the concentration between 2-5 µg of total RNA, onto 1% RNA-agarose gels (RNA gels), prepared as described in *section 3.5.3.2*, and electrophoresis was performed. As molecular markers, both, ssRNA Ladder and GeneRuler 1kb DNA Ladder were used. Generally, the ratio of 28S and 18S eukaryotic ribosomal RNAs should be around 2:1 when using ethidium bromide staining agarose gels, indicating intact and pure RNAs. For degraded samples, the RNAs would appear as lower molecular weight smear according the "*SV Total RNA Isolation System*" (Promega GmbH, Mannheim, Germany). As an example, *figure 4.4* shows the results of an integrity assessment by the use of RNA gels.

Overall, RNAs could be successfully isolated and purified from the samples taken during both cultivation processes. Besides, intact and high quality RNAs could be confirmed by evaluating RNA gels. Unfortunately, although all preparation steps were performed according to the manufacture's protocol, the ssRNA Ladder could not be effectively separated.

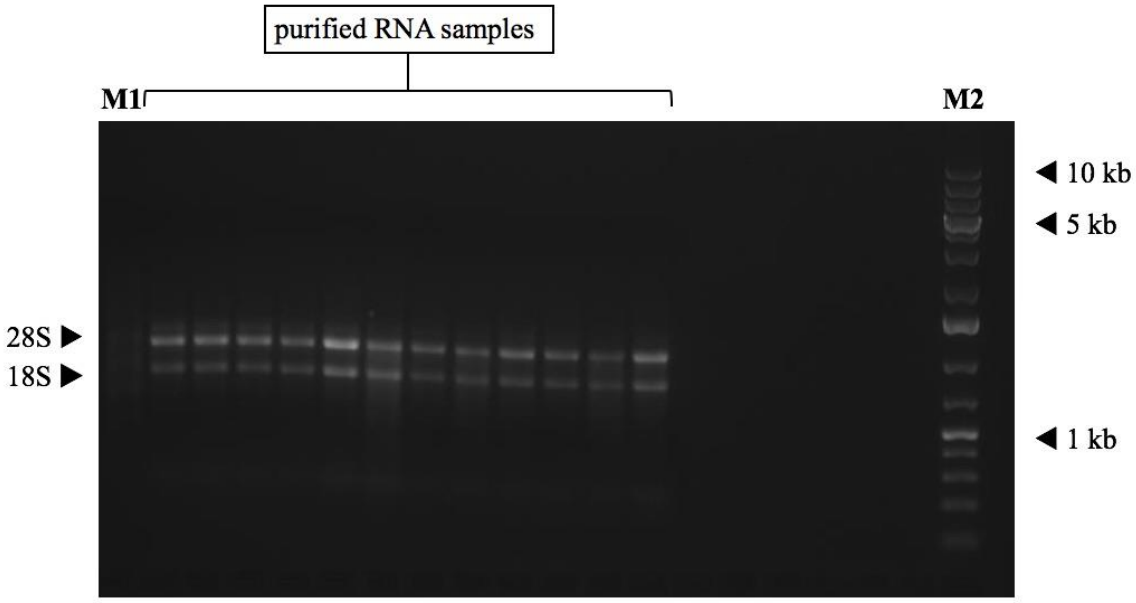

Figure 4.4 Agarose gel electrophoresis for integrity assessment of isolated and purified RNA samples of target *P. pastoris* strains.
In the course of this diploma thesis, hundreds of samples were taken during the cultivation processes (round 1 and 2) and isolated and purified using the SV Total RNA Isolation System from Promega. To confirm their quality and integrity, each isolated RNA sample was applied onto RNA-agarose gels. This figure shows the result of an example RNA gel applied for parental P_{CTA1} samples at timepoints t6 and t7 in BMD and BMG, respectively. The 18S and 28S ribosomal RNA bands are clearly visible in all samples, indicating intact and high quality RNAs. Degraded RNA would appear as lower molecular weight smear. In case of gDNA contaminations, bands would appear at approximately 10 kbs (M2 at 10 kb), which is visible here. As molecular markers, we used ssRNA Ladder from New England Biolabs (M1) and GeneRuler 1 kb DNA Ladder from Thermo Fisher Scientific (M2). 1% (w/v) RNA-agarose gels are shown.

In order to test the isolated and purified RNA samples for possible genomic DNA contaminations, PCR controls were carried out and applied onto 1% agarose gels. To compare the target RNAs, a positive control was prepared and amplified via PCR, together with the samples as described in *section 3.5.3.2*. Through the use of agarose gel electrophoresis, DNA fragments can be separated by size within the gels resulting in visible bands at the expected sizes (Lee, Costumbrado, Hsu, & Kim, 2012). Under ideal requirements, only the positive control should result in a visible band at the expected size as plasmid DNA was amplified during PCR. Since the RNA samples should only be composed of RNA, no bands should be visible here unless genomic DNA contaminations were present, which in turn would have been amplified in the course of the PCR reaction. As an example, *figure 4.5* shows the results of PCR controls to check for genomic DNA contaminations left in the RNA samples. As molecular maker, GeneRuler 1kb DNA Ladder was used.

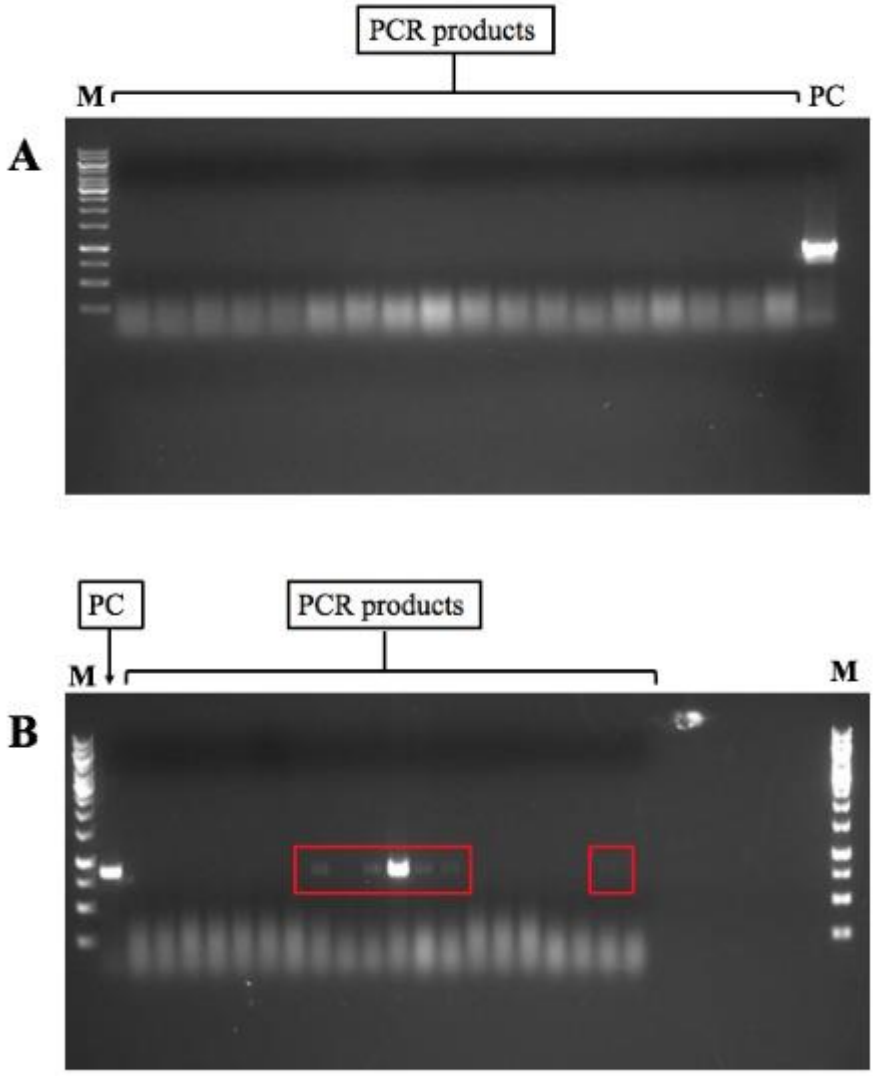

Figure 4.5 PCR controls to check for gDNA contaminations left in the isolated and purified RNA samples of target *P. pastoris* strains.
In the course of quality and integrity assessment of the total RNAs obtained, they were also tested for possible gDNA contaminations left as qPCR mainly depends on high-quality RNA samples. *A* shows high-quality RNA samples as only the positive control (PC) resulted in an amplification product. On the other hand, some samples were contaminated with gDNA as their PCR reactions resulted in an amplification product, as can be seen in *B*. Samples containing gDNA contaminations are highlighted by the red rectangles. For *A* and *B*, GeneRuler 1 kb DNA Ladder from Thermo Fisher Scientific was used as molecular marker (M). In *A* and *B*, 1% (w/v) agarose gels are shown.

Overall, high-quality RNA samples were isolated and purified. Additional gDNA digestion was only performed for those samples where PCR controls showed amplification products according to the RQ1 RNase-free DNase Kit from Promega (Promega GmbH, Mannheim, Germany). The digested RNA samples were used for expression level determination via real-

time PCR without any further treatment. Generally, additional DNase digestion was not necessary for most of the samples as the isolation and purification process worked flawless.

4.3 Functional expression of CalB in target *P. pastoris* strains

In the course of this study, two rounds of cultivation were executed in which cells were harvested at different timepoints (*figure 4.2*), centrifuged and further processed. The cell pellets were used to isolate and purify RNA, whereas the culture supernatants were analyzed for enzymatic activity to screen for functional CalB expression.

Enzymatic activity of the target protein was determined spectrophotometrically by using *p*-nitrophenyl butyrate (*p*NPB) as substrate. The assay is based on the principle that lipases catalyze the hydrolysis of triacylglycerols at the interface between water and a hydrophobic substrate. Furthermore, reactions like esterification, transesterification and interesterification of lipids are catalyzed by lipases (Soares, Facundes, Chagas Júnior, & Da Silva, 2015; Su et al., 2010). Based on this, the activity of CalB was assayed due to the release of *p*-nitrophenol during the hydrolysis of *p*-nitrophenyl butyrate using a kinetic microplate reader as described in *section 3.5.4.1*.

However, according to the cultivation strategies, cell growth and product formation were two separated processes as recombinant protein production should have been repressed during excess of glucose or glycerol, de-repressed upon their depletion and initiated through methanol induction or the limited availability of glycerol. As two cultivation processes were executed, samples were taken over different timepoints according to the data provided by the PreSens online monitoring system (see *table 3.8* and *3.9*).

4.3.1 Screening for protein activity during first round of cultivation

In cultivation round one, cells were harvested shortly after inoculation of the main cultures in the lag phase (t1); 16 h after inoculation in the log phase (t2); 30 h after inoculation, shortly after first induction/feed with methanol/glycerol feed discs, respectively (t3); after 42 h of inoculation and 24 h of first induction/feed (t4); and finally, after 65 h of inoculation, 47 h of feed and 22 h after pulsing with methanol (t5). Cultures containing BMD medium had to be pulsed again after approximately 42 h of inoculation to keep the methanol levels constant. BMG cultures were fed with glycerol feed discs after around 27 h of inoculation and thereby,

constantly providing a carbon source for the entire cultivation time without the necessity of a second feed.

Figure 4.6 shows the results of the screening assays performed on the target *P. pastoris* strains during the first round of cultivation. As visualized by the graph, functional enzymatic expression and secretion was possible by using the novel synthetic P_{CTA1} variants. Overall, no protein activity was detected when the cells were grown on high amounts of glucose or glycerol. This was presumed to the fact that an excess of glucose or glycerol in the media represses heterologous gene expression of the reporter protein CalB, which is under control of a presumably carbon source-dependent promoter. As the available carbon sources were consumed by the cells to produce biomass, the amount of glucose and glycerol decreased over time, which in turn triggered the so-called de-repression state. Thereby, recombinant protein expression could have been initiated during substrate-limited cultivation with glucose or glycerol. Theoretically, based on previously published data (Vogl et al., 2016), the P_{CTA1} and its variants should have been switched on upon carbon source depletion, to prepare the cells for the potential occurrence of hydrogen peroxide from oxidative metabolization of various alternative carbon sources (methanol degradation, ß-oxidation). Based on the known expression pattern of catalases, the de-repression sate of the cells should have led to activation of the transcription process and thereby, to heterologous gene expression. However, no protein activity was determined before induction with methanol *(figure 4.6 A)* or feed with glycerol feed discs *(figure 4.6 B)*. This result indeed could be caused by the fact that secretion of the target protein could have taken some time, which is why no protein activity could be measured. Anyway, once the promoters were turned on upon induction and feed with methanol and glycerol, respectively, CalB was functionally expressed and its production significantly increased over time.

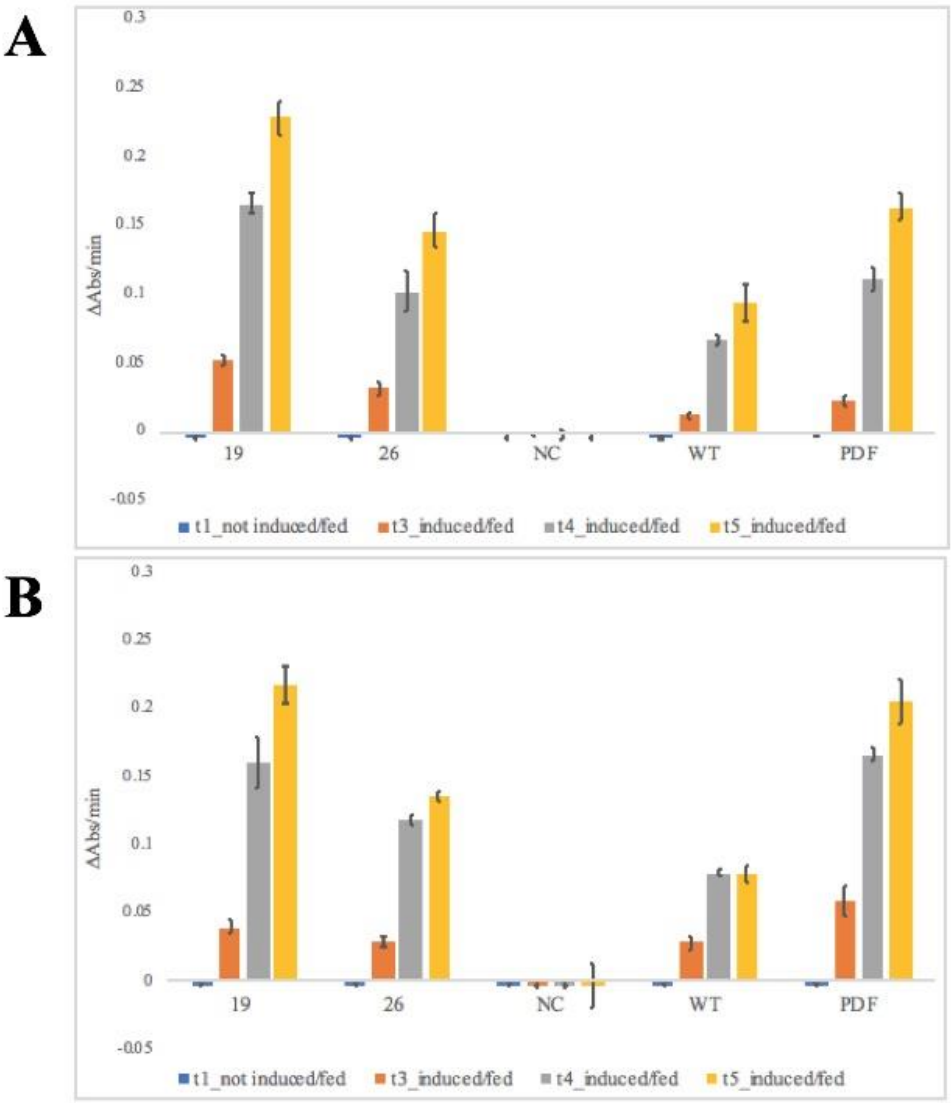

Figure 4.6 Absolute values of esterase activity determination of secreted CalB during first round of cultivation in target *P. pastoris* strains cultivated in BMD (*A*) and BMG (*B*).
Timepoints t1, t3, t4 and t5 are shown; t2 is excluded as the activities obtained resulted in negative values. Before induction (t1), heterologous gene expression was repressed caused by an excess of glucose or glycerol in the media. Here, no protein activity could be detected. After induction and feed (t3) with methanol (*A*) and glycerol feed discs (*B*), recombinant protein production was initiated and thereby, leading to measurable lipase activities for variants 19, 26, WT and P_{DF}. The negative control (NC) was not regulated by any promoter, thus no protein activity could be measured at all. Analytical polyacrylamide gel electrophoresis (SDS-PAGE) was performed for all target strains reflecting the very same results as obtained by screening for lipase activity. Moreover, Bradford assays were executed to determine total protein content of the culture supernatants resulting in 0.07-0.1 mg/mL of total protein at timepoint t4. Protein concentration before and after t4 was not measurable. To calculate these results, v_{max} values were used. In this graph, averaged values of biological duplicates are visualized.

The screening of the target *P. pastoris* strains resulted in the detection of stronger protein activities for P_{CTA1} variants 19 and 26 in BMD and BMG cultures in comparison to the parental promoter. Variant 19 cultivated in BMD resulted in almost twice as high CalB activity as the

WT strain immediately after induction and later on in the cultivation process, and even surpassed the benchmark (P_{DF}). In contrast, the negative control (plasmid encoding CalB but lacking a promoter) did not show any protein activity at all, which in turn proved that without a promoter, heterologous gene expression cannot occur. Variant 19 showed almost twice as much protein in BMG cultures compared to the WT and also, variant 26 confirmed to have enhanced lipase activities with both carbon sources, glucose and glycerol. However, for the P_{DF} higher protein activity was measured than for variant 26 in cultures containing either carbon source. Overall, variant 19 showed stronger lipase activities compared to variant 26 in both, BMD and BMG. Moreover, only for samples at time point t4, total protein content (0.07-0.1 mg/mL) was measurable. Since the esterase activity assay of secreted CalB already showed protein activity before time point t4, as well as afterwards, confirmed by the very same outcome via SDS-PAGE, it seems that the results of the Bradford assay were not reliable. Overall, the shake flask cultures with glycerol and methanol feed confirmed the results observed previously in 96 well plate experiments.

4.3.2 Analysis of lipase activity during a second round of cultivation

In cultivation round two, cells were harvested at different timepoints according to the same requirements as in the first round (see *table 3.9*). Samples were taken shortly after inoculation of the main cultures in the lag phase (t1) in media containing glucose or glycerol in 0.5% end concentration; 16 h after inoculation in the log phase (t2); 20 h after inoculation at the end of the log phase (t3), after around 22.5 h of inoculation, shortly after starting the induction phase/feed with methanol/glycerol feed discs, respectively (t4); after 44 h of inoculation and 21.5 ho after the first induction/feed (t5); after 46 h of inoculation, 23.5 h of feed and shortly after second methanol pulse (t6); and finally, after 68 h of inoculation, 49.5 h after addition of the glycerol feed discs/the first and 23.5 h after the second methanol pulse (t7).

Figure 4.7 shows the results of the enzyme assays performed on the target *P. pastoris* strains during the second round of cultivation. As can be obtained from the graph, the target protein CalB was functionally expressed and secreted by using the novel synthetic P_{CTA1} promoter variants. Overall, no lipase activity was measurable when the cells were grown on high amounts of glucose or glycerol as recombinant protein production was repressed. After carbon source depletion, a de-repression state has set, which in turn led to the activation of the promoters, as enzymatic activity was measured at timepoint t3, although induction and feed actually started

afterwards. As transcription at the promoters was initiated through induction and feed with methanol and glycerol, respectively, CalB was functionally expressed and its production significantly increased over time.

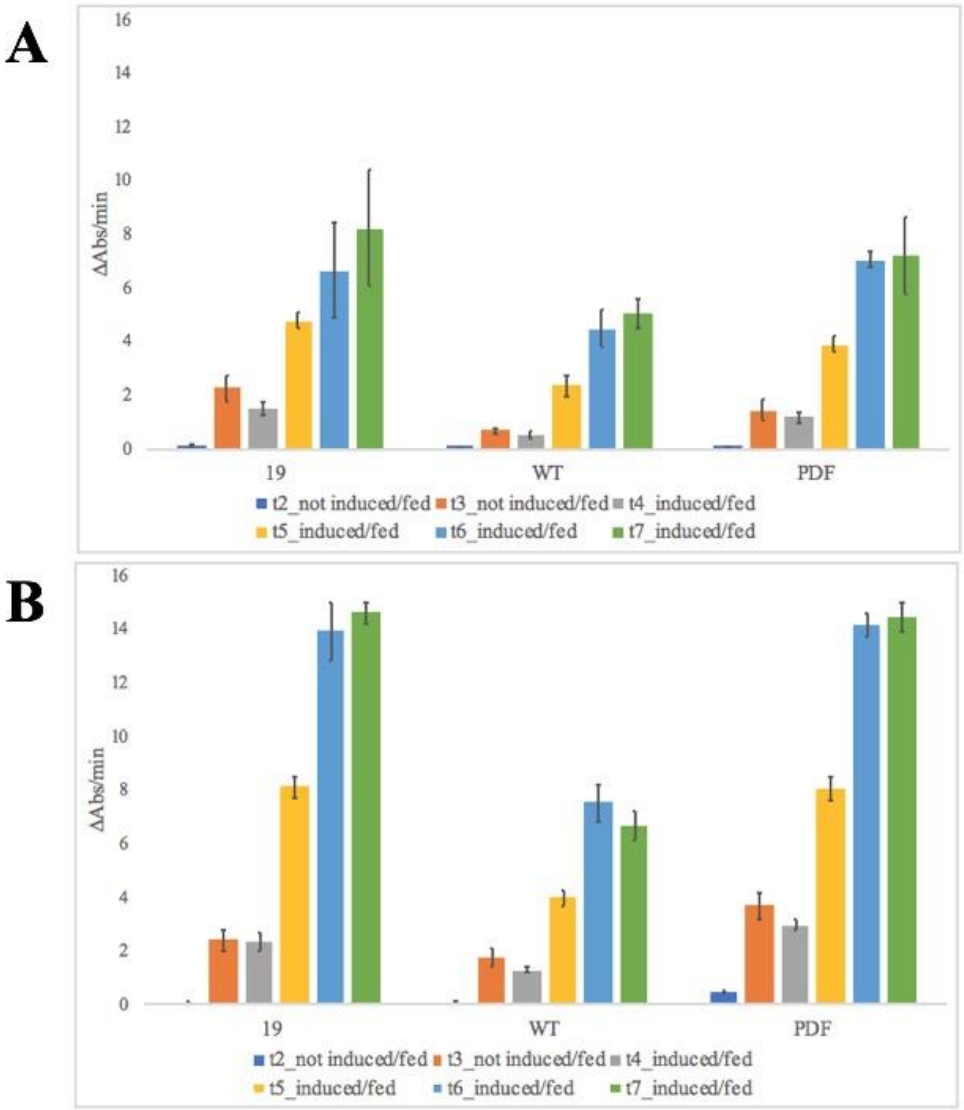

Figure 4.7 Absolute values of esterase activity determination of secreted CalB during second round of cultivation in target *P. pastoris* strains cultivated in BMD (*A*) and BMG (*B*).
Timepoints t2, t3, t4, t5, t6 and t7 are shown; t1 is excluded as the activities obtained resulted in negative values. Before induction (t2), heterologous gene expression was repressed caused by an excess of glucose or glycerol in the media. Here, no protein activity could be detected. For *A* and *B*, protein activity was already measurable in the de-repression state (t3) after carbon source depletion. Upon induction and feed (t4) with methanol (*A*) and glycerol feed discs (*B*), recombinant protein production was initiated and thereby, leading to measurable lipase activities for variant 19, WT and P_{DF}. To calculate these results, v_{mean} values were used. In this graph, averaged values of biological triplicates are shown.

The results of the lipase activity assays performed in the course of the second cultivation and validation round are shown in *figure 4.7*. As can be obtained from the graph, the target protein CalB was successfully expressed and secreted with the use of the P_{CTA1} promoter variant 19, the parental P_{CTA1} (WT) and benchmark (P_{DF}). Furthermore, lipase activity was already detectable in the de-repression state (t3), thereby leading to the assumption that the promoters were turned on as a consequence of substrate-limitation, either directly or indirectly. By comparing the cultivation conditions of variant 19, results indicated similar lipase activity for both cultivation strategies in the de-repression state (at timepoint t3). However, lipase activity seems be stronger under conditions of starvation for cells cultivated in BMG (*figure 4.7 B*). Besides, they exhibited higher protein activity over the entire cultivation process compared to those strains grown on BMD, which is highly desirable when it comes to up-scaling in bioreactors. Although induction with methanol proved to strongly initiate heterologous gene expression, the activity assays revealed that a constant supply of glycerol via feed discs may result in greater product formation as an overall higher protein activity was measured. Stronger enzymatic activities were measured for promoter variant 19 in both, BMD and BMG cultures, compared to the parental $P_{CTA1.}$ Moreover, the P_{CTA1} variant 19 seems to be less tightly regulated by the amount of carbon source present as lipase activity was already detected upon the de-repression state. Thereby, this variant could potentially offer methanol-free regulation possibility caused by the modifications in the promoter region.

4.4 Expression level determination of P_{CTA1} variants by means of qPCR

To characterize the novel synthetic P_{CTA1} variants, selected constructs were further investigated for their promoter activity by measuring the amount of gene product expressed using qPCR. During both cultivation processes, cells were harvested at different timepoints (*figure 4.2*) and RNA samples were isolated, purified and assessed for their quality and integrity (*section 3.5.3*) before finally using them for qPCR. By applying a relative quantification, changes in the expression of the target gene, relative to a reference group, were intended to examine, such as an untreated control or a control at time zero. As samples were taken over the entire cultivation processes (*figure 4.2*), it was attempted to get an overview about the differences in the expression levels under the given conditions such as growth and production phase, cultivation medium (BMD and BMG), repressed and induced state. By using the comparative $\Delta\Delta C_T$ method ($2^{-\Delta\Delta C_T}$), the relative changes in gene expression were calculated for the analyzed *P. pastoris* reporter strains.

By taking samples during the entire cultivation process and subsequent isolation of the mRNAs, the amount of transcript was determined for each time point, in order to infer the regulation pattern of the promoters and see when the promoters were switched on and whether there was already transcript produced in the de-repression state. Through comparisons of the P_{CTA1} variants with the respective wild type, it was aimed to identify whether or not the introduced mutations led to an altered transcript level. According to the performed qPCR experiments, results indicated higher transcript amounts for the promoter variants compared to the WT. However, it is difficult to assess, if the P_{CTA1} was producing more transcript as a result of enhanced promoter strength caused by the modifications, or if the synthetic sequence motifs located upstream the promoter had a stabilizing effect on the transcript itself. Besides, as gene expression is regulated on transcriptional basis and transcribed mRNA functions to be further translated into the coding protein, correlations between the mRNA expression levels and the amount of target protein were investigated

4.4.1 Validation of the $2^{-\Delta\Delta C_T}$ method

qPCR offers a great way to quantify target nucleic acids in real-time based on fluorescence technologies. In a thermocycler, the amplified products are detected as the reaction progresses through binding of intercalating dyes to the DNA strands leading to measurable fluorescent signals. Due to the sensitivity of the method, occurring errors caused by variations in the amount of starting material, need to be minimized and corrected in order to obtain precise and reliable results. To do so, the target gene gets simultaneously amplified along with an internal reference gene, HKG, to normalize RNA input amounts (Jensen, 2012). Many articles and studies recommend the use of different HKGs including GAPDH, β-actin, *ARG4*, β_2-microglobulin, rRNA (18S and 28S) and histone H3 as the most frequently used ones (Abad et al., 2010; Jensen, 2012; Krainer et al., 2012; Livak & Schmittgen, 2001; Nazari, Parham, & Maleki, 2015). Important to consider when choosing a proper HKG is that its expression level should remain constant among different tissues and unaffected by experimental conditions. Moreover, expression of target and HKG should be similar strong (Jensen, 2012). However, studies showed that experimental treatment as well as cell source indeed affect the expression levels of HKGs. Therefore, it is especially important to evaluate their stability based on the host organism (Nazari et al., 2015).

In the course of a validation experiment, different HKGs were tested and analyzed to determine if amplification of both genes (target and HKG) result in comparable efficiencies. To obtain accurate and precise quantification results, the validation experiment had to be performed prior to experimentation in order to ensure that the method of choice, $2^{-\Delta\Delta C_T}$ method, was valid or not. For valid $2^{-\Delta\Delta C_T}$ calculations, the amplification efficiencies of target and HKG must have been approximately equal. Therefore, the PCR efficiencies of both amplification reactions were calculated and determined by preparing dilution series from the specific target to look at how ΔC_T varied with template dilution (*section 3.5.5.2*). A passing validation experiment results in a slope of less than 0.1 composing of absolute values of ΔC_T versus log input according the "*Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR*" (Applied Biosystems, Inc., Foster City, CA, Unites States). The PCR efficiencies were calculated according to the formula described in *section 3.5.5.2*. Generally, through the calculation of a PCR efficiency, the performance of real-time PCR reactions can be estimated. However, the efficiency of a reaction changes with time as in the early exponential phase the PCR components are still sufficiently abundant, but towards the end, they have mostly been depleted which in turn decreases the stability of the reaction (Wong & Medrano, 2005).

Furthermore, by using the $2^{-\Delta\Delta C_T}$ method, the outcome is presented as the fold change in gene expression normalized to an internal reference gene and relative to an untreated control (calibrator). As the basis for comparative expression results, calibrators usually represent samples at time zero or without any experimental treatment. However, in this study samples taken at the beginning of cultivation (without any experimental treatment such as induction or feed) were used as calibrators.

4.4.1.1 Validation of the HKGs *ARG4* and *ACT1*

In the course of qPCR, two different HKGs were tested for the normalization of target input amounts, thus two independent validation experiments were executed. The resulting graphs (*figure 4.8 and 4.9*) represent the C_T values of each gene (y-axis) plotted against the logarithmic RNA input amounts (x-axis) from the dilution series of the specific target gene. With the resulting slope of the calibration curve, the PCR efficiency was calculated for the target and reference genes as described in *section 3.5.5.2*. To obtain valid $2^{-\Delta\Delta C_T}$ calculations, the calculated PCR efficiency of target gene (CalB) was compared with those of the HKGs (*ARG4*, *ACT1*). The more similar the PCR efficiencies of target and reference genes, the more comparable are their expression levels, which in turn is key when choosing a suitable HKG. However, as reference genes are not always constantly expressed and unaffected by experimental conditions, PCR efficiencies can vary. Therefore, it is important that the values are comparable and do not differ greatly from each other.

Figure 4.8 shows the results of the validation experiment performed in the course of cultivation round 1. To normalize RNA input amounts, the primer pairs for the HKGs *ARG4* and *ACT1* were validated. For each reference gene, different sets of primer pairs were designed, tested and analyzed in order to ensure accurate and reliable validation results. All primer sets are listed in *table 3.3.*

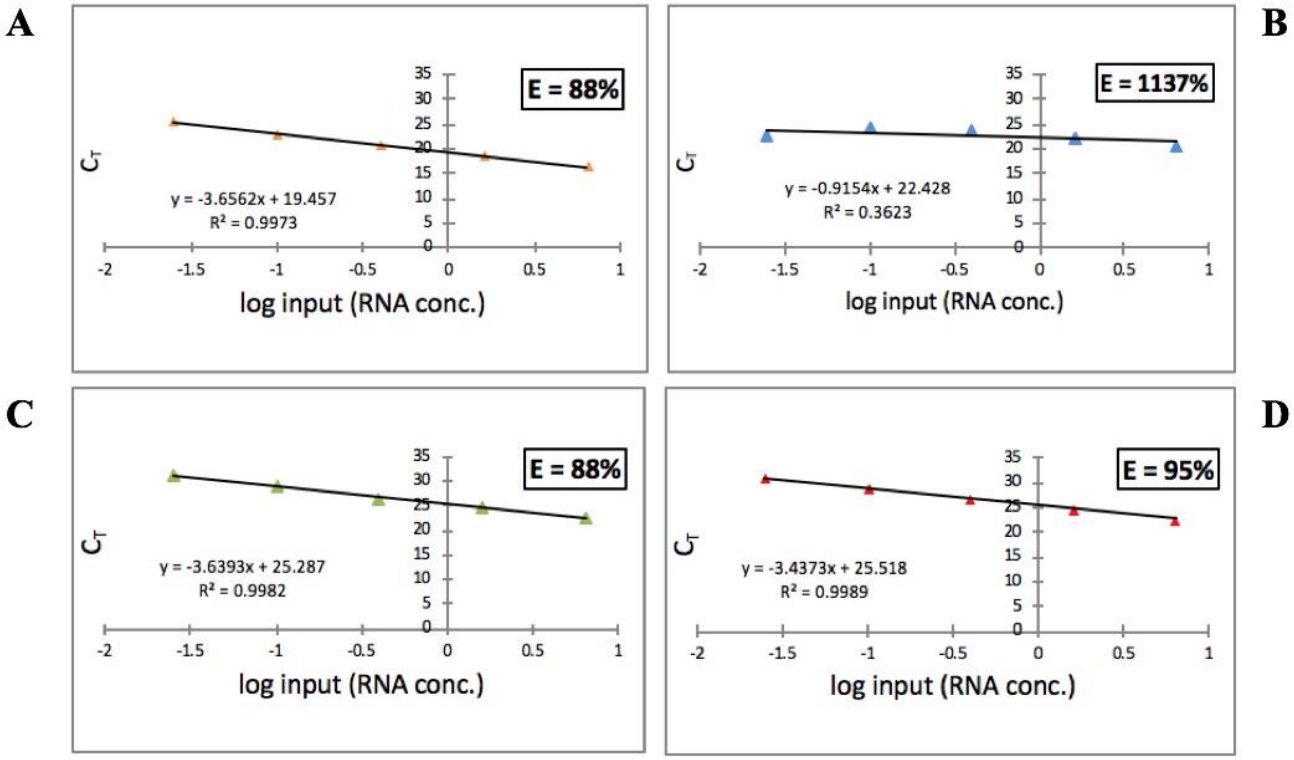

Figure 4.8 Results of the validation experiment performed in the course of cultivation round 1. In this graph, all C_T values (y-axis) are plotted against the log input amounts (x-axis) of the specific target gene (CalB). Variant 26 at timepoint t4 cultivated in BMG was used to execute the validation experiment. By using this strain, dilution series with varying target input materials were prepared to obtain a standard curve encompassing the following amounts of absolute RNAs: 6.4 ng, 1.6 ng, 0.4 ng, 0.1 ng and 0.025 ng of absolute RNA per reaction. *A* shows the C_T values of the target gene CalB (y-axis), *B* the C_T values of the HKG *ACT1*, *C* the C_T values of the HKG *ARG4*_a and *D* shows the C_T values of the HKG *ARG4*_b. The PCR efficiency E was calculated for all genes leading to the following values: 88%, 1137%, 88% and 95% from *A* to *D*, respectively. Based on these results, *ARG4*_a (*C*) was used as internal reference gene, as calculations of both, target and HKG, resulted in the similar PCR efficiencies.

As can be obtained from the graph, the PCR efficiency E of CalB was 88% (*figure 4.8 A*), same as the PCR efficiency of *ARG4*_a (*figure 4.8 C*). The efficiency of *ACT1* (*figure 4.8 B*) resulted in 1137%, which in turn indicated unspecific amplification products. Therefore, this set of primer pair was not used for any qPCR experiment. As the PCR efficiency of *ARG4*_b (*figure 4.8 D*) resulted in 95%, this pair could have been used as well, but since the efficiencies of CalB and *ARG4*_a resulted in the very same values, all qPCR experiments were executed by using these primer pairs (*ARG4*_a corresponds to *ARG4*_qPCR_fw_3 and *ARG4*_qPCR_rev_3 as can be obtained from *table 3.3*).

To determine the expression levels of the samples obtained during cultivation round 2, the primer pair of the HKG *ACT1* was validated. For this purpose, different primer pairs were specifically designed for the reference gene and analyzed in the course of a validation experiment. All primer sets are listed in *table 3.3*. *Figure 4.9* shows the results of the validation experiment performed in the course of cultivation round 2.

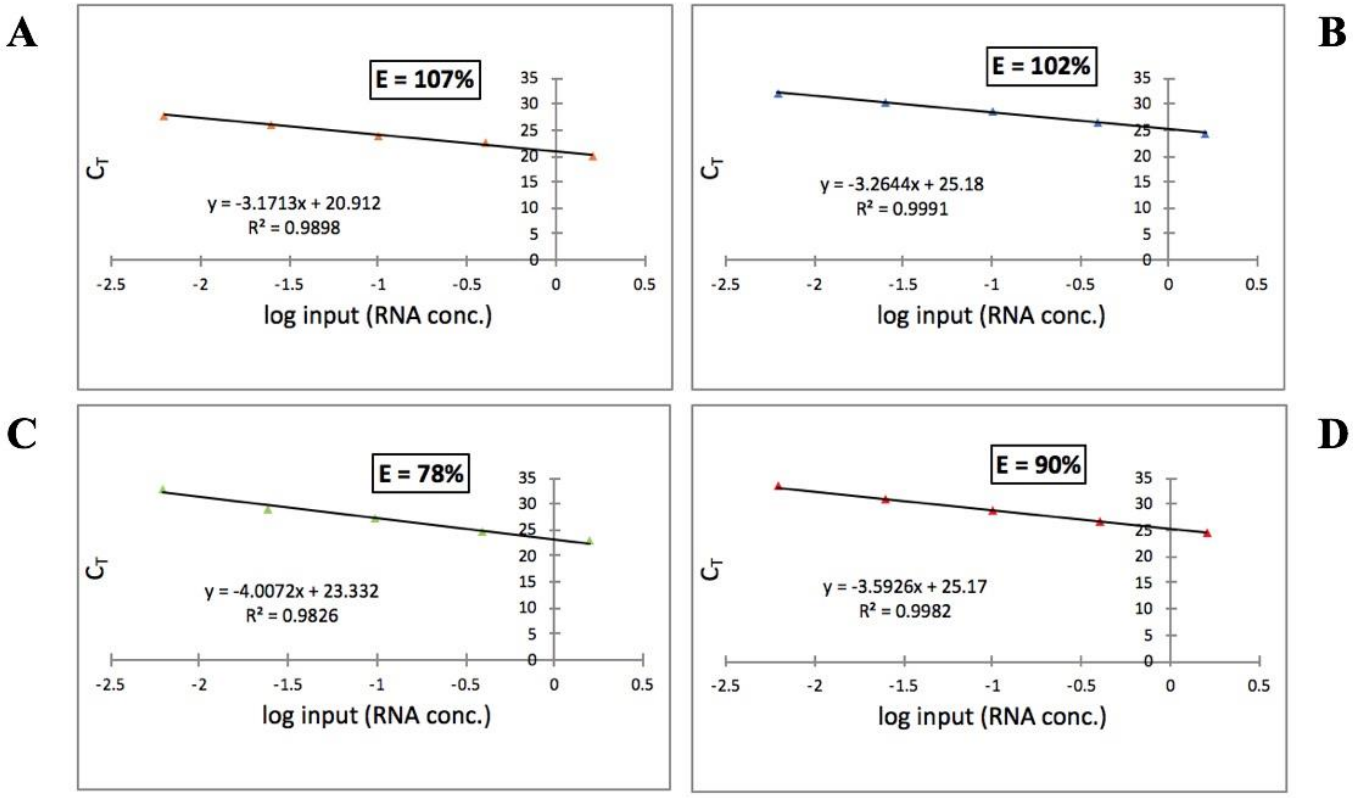

Figure 4.9 Results of the validation experiment performed in the course of cultivation round 2. In this graph, all C_T values (y-axis) are plotted against the log input amounts (x-axis) of the specific target gene (CalB). WT strain at timepoint t3 cultivated in BMD was used to execute the validation experiment. By using this strain, dilution series with varying target input materials were prepared to obtain a standard curve encompassing the following amounts of absolute RNAs: 1.6 ng, 0.4 ng, 0.1 ng, 0.025 ng and 0.00625 ng of absolute RNA per reaction. *A* shows the C_T values of the target gene CalB (y-axis), *B* the C_T values of the HKG *ACT1*_a, *C* the C_T values of the HKG *ACT1*_b and *D* shows the C_T values of the HKG *ACT1*_c. The PCR efficiency E was calculated for all genes leading to the following values: 107%, 102%, 78% and 90% from *A* to *D*, respectively. Based on these results, *ACT1*_a (*B*) was used as internal reference gene, as calculations of both, target and HKG, resulted in similar PCR efficiencies.

As can be obtained from *figure 4.9*, the PCR efficiency E of the target gene CalB resulted in 107% (*figure 4.9 A*). The primer set of the reference gene *ACT1*_a (*figure 4.9 B*) led to a similar outcome of 102%, which is why this primer pair was chosen for the execution of the qPCR

experiments. The efficiency of *ACT1*_b resulted in 78% (*figure 4.9 C*), that of *ACT1*_c in 90% (*figure 4.9 D*).

4.4.2 Transcript level determinations of the target gene CalB under the control of P_{CTA1} variants, P_{CTA1} and P_{DF}

By means of qPCR, transcript level changes were determined over time during the entire cultivation processes. To compare the novel synthetic P_{CTA1} variants, the transcripts produced by the parental promoter were furthermore analyzed as well as the benchmark P_{DF}. In a qPCR experiment, the increase of fluorescence emission, caused by the binding of intercalating dyes, can be detected during the whole quantification reaction. As only during the exponential phase the reaction provides the most accurate and precise data, qPCR only focuses on this state. Two values are calculated by the instrument within this phase, namely the "threshold line" as the level of detection at which the qPCR reaction crosses a fluorescent intensity above the background, and the "cycle threshold" (C_T) at which the fluorescent signal of the sample passes this threshold (Bustin, Benes, Nolan, & Pfaffl, 2005; Giulietti et al., 2001; Jensen, 2012).

Figure 4.10 shows an example of a typical amplification plot generated in a real-time PCR experiment along with a melt curve analysis used to evaluate amplification specificity.

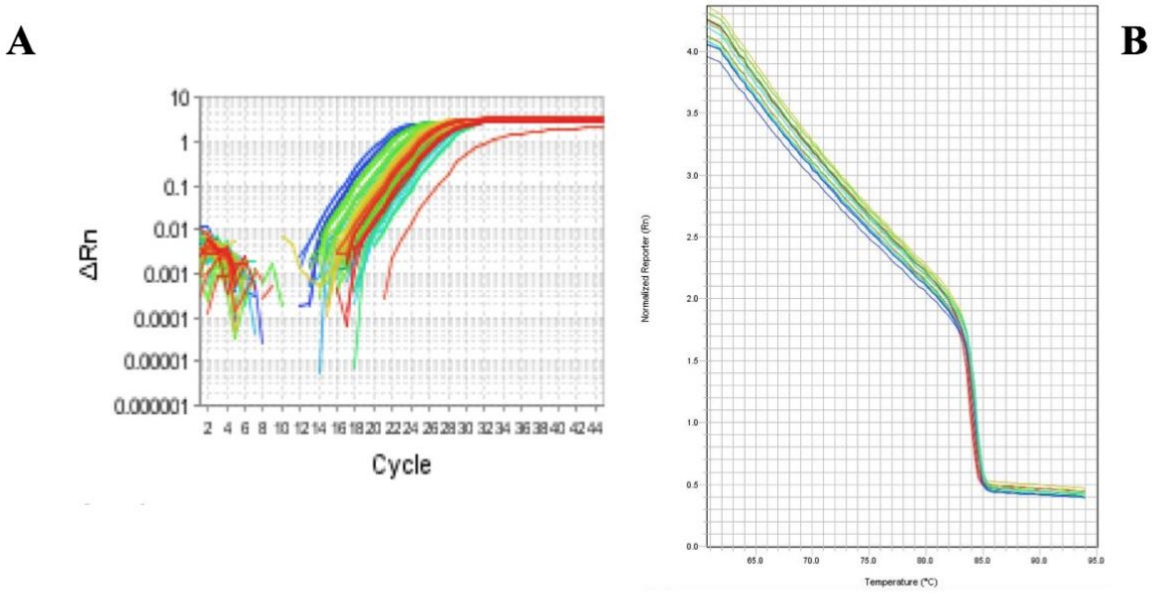

Figure 4.10 Screenshot of a typical amplification plot (*A*) and melt curve (*B*) in qPCR experiments. *A* represents an example amplification plot by using SYBR Green I for the detection of the target nucleic acids. Via reverse transcription, the mRNA strand is first reversed transcribed into its complementary DNA (cDNA) to which the fluorescent dye hybridizes and thereby, emitting a fluorescent signal which can be detected. Generally, the higher the target input amount of nucleic acid, the more fluorescent dye can bind. *B* shows the result of an example melt curve analysis which was performed to evaluate the amplification specificity. This graph represents an optimal result of a melt curve analysis. In contrast to that, unspecific products like primer dimers would appear as small peaks before or after the melting temperature of the PCR product itself (in multiple, early or delayed peaks). As SYBR Green I intercalating dyes bind unspecific to all dsDNA-molecules, melting curve analyzes were added after each cycle to overcome the problem of extenuated specificity and inaccuracy.

Melting curve analyses were added after each quantification cycle to overcome the drawbacks of extenuated specificity and inaccuracy SYBR Green I dyes bring with. However, it is crucial to include such analyses as otherwise the outputs (concentration of target nucleic acid) can be overestimated. Accordingly, detection with SYBR Green I requires extensive optimization and validation of the results, as well as the incorporation of melt curve analyzes.

4.4.3 Results of qPCR experiments executed in the course of both cultivation processes

Previous results in this study indicated that P_{CTA1} variants might have a preceding induction mechanism as we already detected protein activity upon the de-repression state. Overall, stronger enzymatic activities could be observed for the promoter variants than for the parental P_{CTA1}. Thus, enhanced variants of the P_{CTA1} at least on a protein basis had been found. To see if these results correlate with the amount of transcript, selected constructs were further investigated to find out, whether the P_{CTA1} variants led to production of more transcript compared to the parental promoter, or if they simply produced more stable mRNAs, which in turn caused an increased formation of the target protein. Alternatively, also less transcript might lead to higher enzyme activity due to a reduced cellular stress caused by high transcript levels, finally leading to better secretion, folding and measured activity in the culture supernatant.

The fold changes in the target gene, normalized to an internal reference control and relative to the expression of an untreated sample, were calculated and plotted against time as can be seen in *figure 4.11* and *4.12*

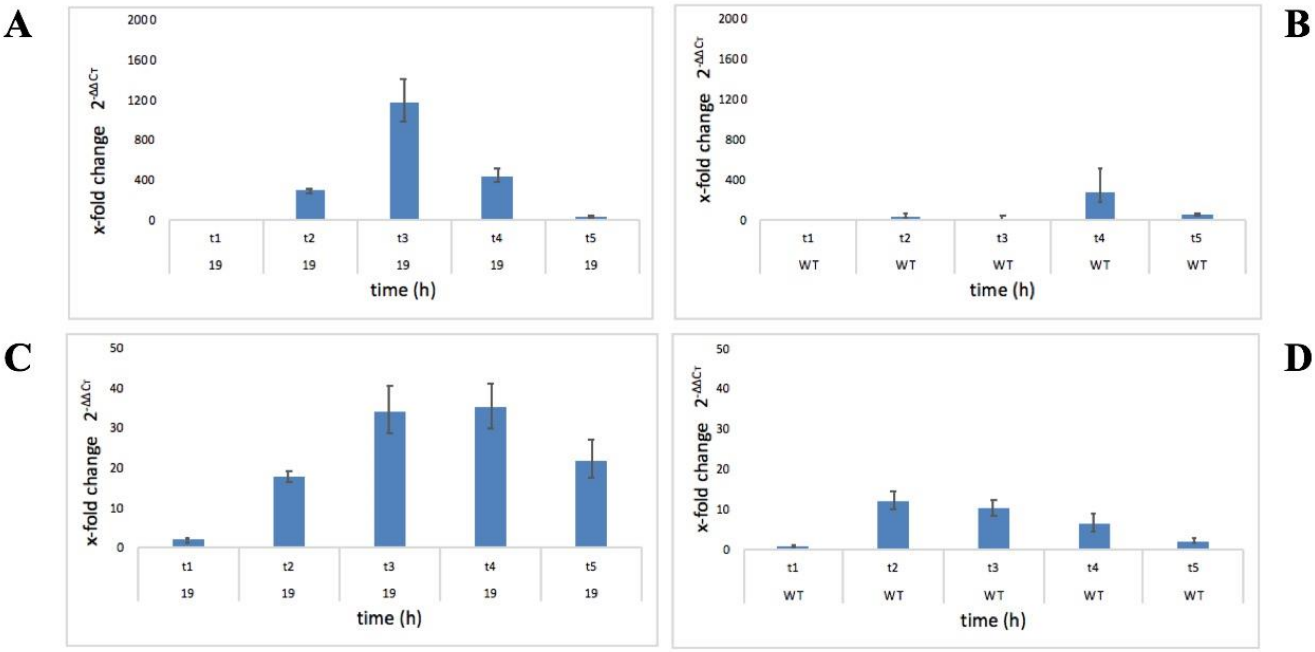

Figure 4.11 Gene expression analysis of P_{CTA1} variant 19 (*A, C*) in comparison to the P_{CTA1} promoter, WT (*B, D*) cultivated in BMD (*A, B*) and BMG (*C, D*) for cultivation round 1. X-fold changes in the target gene expression, normalized to an HKG and relative to a calibrator, were calculated.

In these graphs, the fold changes in the target gene (y-axis), normalized to an internal reference control and relative to the expression of an untreated sample, are plotted against time (x-axis). To normalize RNA input amounts, we used the validated HKG *ARG4*. Depending on the culture (BMD or BMG) and strain (19 or WT), we decided to use variant 19 or WT at time point t1 as calibrator. As can be obtained from all four diagrams, amount of transcript present increased over time, which in turn reflects the fact that cells were initially producing biomass while consuming the excess of glucose or glycerol. Once carbon sources were depleted (around timepoint t2), increased fold changes could be observed, suggesting that P_{CTA1} was activated. Through induction with methanol (*A, C*) and feed with glycerol feed discs (*B, D*), we started recombinant protein production, which in turn reflects the increase in transcript amount with timepoint t3. Averaged values of biological duplicates are shown.

Figure 4.11 shows higher transcript levels for the P_{CTA1} variant 19 (*A, C*) compared to that of the parental promoter (*B, D*). Variant 19 cultivated in BMD at timepoint t2 showed approximately 400-fold higher transcript levels than the parental promoter, thereby indicating that the P_{CTA1} variant was already switched on upon carbon source depletion (probably already in presence of some residual repressing carbon source) in the de-repression state (*figure 4.11 A* and *C*). Through the induction with methanol (*figure 4.11 A*), shortly before timepoint t3, there was a rapid increase in the transcript level up to 1200-fold changes for 19, whereas for the parental promoter no immediate changes in mRNA level could be observed (*figure 4.11 B*). On the other hand, cells cultivated in BMG (*figure 4.11 C and D*) indicated smaller fold changes for the respective timepoints. However, cells in BMG expressed the target gene more constantly over time, which in turn reflects the fact that the glycerol feed discs continuously supplied the cells with nutrients. In comparison to that, BMD cultures showed a much greater increase in gene expression due to the induction with methanol, which decreased shortly thereafter. At timepoint t3, variant 19 cultivated in BMG (*figure 4.11 C*) resulted in fold changes of approximately 35 compared to the WT with fold-changes of around 10. Overall, more constant transcriptional levels could be observed for cells cultivated in BMG than BMD, as after roughly 60 h of cultivation, they were still expressing the target gene, whereas levels for BMD cultures dropped to approximately zero.

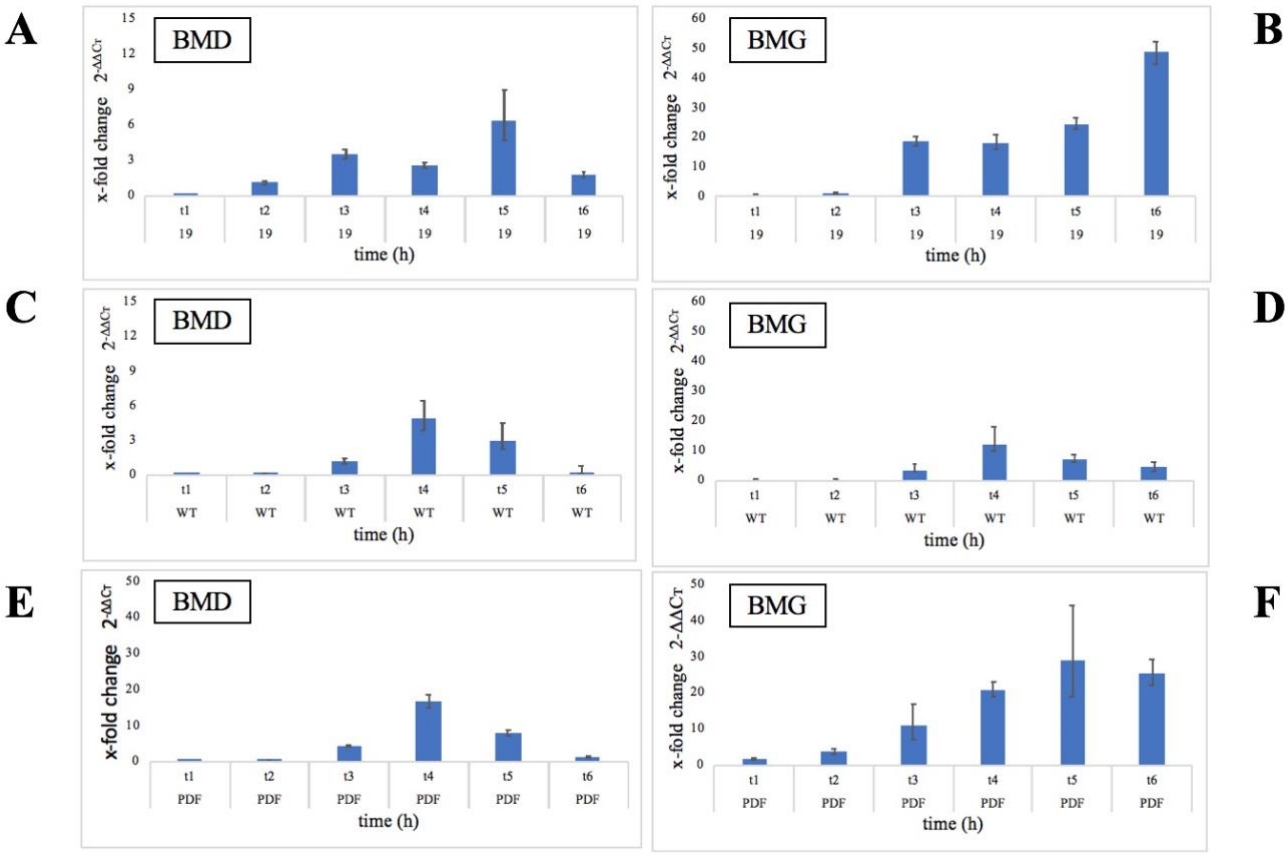

Figure 4.12 Gene expression analysis of P_{CTA1} variant 19 (*A*, *B*) in comparison to the parental P_{CTA1} (*C*, *D*) and P_{DF} promoter (*E*, *F*) cultivated in BMD (*A*, *C*, *E*) and BMG (*B*, *D*, *F*) for cultivation round 2. X-fold changes in the target gene expression, normalized to an HKG and relative to a calibrator, were calculated.

In these graphs, the fold changes in the target gene (y-axis), normalized to an internal reference control and relative to the expression of an untreated sample, are plotted against time (x-axis). To normalize target input amounts, we used the validated HKG *ACT1*. Depending on the culture (BMD or BMG), we used the P_{CTA1} variant 19 at timepoint t2 as calibrator. As can be seen in each diagram, no transcript was produced at the initial stages of cultivation as cells were primarily producing biomass while consuming the excess of glucose or glycerol. Once carbon sources were depleted, approximately around timepoint t3, we were able to observe fold changes in the transcript level of all tested promoters, indicating that gene expression was already initiated upon the de-repression state. Through induction with methanol (*A, C, E*) and feed with glycerol feed discs (*B, D, F*), we started recombinant protein production, which in turn reflects the increase in transcript amount at timepoint t4. Given values are the mean of biological triplicates.

In the course of cultivation round 2, qPCR experiments were performed for promoter variant 19 in comparison to the parental P_{CTA1} and the P_{DF} based strains. As can be seen in *figure 4.12*, higher transcript levels were achieved for the P_{CTA1} variants 19 (*A*, *B*) than for the parental promoter (*C*, *D*). Under certain experimental conditions, variant 19 even outperformed the benchmark P_{DF} (*E*, *F*). However, for variant 19 cultivated in BMD, fold changes of

approximately 3 were achieved upon carbon source depletion. In addition, P_{CTA1} variant showed higher transcript amounts in the de-repression state (*figure 4.12 A* at t3) than the parental strain (*figure 4.12 B*). The values were much lower compared to those obtained from cultivation round 1 (*figure 4.11*), probably caused by the fact that different HKGs were used for normalization of target RNA input amounts.

Furthermore, comparing cell density measurements (*figure 4.3*) of both cultivation processes, it can be seen that cell growth proceeded with similar patterns over time. Especially for BMD cultures, the strong increase in cells proceeded similarly at time points before induction and tended to continue somewhat flatter afterwards. Interestingly, variant 19 resulted in lower cell densities for cells cultivated in BMD at timepoint t3 (*figure 4.3*) compared to those of the P_{CTA1} and the P_{DF}, which may be caused by the fact, that the P_{CTA1} variant behaved differently as a response to methanol induction. Despite these circumstances, it was possible to measure higher fold changes in the transcript level of variant 19 at timepoint 3 compared to P_{CTA1} and P_{DF} based strains (*figure 4.12 A, C* and *E*). This result indicates that the differences in the amount of cells present after induction with methanol between 19, P_{CTA1} and P_{DF} do not directly reflect the levels of mRNA transcribed. Moreover, since the growth rate was lower in cells grown with methanol, it appears that the promoters are likely to have an direct related to biomass yield.

Anyway, biological samples are often difficult to compare, especially as the exact same external conditions could not be assured. In addition, different internal reference genes were used for both qPCR experiments, thereby it is difficult to say how normalizations by the respective HKG affect the total outcomes. However, the validation experiments resulted in acceptable PCR efficiencies and have met all the requirements necessary for their particular use.

On the other hand, cells cultivated in BMG (*figure 4.12 C, D* and *F*) showed higher fold changes in the transcript levels compared to those in BMD cultures (*figure 4.12 A, B* and *E*). Variant 19 achieved 20-fold transcript changes for cells fed with glycerol feed discs (*figure 4.12 C*) and showed little transcript changes in the de-repression state upon carbon source depletion compared to the WT (*figure 4.12*). Over time, it seems that cells cultivated in BMG were still expressing the target gene as results indicated transcript changes of 50-fold for timepoint t6. The benchmark P_{DF} resulted in higher fold changes for cells cultivated in BMG than BMD (*figure 4.12 E* and *F*). Anyway, the P_{CTA1} variant 19 outperformed P_{DF} in amount of transcript present after addition of glycerol feed discs at timepoint t3 and at timepoint t6 after around 70 h of cultivation.

All in all, in the course of cultivation round 2, each qPCR experiment was executed by using the HKG *ACT1* to normalize target RNA input amounts. Accordingly, results differ compared to those of cultivation round 1. Through comparisons of the fold changes in the transcript levels for selected promoter variants, values obtained are lower in cultivation round 2 than 1, especially for cells cultivated in BMD. As discussed above, maybe induction was started too early when cells were still consuming glucose, whereby recombinant gene expression was repressed instead of induced. However, as glucose should have been consumed from a certain time, gene expression should have been initiated as methanol was present in the cell cultures, thereby leading to higher transcript levels. Although changes in the target gene were measured, similar high values as for cultivation round 1 were not achieved. Nevertheless, fold-changes in the de-repression state were observed upon carbon source depletion.

5 FINAL CONCLUSION

In the course of this thesis, *P. pastoris* strains with integrated expression cassettes were cultivated in different cultivation media, namely BMD and BMG, to investigate the effects of the respective carbon sources on the regulatory behavior of the individual promoter variants. The aim of this study was to evaluate new catalase promoter variants in reliable shake flask cultures and to understand the correlations between reporter enzyme activity and transcript levels with the goal to identify a promoter with a tight regulatory mechanism that does not require methanol as inducer but still achieves high product yields. In this thesis project, two different promoter variants, namely 19 and 26, were tested in comparison to the parental P_{CTA1} and the P_{DF}, not only on a protein basis, but also at transcriptional levels by means of qPCR. This technique allowed to quantify target nucleic acids in real-time and collects the data during the entire amplification process. By using qPCR, changes in the amount of the target transcript based on the cultivation conditions of repression, de-repression and induction were examined. By comparing the P_{CTA1} variants with the parental promoter, the effects of the inserted mutations on promoter strength and regulation were evaluated on a transcription level in comparison to the posttranslational level. Therefore, mRNA levels were determined and compared with the enzymatic activities of the reporter protein.

Generally, many studies describe that cell growth and recombinant protein production can be separated in *P. pastoris* based on the strong and tightly controllable promoters of the MUT pathway. As heterologous gene expression is strictly regulated by the presence of certain carbon sources in the media, it is repressed at excess of glucose or glycerol, de-repressed upon their depletion and initiated through the induction with methanol (V. Looser et al., 2014; Verena Looser et al., 2017). Besides, it turned out that not only carbon sources themselves, but also the source of nitrogen has a significant impact on gene expression. Studies investigating the effect of different nitrogen sources on the expressional behavior of the yeast *P. pastoris* revealed, that expression of genes (*AOX1*, *AOX2* and *CTA1*) encoding key enzymes of the MUT pathway, is not entirely regulated by the carbon sources available. It has been shown that these genes are further regulated by an additional system based on the source of nitrogen (Rumjantsev, Padkina, & Sambuk, 2013). However, the use of regulatable promoters enables initial biomass gain without heterologous gene expression, thereby allowing fine-tuning of the production process and even methanol independent production processes (Prielhofer et al., 2013). As previously described, the P_{CTA1} promoter is controlled by substrate (glucose or glycerol) availability and

can be activated by simple de-repression due to carbon source depletion or limited feed (Fischer et al., 2019).

Concerning this thesis: for both P_{CTA1} variants, complete repression of product formation was achieved during cultivation with excess of glucose or glycerol, since CalB activity was not detected in the supernatants. These results were also be confirmed by the determination of the transcription levels of the target gene. However, once carbon sources were depleted, P_{CTA1} variants were switched on as reporter protein activity was measured for the de-repression state before the actual induction and feed with methanol and glycerol feed discs, respectively. On a transcriptional basis, the promoter variant 19 led to an improved promoter activity compared to the parental P_{CTA1}. Significantly more transcript was produced and transcription, regulated by promoter variant 19, was already initiated upon the de-repression, seemingly already at the end of the batch phase where carbon source levels get low. In the de-repression state, upon carbon source depletion but without induction or feed with methanol or glycerol feed discs, respectively, changes in the transcript level of up to 4-fold were measured between time point t4 and start of cultivation for cells cultured in BMD and up to 20-fold changes between the very same time points for cells cultivated in BMG for cultivation round 2. Similar fold-changes were determined for the same state of cultivation for variant 19 cultivated in BMG in round 1, under carbon source depletion and without feed of glycerol feed discs. However, higher fold-changes in the transcript level were observed for cells cultivated in BMD in the very same cultivation process. In general, higher fold-changes were observed in the transcript level for cells cultivated in BMD for cultivation round 1 compared to the fold-changes determined for the cells cultivated in BMD in the second cultivation process. Since different HKGs were used for normalization in both cultivation processes (for cultivation round 1, the HKG *ARG*4; for cultivation round 2 the HKG *ACT*1 was used), it seems that the use of varying HKGs strongly affected the final results. Usually, HKGs are anticipated to be unaffected by external conditions and expressed at constant levels (Løvdal & Lillo, 2009). In the course of validation experiments, the primer sets of the HKGs were tested against the primer pair of the reporter gene CalB, in order to calculate how similar, and therefore comparable, their PCR efficiencies were. Anyway, only the primer pairs were validated and not the HKGs itself, which is why no conclusions can be drawn about their actual expression behavior in the respective *P. pastoris* strains. The data of this diploma thesis indicated that the HKGs were expressed at different levels as the fold-changes in the transcript levels of the target gene led to different results between the used HKGs. The *P. pastoris* strain employing the P_{CTA1} variant 19 cultivated in BMD in cultivation round 2, resulted in a difference of 4-fold for the transcriptional level between the sample shortly taken after

cultivation start (time point t1) and that after 22.5 h of inoculation shortly after induction with methanol (time point t4). In cultivation round 1, the very same *P. pastoris* strain cultivated in BMD showed a fold-change of 1200 in the transcription level between the sample taken shortly after cultivation start (time point t1) and that after approximately 30 h of cultivation shortly after induction with methanol (time point t3). Since both cultivation processes were performed under the same requirements and it could be demonstrated that cellular growth behavior was comparable, it can be concluded that the use of different HKGs to normalize RNA input amounts strongly effects the outcome. As for cultivation round 1 (in BMD), the samples were normalized to the HKG *ARG*4 and overall higher fold-changes in the transcript levels were observed between the respective time points of sampling and the cultivation start (time point t1), compared to the samples in cultivation round 2 which were normalized to *ACT*1 resulting in lower fold-changes for the transcript levels between the very same time points. For BMG cultures, similar values were determined for both cultivation processes regarding the fold-changes in the transcript levels between the respective time points and cultivation start, although different HKGs were used. Based on this fact, it seemed that the choice of carbon source strongly influenced expression of the certain HKG and thus, led to different outcomes. Anyway, as the main focus of this study was to show whether the P_{CTA1} variants provoked any changes in expression compared to the parental P_{CTA1}, i.e. higher transcript levels, the absolute values were negligible. In fact, the main focus was to evidence, regardless of the HKG, that the trend remained the same over both cultivation processes. According to the results obtained in this study, it was possible to demonstrate that no matter which HKG was used for normalization, the P_{CTA1} variant yielded in higher transcript amounts at the respective time points than the strain employing the parental P_{CTA1} promoter.

In the experiments of this thesis, employing a Mut^S strain background, cells cultivated in BMG were producing more biomass after feed with glycerol feed discs compared to those cultured in BMD in the presence of methanol. However, a stronger change in biomass between growth and production phase was visible for cells in BMD by induction with methanol. The performed epxeriments also confirmed, that P_{CTA1}, as well as the tested variants, was still inducible with methanol. As expected for the used Mut^S strains, only minor increases in biomass formation were observed when using methanol as a sole carbon source during the induction of cells after prior growth on glucose. According to the importance of catalases regarding cellular viability, compontents such as methanol, fatty acids, uric acids or amines lead to their specifc induction (Rußmayer et al., 2015; Verduyn, Giuseppin, Scheffers, & Van Dijken, 1988). The analysis performed with the new promoter variants confirmed that at least induction by methanol was

still retained. Therefore, future studies evaluating the use of such components for alternative induction processes might be intersting also for the promoter variants studied in the course of this thesis.

An important aim of this thesis was to study possible positive or negative correlations between mRNA levels and reporter enzyme activates resulting from the individual promoter variants. For P_{CTA1} variant 19, changes in the transcript level of approximately 4-fold were determined between the sample shortly taken after cultivation start (time point t1) and that after 22.5 h of cultivation shortly after induction with methanol (time point t4) in BMD, whereas for BMG, 20-fold changes were observed between the very same time points. However, according to the enzymatic activities measured, the amount of protein present at the particular timepoints did not strongly differentiate between the respective cultivation media, despite the high differences on a transcriptional basis. This could be caused by the fact, that the amount of mRNA transcribed does not directly reflect the amount of protein produced as intracellular processes, such as induction of gene expression, translation, protein folding and posttranslational processing, translocation, secretion etc., do significantly influence product formation (Borbolis & Syntichaki, 2015; Verena Looser et al., 2017). The P_{CTA1} variant 19 showed changes in the transcript level of 7-fold after 21.5 h of induction with methanol (BMD, time point t5) and approximately 2-fold changes shortly after pulse with methanol (BMD, time point t6). Compared to this, variant 19 showed fold-changes in the transcript level of approximately 30, compared to the cultivation start, after 44 h of cultivation and 21.5 h after feed with glycerol feed discs (time point t5) in BMG, and even higher changes were observed (50-fold) after around 46 h of cultivation and 23.5 h of feed (time point t6, BMG). These results revealed that the cells cultivated in BMD needed some time to adjust to the changing nutritional state, namely the addition of methanol after glucose depletion, as higher transcript changes were determined after a longer period of induction with methanol compared to the time points immediately taken after induction/pulse (time point t4 and t6). Based on these results, it is assumed that cell physiology had changed upon the induction/pulse with methanol, which in turn caused the target gene expression to be shifted back in time and to start only after adaption to the new conditions. However, as cells cultivated in BMG were fed with glycerol feed discs after glycerol depletion, higher changes in the transcript levels were observed at the respective time points as the same carbon source was kept for heterologous gene expression.

Generally, it is well known that there are many factors existing that affect mRNA-to-protein ratio. Indeed, scientists assume that there are three potential reasons for the lack of this relation,

namely the translation regulation, differences in protein *in vivo* half-lives and the significant amount of errors occurring due to variations in the experimental conditions. Studies investigating the contributions of different biochemical and physical sources to the mRNA-protein correlation proved, that variations in consequence of analytical handling can account for up to 34-44% of the total variation of mRNA-to protein ratio (Nie, Wu, & Zhang, 2006). In respect of this study, P_{CTA1} variant 19 showed higher transcriptional changes in the de-repression state than the parental promoter under both cultivation conditions – BMD and BMG. Furthermore, higher protein activities were measured for the promoter variants than for the parental P_{CTA1}, not only after carbon source depletion, but also after induction and feed with methanol and glycerol feed discs. Overall, the shake flask evaluations and qPCR data confirmed that the novel synthetic P_{CTA1} variants led to stronger results under conditions of de-repression and induction/feed with methanol/glycerol feed discs, respectively, compared to the parental P_{CTA1}, on both transcription level and protein basis.

To conclude, as P_{CTA1} is only expressed after carbon source depletion, it might be initiated as a response to catabolite starvation or a switch to alternative carbon sources, e.g. stored lipids which get used by β-oxidation, which is known to produce hydrogen peroxide, causing a need for catalase expression and activity. In theory, P_{CTA1} regulated expression might also be influenced by the source of nitrogen present as suggested by Rumjantsev *et al*., (2013) but no studies were performed in this thesis to evaluate such possible effects on the novel promoter variants 19 and 26. Furthermore, increased promoter activity could be caused by oxidative stress based on the creation of ROS, reactive oxygen species (Vogl & Glieder, 2013). It is still not clear what regulatory mechanism P_{CTA1} is subjected to, therefore a rational approach to create more active promoter variants is currently not possible. However, the results of this study suggest that for glucose or glycerol de-repression, the lack of transcription factors or their respective binding sites on the promoter could potentially lead to the initiation of transcription and thus, induce gene expression.

6 BIBLIOGRAPHY

Abad, S., Kitz, K., Hörmann, A., Schreiner, U., Hartner, F. S., & Glieder, A. (2010). Real-time PCR-based determination of gene copy numbers in Pichia pastoris. *Biotechnology Journal*. https://doi.org/10.1002/biot.200900233

Ahmad, M., Hirz, M., Pichler, H., & Schwab, H. (2014). Protein expression in Pichia pastoris: Recent achievements and perspectives for heterologous protein production. *Applied Microbiology and Biotechnology*. https://doi.org/10.1007/s00253-014-5732-5

Anamika, K., Verma, S., Jere, A., & Desai, A. (2016). Transcriptomic Profiling Using Next Generation Sequencing - Advances, Advantages, and Challenges. In *Next Generation Sequencing - Advances, Applications and Challenges*. https://doi.org/10.5772/61789

Asada, H., Uemura, T., Yurugi-Kobayashi, T., Shiroishi, M., Shimamura, T., Tsujimoto, H., … Kobayashi, T. (2011). Evaluation of the Pichia pastoris expression system for the production of GPCRs for structural analysis. *Microbial Cell Factories*. https://doi.org/10.1186/1475-2859-10-24

Blazeck, J., Alper, H. S., & Nagamune, T. (2013). Promoter engineering: Recent advances in controlling transcription at the most fundamental level. *Biotechnology Journal*. https://doi.org/10.1002/biot.201200025

Borbolis, F., & Syntichaki, P. (2015). Cytoplasmic mRNA turnover and ageing. *Mechanisms of Ageing and Development*. https://doi.org/10.1016/j.mad.2015.09.006

Bustin, S. A., Benes, V., Nolan, T., & Pfaffl, M. W. (2005). Quantitative real-time RT-PCR – a perspective. *Journal of Molecular Endocrinology*. https://doi.org/10.1677/jme.1.01755

Bustin, S. A., & Pfaffl, M. W. (n.d.). *A-Z of quantitative PCR (Editor: Markers of a Successful Real-Time RT-PCR Assay 2 3.2.1. RNA Extraction 2 3.2.2. Reverse Transcription 3 3.2.3. Comparison of Real-Time RT-PCR with Classical End-Point Detection Method 4 3.2.4. Chemistry Developments for Real-Time*.

Cereghino, G. P. L., & Cregg, J. M. (1999). *Applications of yeast in biotechnology: protein production and genetic analysis Lin Cereghino and Cregg 423*.

Cereghino, J. L., & Cregg, J. M. (2000). Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiology Reviews*. https://doi.org/10.1016/S0168-6445(99)00029-7

Cregg, J. M., Madden, K. R., Barringer, K. J., Thill, G. P., & Stillman, C. A. (1989). Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. *Molecular and Cellular Biology*. https://doi.org/10.1128/MCB.9.3.1316

Engstrom, M. D., & Pfleger, B. F. (2017). Transcription control engineering and applications in synthetic biology. *Synthetic and Systems Biotechnology*. https://doi.org/10.1016/j.synbio.2017.09.003

Fischer, J. E., Hatzl, A.-M., Weninger, A., Schmid, C., & Glieder, A. (2019). Methanol Independent Expression by Pichia Pastoris Employing De-repression Technologies. *Journal of Visualized Experiments*, (143). https://doi.org/10.3791/58589

Gellissen, G., & Hollenberg, C. P. (1997). *GENE Application of yeasts in gene expression studies: a comparison of Saccharomyces cerevisiae, Hansenula polymorpha and Kluyveromyces lactis-a review 1*. *Gene* (Vol. 190).

Ginzinger, D. G. (2002). *Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream*. *Experimental Hematology* (Vol. 30).

Giulietti, A., Overbergh, L., Valckx, D., Decallonne, B., Bouillon, R., & Mathieu, C. (2001). An overview of real-time quantitative PCR: Applications to quantify cytokine gene expression. *Methods*. https://doi.org/10.1006/meth.2001.1261

Gong, H., Sun, L., Chen, B., Han, Y., Pang, J., Wu, W., … Zhang, T. (2016). Evaluation of candidate reference genes for RT-qPCR studies in three metabolism related tissues of mice after caloric restriction. *Scientific Reports*. https://doi.org/10.1038/srep38513

Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/s41580-018-0028-8

Hartner, F. S., & Glieder, A. (2006). Regulation of methanol utilisation pathway genes in yeasts. *Microbial Cell Factories*. https://doi.org/10.1186/1475-2859-5-39

Horiguchi, H., Yurimoto, H., Goh, T. K., Nakagawa, T., Kato, N., & Sakai, Y. (2001).

Peroxisomal catalase in the methylotrophic yeast Candida boidinii: Transport efficiency and metabolic significance. *Journal of Bacteriology*. https://doi.org/10.1128/JB.183.21.6372-6383.2001

Jensen, E. C. (2012). Real-time reverse transcription polymerase chain reaction to measure mRNA: Use, limitations, and presentation of results. *Anatomical Record*. https://doi.org/10.1002/ar.21487

Klose, C., Surma, M. A., Gerl, M. J., Meyenhofer, F., Shevchenko, A., & Simons, K. (2012). Flexibility of a eukaryotic lipidome - insights from yeast lipidomics. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0035063

Koleva, D., Petrova, V., Hristozova, T., & Kujumdzieva, A. (2008). Study of catalase enzyme in methylotrophic yeasts. *Biotechnology and Biotechnological Equipment*. https://doi.org/10.1080/13102818.2008.10817548

Koutz, P., Davis, G. R., Stillman, C., Barringer, K., Cregg, J., & Thill, G. (1989). Structural comparison of the Pichia pastoris alcohol oxidase genes. *Yeast*. https://doi.org/10.1002/yea.320050306

Kozera, B., & Rapacz, M. (2013). Reference genes in real-time PCR. *Journal of Applied Genetics*. https://doi.org/10.1007/s13353-013-0173-x

Krainer, F. W., Dietzsch, C., Hajek, T., Herwig, C., Spadiut, O., & Glieder, A. (2012). Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway. *Microbial Cell Factories*. https://doi.org/10.1186/1475-2859-11-22

Kurtzman, C. P. (2009). Biotechnological strains of Komagataella (Pichia) pastoris are Komagataella phaffii as determined from multigene sequence analysis. *Journal of Industrial Microbiology and Biotechnology*, *36*(11), 1435–1438. https://doi.org/10.1007/s10295-009-0638-4

Lee, P. Y., Costumbrado, J., Hsu, C.-Y., & Kim, Y. H. (2012). Agarose Gel Electrophoresis for the Separation of DNA Fragments. *Journal of Visualized Experiments*. https://doi.org/10.3791/3923

Liang, S., Zou, C., Lin, Y., Zhang, X., & Ye, Y. (2013). Identification and characterization of PGCW14: A novel, strong constitutive promoter of Pichia pastoris. *Biotechnology*

Letters, *35*(11), 1865–1871. https://doi.org/10.1007/s10529-013-1265-8

Livak, K. J., & Schmittgen, T. D. (2001). Analyzes of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2-ΔΔCT Method. *Analyzes of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2-ΔΔCT Method*. https://doi.org/10.1006/meth.2001.1262

Looser, V., Bruhlmann, B., Bumbak, F., Stenger, C., Costa, M., Camattari, A., … Kovar, K. (2014). Cultivation strategies to enhance productivity of Pichia pastoris: A review. *Biotechnology Advances*. https://doi.org/10.1016/j.biotechadv.2015.05.008

Looser, Verena, Lüthy, D., Straumann, M., Hecht, K., Melzoch, K., & Kovar, K. (2017). Effects of glycerol supply and specific growth rate on methanol-free production of CALB by P. pastoris: functional characterisation of a novel promoter. *Applied Microbiology and Biotechnology*. https://doi.org/10.1007/s00253-017-8123-x

Løvdal, T., & Lillo, C. (2009). Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress. *Analytical Biochemistry*. https://doi.org/10.1016/j.ab.2009.01.024

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1005457

Lubliner, S., Regev, I., Lotan-Pompan, M., Edelheit, S., Weinberger, A., & Segal, E. (2015). Core promoter sequence in yeast is a major determinant of expression level. *Genome Research*, *25*(7), 1008–1017. https://doi.org/10.1101/gr.188193.114

Macauley-Patrick, S., Fazenda, M. L., McNeil, B., & Harvey, L. M. (2005). Heterologous protein production using the Pichia pastoris expression system. *Yeast*. https://doi.org/10.1002/yea.1208

Malak, A., Baronian, K., & Kunze, G. (2016). Blastobotrys (Arxula) adeninivorans: a promising alternative yeast for biotechnology and basic research. *Yeast*, *33*(10), 535–547. https://doi.org/10.1002/yea.3180

Nakagawa, T., Mizumura, T., Mukaiyama, H., Miyaji, T., Yurimoto, H., Kato, N., … Tomizuka, N. (2002). Physiological role of the second alcohol oxidase gene MOD2 in

the methylotrophic growth of Pichia methanolica. *Yeast*, *19*(12), 1067–1073. https://doi.org/10.1002/yea.896

Nakagawa, T., Sakai, Y., Mukaiyama, H., Mizumura, T., Miyaji, T., Yurimoto, H., … Tomizuka, N. (2001). Analysis of alcohol oxidase isozymes in gene-disrupted strains of methylotrophic yeast Pichia methanolica. *Journal of Bioscience and Bioengineering*, *91*(2), 225–227. https://doi.org/10.1016/S1389-1723(01)80071-2

Nakagawa, T., Yamada, K., Fujimura, S., Ito, T., Miyaji, T., & Tomizuka, N. (2005). Pectin utilization by the methylotrophic yeast Pichia methanolica. *Microbiology*, *151*(6), 2047–2052. https://doi.org/10.1099/mic.0.27895-0

Nakagawa, T., Yoshida, K., Takeuchi, A., Ito, T., Fujimura, S., Matsufuji, Y., … Hayakawa, T. (2010). The Peroxisomal Catalase Gene in the Methylotrophic Yeast Pichia methanolica. https://doi.org/10.1271/bbb.100329

Nazari, F., Parham, A., & Maleki, A. F. (2015). GAPDH, β-actin and β2-microglobulin, as three common reference genes, are not reliable for gene expression studies in equine adipose- and marrow-derived mesenchymal stem cells. *Journal of Animal Science and Technology*. https://doi.org/10.1186/s40781-015-0050-8

Nie, L., Wu, G., & Zhang, W. (2006). Correlation of mRNA expression and protein abundance affected by multiple sequence features related to translational efficiency in Desulfovibrio vulgaris: A quantitative analysis. *Genetics*. https://doi.org/10.1534/genetics.106.065862

Portela, R. M. C., Vogl, T., Ebner, K., Oliveira, R., & Glieder, A. (2018). Pichia pastoris Alcohol Oxidase 1 (AOX1) Core Promoter Engineering by High Resolution Systematic Mutagenesis. *Biotechnology Journal*. https://doi.org/10.1002/biot.201700340

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

Rebnegger, C., Graf, A. B., Valli, M., Steiger, M. G., Gasser, B., Maurer, M., & Mattanovich, D. (2014). In Pichia pastoris, growth rate regulates protein synthesis and secretion, mating and stress response. *Biotechnology Journal*.

https://doi.org/10.1002/biot.201300334

Rumjantsev, A. M., Padkina, M. V., & Sambuk, E. V. (2013). Effect of nitrogen source on gene expression of first steps of methanol utilization pathway in Pichia pastoris. *Russian Journal of Genetics*. https://doi.org/10.1134/S102279541304011X

Rußmayer, H., Buchetics, M., Gruber, C., Valli, M., Grillitsch, K., Modarres, G., … Gasser, B. (2015). Systems-level organization of yeast methylotrophic lifestyle. *BMC Biology*, *13*(1). https://doi.org/10.1186/s12915-015-0186-5

Sakai, Y., Yurimoto, H., Matsuo, H., & Kato, N. (1998). Regulation of peroxisomal proteins and organelle proliferation by multiple carbon sources in the methylotrophic yeast, Candida boidinii. *Yeast*, *14*(13), 1175–1187. https://doi.org/10.1002/(SICI)1097-0061(19980930)14:13<1175::AID-YEA319>3.0.CO;2-7

See Kobs, G. (1998). isolation-of-rna-from-plant-yeast-and-bacteria.

Sekar, R. B., & Periasamy, A. (2003). Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *Journal of Cell Biology*. https://doi.org/10.1083/jcb.200210140

Sinha, P., Singh, V. K., Suryanarayana, V., Krishnamurthy, L., Saxena, R. K., & Varshney, R. K. (2015). Evaluation and validation of housekeeping genes as reference for gene expression studies in pigeonpea (cajanus cajan) under drought stress conditions. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0122847

Soares, M. de K. G., Facundes, B. C., Chagas Júnior, A. F., & Da Silva, E. M. (2015). Assessment of lipolytic activity of isolated microorganisms from the savannah of the Tocantins. *Acta Scientiarum - Biological Sciences*. https://doi.org/10.4025/actascibiolsci.v37i4.24324

Su, G. D., Huang, D. F., Han, S. Y., Zheng, S. P., & Lin, Y. (2010). Display of Candida antarctica lipase B on Pichia pastoris and its application to flavor ester synthesis. *Applied Microbiology and Biotechnology*. https://doi.org/10.1007/s00253-009-2382-0

Subramaniyan, S., Alugoju, P., SJ, S., Veerabhadrappa, B., & Dyavaiah, M. (2019). Magnolol protectsSaccharomyces cerevisiae antioxidant-deficient mutants from oxidative stress and extends yeast chronological life span. *FEMS Microbiology Letters*.

https://doi.org/10.1093/femsle/fnz065

Teste, M. A., Duquenne, M., François, J. M., & Parrou, J. L. (2009). Validation of reference genes for quantitative expression analysis by real-time RT-PCR in Saccharomyces cerevisiae. *BMC Molecular Biology*. https://doi.org/10.1186/1471-2199-10-99

Thelwell, N., Brown, Millington, S., Solinas, A., Booth, J., & Brown, T. (2000). Mode of action and application of Scorpion primers to mutation detection. *Astronomical Journal*. https://doi.org/10.1086/300550

Tyagi, S., & Kramer, F. R. (1995). Molecular Beacons: Probes that Fluoresce upon Hybridization.

Valli, M., Tatto, N. E., Peymann, A., Gruber, C., Landes, N., Ekker, H., … Graf, A. B. (2016). Curation of the genome annotation of Pichia pastoris (Komagataella phaffii) CBS7435 from gene level to protein function. *FEMS Yeast Research*, *16*(6). https://doi.org/10.1093/femsyr/fow051

Verduyn, C., Giuseppin, M. L. F., Scheffers, W. A., & Van Dijken, J. P. (1988). Hydrogen peroxide metabolism in yeasts. *Applied and Environmental Microbiology*.

Vogl, T., & Glieder, A. (2013). Regulation of Pichia pastoris promoters and its consequences for protein production. *New Biotechnology*. https://doi.org/10.1016/j.nbt.2012.11.010

Vogl, T., Sturmberger, L., Fauland, P. C., Hyden, P., Fischer, J. E., Schmid, C., … Glieder, A. (2018). Methanol independent induction in Pichia pastoris by simple derepressed overexpression of single transcription factors. *Biotechnology and Bioengineering*, *115*(4), 1037–1050. https://doi.org/10.1002/bit.26529

Vogl, T., Sturmberger, L., Kickenweiz, T., Wasmayer, R., Schmid, C., Hatzl, A. M., … Glieder, A. (2016). A Toolbox of Diverse Promoters Related to Methanol Utilization: Functionally Verified Parts for Heterologous Pathway Expression in Pichia pastoris. *ACS Synthetic Biology*. https://doi.org/10.1021/acssynbio.5b00199

Wolf, J. B. W. (2013). Principles of transcriptome analysis and gene expression quantification: An RNA-seq tutorial. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.12109

Wong, M. L., & Medrano, J. F. (2005). Real-time PCR for mRNA quantitation.

BioTechniques. https://doi.org/10.2144/05391RV01

Yella, V. R., & Bansal, M. (2017). DNA structural features of eukaryotic TATA-containing and TATA-less promoters. *FEBS Open Bio*, *7*(3), 324–334. https://doi.org/10.1002/2211-5463.12166

Zhang, H., Loovers, H. M., Xu, L. Q., Wang, M., Rowling, P. J. E., Itzhaki, L. S., … Perrett, S. (2009). Alcohol oxidase (AOX1) from Pichia pastoris is a novel inhibitor of prion propagation and a potential ATPase. *Molecular Microbiology*, *71*(3), 702–716. https://doi.org/10.1111/j.1365-2958.2008.06557.x

Zhang, N., Suen, W.-C., Windsor, W., Xiao, L., Madison, V., & Zaks, A. (2003). Improving tolerance of Candida antarctica lipase B towards irreversible thermal inactivation through directed evolution. *Protein Engineering Design and Selection*. https://doi.org/10.1093/protein/gzg074

7 LIST OF FIGURES

8 LIST OF TABLES

9 APPENDIX

9.1 Expression vector constructs

Table 9.1 List of plasmid maps and their corresponding SnapGene file names.

Plasmid	Map	Nr. of figure	Name of figure	Corresponding SnapGene file name
1.a	pKEB#1-PCTA1_variant_19	3.1	Expression construct of P_{CTA1} promoter variant 19	P_{CTA1}_variant_19
1.b	pKEB#1-PCTA1_variant_26	3.2	Expression construct of P_{CTA1} promoter variant 26	P_{CTA1}_variant_26
2.0	pKEB#1-PCTA500_WT	9.1	Expression construct of WT CTA500	P_{CTA1}_WT
3.0	pKEB#1-PDF	9.2	Expression construct of the benchmark PDF	P_{CTA1}_P_{DF}
4.0	pKEB#1-NC	9.3	Expression construct of NC (negative control)	P_{CTA1}_NC

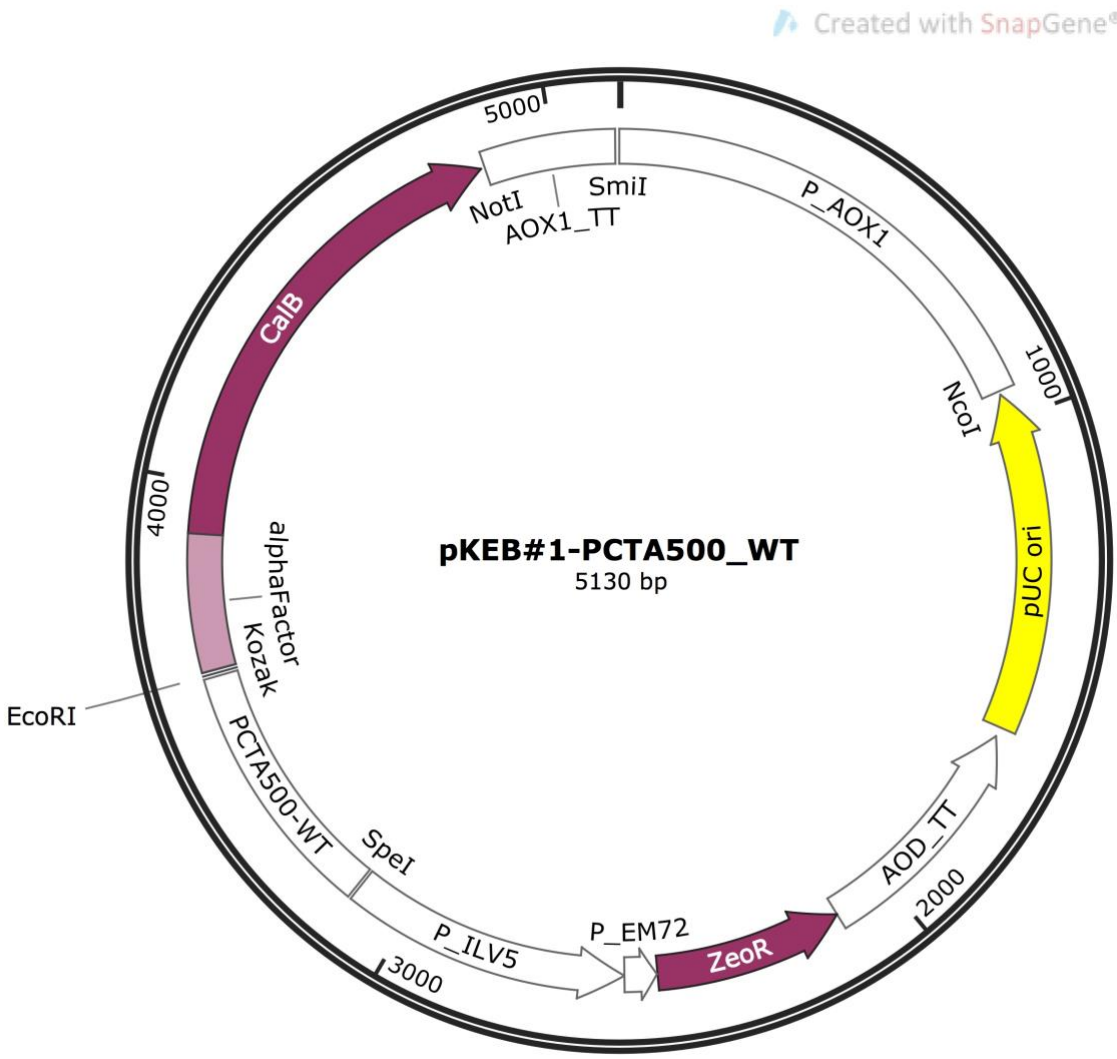

Figure 9.1 Expression construct of WT CTA500.
This vector encodes the target protein CalB under the control of the parental P_{CTA1}. P_{AOX1}: 5´*AOX*1 promoter; *AOX*1_TT: *AOX*1 transcription termination region; pUC ori: pUC replication origin for *E. coli*; ZeoR: Zeocin resistance gene; alpha factor: α-signal sequence for protein secretion.

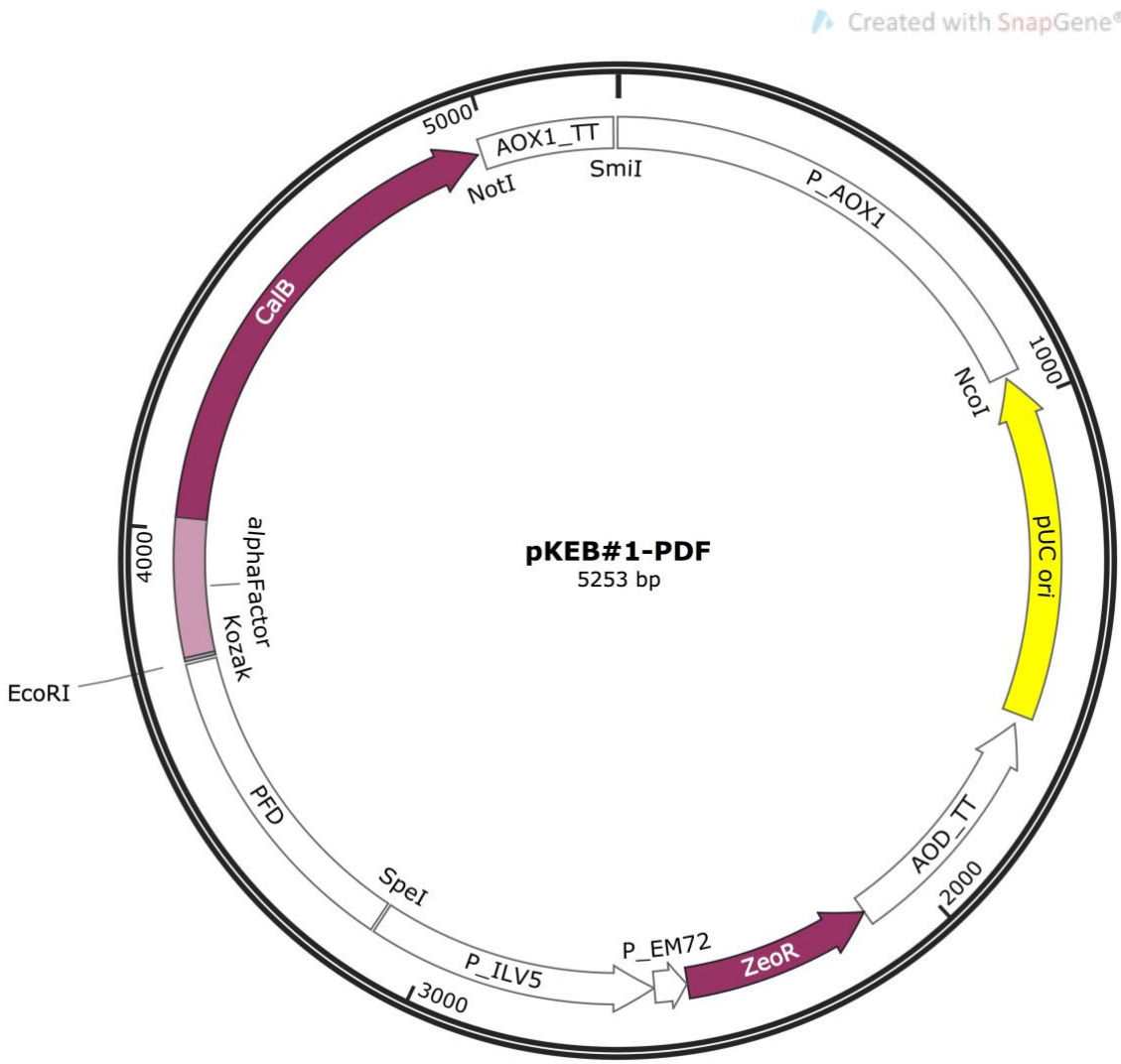

Figure 9.2 Expression construct of the benchmark P_{DF}.
This vector encodes the target protein CalB under the control of the P_{DF} promoter. P_{AOX1}: 5´*AOX*1 promoter; *AOX*1_TT: *AOX*1 transcription termination region; pUC ori: pUC replication origin for *E. coli*; ZeoR: Zeocin resistance gene; alpha factor: α-signal sequence for protein secretion.

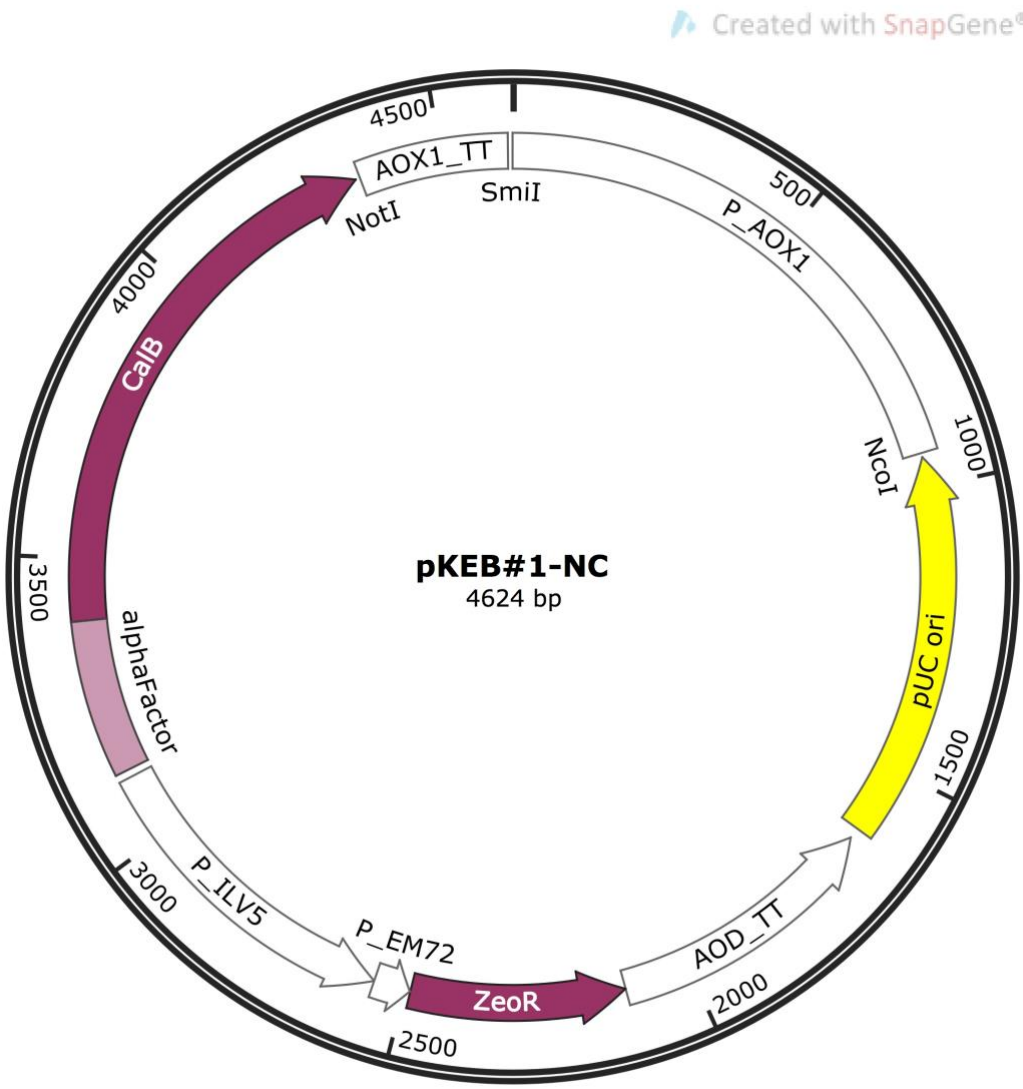

Figure 9.3 Expression construct of NC (negative control).
This vectors encodes the target protein CalB lacking of any controllable promoter. P_{AOX1}: 5´*AOX*1 promoter; *AOX*1_TT: *AOX*1 transcription termination region; pUC ori: pUC replication origin for *E. coli*; ZeoR: Zeocin resistance gene; alpha factor: α-signal sequence for protein secretion.

9.2 Coding sequences

9.2.1 pPp_AOX1-pUC-ZeoR-EcoRI-CalB-TT_AOX1.gbk (4624 bp)

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCCGACATC
CACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAGGAGGGGATACACTAG
CAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTTCTCCTCAACACCCACTT
TTGCCATCGAAAAACCAGCCCAGTTATTGGGCTTGATTGGAGCTCGCTCATTCCA
ATTCCTTCTATTAGGCTACTAACACCATGACTTTATTAGCCTGTCTATCCTGGCCC
CCCTGGCGAGGTTCATGTTTGTTTATTTCCGAATGCAACAAGCTCCGCATTACACC
CGAACATCACTCCAGATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTT
CCCCAAATGGCCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACA
AAAGCGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACGGC
CAGTTGGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTCTTGTTTGGTA
TTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAGCGCAGTCTCTCTATC
GCTTCTGAACCCCGGTGCACCTGTGCCGAAACGCAAATGGGGAAACACCCGCTTT
TTGGATGATTATGCATTGTCTCCACATTGTATGCTTCCAAGATTCTGGTGGGAATA
CTGCTGATAGCCTAACGTTCATGATCAAAATTTAACTGTTCTAACCCCTACTTGAC
AGCAATATATAAACAGAAGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATC
ATTATTAGCTTACTTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTT
AACGACTTTTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAAC
GCCATGGGTGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCG
CGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCG
ACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTT
TCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGAT
ACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGT
AGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAAC
CCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAA
CCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAG
CAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTA
CGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACC
TTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC
GGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAG
AAGATCCTTTGATCTTTTCTACGGGGGATCCGTCGACGAGCTCCTGCAGCTAAGG
TAATCAGATCCAAGTTTCCCCAATCTTCCGGGGCAAACCAATTCAGGCAATGTGG

TCCCACGATAAGGCCGAATATAACGGCCACAGAAGCTTCTCCGATATAGAGCTTC
TCTTTAATAAAAAGAGAGCACAGAGAAAATATTGTGGAAAAAACACCTATACAT
GCGTAGGCAATATGGGTTTTGGAGGGCGCTAAGTGCTCCCATATAGCCATGTTGT
AAAGGATTATGCAAAATTTGGCAGGAGAATCTGTGGATGCAAGGTATAATTATA
CCGAAGGGAAAAAAAGAGAGAAGAGCATAAAAAGCTTATTTTTAGTTGACGTTC
GCGGACATAGATTGCTGGCTTGATAAGATAGGACTTTTGGATAGTTACAGAATAT
AAGGGCATAAATAGATAATAAAACATCCGTATACATACAATAAAACTAATAAAT
ACTCTCTAAATAATCGTAAGGTGTCAATTTTAGTCCTGCTCTTCTGCGACGAAATG
CACGCAGTTACCAGCTGGATCACGCAGTGCAAACTCACGACCCCAGGGCTGTTCA
CCGATCTCGGTCATAGCTGGACCAGATGCATCACGGAAGTTGGTAGACACGACCT
CAGACCACTCAGCGTACAGTTCGTCCAGACCACGAACCCATACCCATGCCAGAGT
GTTGTCTGGCACAACCTGGTCCTGAACTGCGGAGATGAACAGGGTAACGTCGTCA
CGTACAACACCGGCAAAGTCGTCCTCTACGAAGTCACGGGAGAAACCGAGTCTA
TCAGTCCAGAACTCAACAGCACCAGCAACATCACGAGCAGTCAGGACTGGAACA
GCAGAGGTGAGTTTAGCCATGGTTTAGTCCTCCTTACACCTTGTCGTATTATACGA
GCCGGAAGTATAAAGTAACAACACTCTAGAAAATTTTTTCGGATAATTTTTTAAA
GCGAGAAAGTCACCCAGCGGATATGAACGCTTATTGTGTGCCGCACATCATATTG
GCGTTTGAAGTCGCGGTTTCGAGATTATCTCTTAATTTGTTCGCATTTTACTTTCA
AGCGCAACACAATGAAGTGCTCTAATATTACAAAGATTGTTGGAACTGGCTCACT
TGGAATTGCAACAGGCGTCCTTGGGTATGCATGCTTACGCGGAATCGACAACCTT
GTCACGTTGATCGATAACAACAAACTGGACCAACGAATGCGAAGGTTAGTAGGA
CGCGTTCTTAGATCCACACGGTACATTGTTGCCATATGCATGACTACTGGATCTTA
TCTGCTAGTAGAAACTTTCAGAAGATCACCTGCCAGTGCTCGACACTTATACCTG
ATTTATGCCAGTCTTGGTCTTCCTACGCTTGCACTTTACTACGGGTGGAATGTTTG
GCCTCTGGAAACAACTATTCTAAAGAAACTTTCACGAAGTCAAAATGGCTCACCA
CTGCAACAAAAGAAACAAGACATTACTGAAGGATCGAATTCCGAAACGATGAGA
TTCCCATCTATTTTCACCGCTGTCTTGTTCGCTGCCTCCTCTGCATTGGCTGCCCCT
GTTAACACTACCACTGAAGACGAGACTGCTCAAATTCCAGCTGAAGCAGTTATCG
GTTACTCTGACCTTGAGGGTGATTTCGACGTCGCTGTTTTGCCTTTCTCTAACTCC
ACTAACAACGGTTTGTTGTTCATTAACACCACTATCGCTTCCATTGCTGCTAAGGA
AGAGGGTGTCTCTCTCGAGAAGAGAGAGGCCGAAGCTTTGCCTTCAGGTTCAGA
CCCAGCCTTCTCACAGCCTAAATCAGTTTTGGACGCAGGTTTGACTTGCCAGGGT
GCTTCTCCTTCCTCCGTTTCCAAGCCTATTTTGCTTGTCCCAGGTACTGGAACTAC
AGGTCCTCAATCTTTTGATTCCAACTGGATTCCATTGAGTACTCAGCTTGGATACA

CACCTTGTTGGATCTCTCCACCTCCATTCATGTTGAACGACACACAAGTTAATACC
GAATACATGGTCAATGCAATTACTGCTTTGTATGCCGGTTCCGGAAACAATAAAT
TGCCAGTTCTTACTTGGTCACAAGGTGGATTGGTCGCTCAGTGGGGTCTTACATTT
TTCCCATCAATCCGTAGTAAGGTTGATAGATTGATGGCATTTGCTCCTGACTATA
AAGGTACTGTCTTGGCTGGACCATTGGATGCCCTTGCAGTTTCTGCCCCTTCCGTC
TGGCAACAGACCACTGGAAGTGCCTTGACAACCGCACTTAGAAACGCTGGTGGA
TTGACACAAATTGTTCCAACTACAAATCTTTACTCAGCTACCGATGAGATCGTTC
AACCTCAGGTCTCAAACAGTCCATTGGACTCTTCCTATCTTTTCAACGGAAAGAA
TGTTCAAGCTCAGGCCGTCTGCGGTCCTTTGTTTGTTATTGATCATGCTGGATCAT
TGACTAGTCAATTCTCTTACGTTGTCGGAAGATCCGCTTTGAGATCAACCACTGG
TCAGGCAAGATCTGCTGATTATGGAATTACCGACTGTAACCCTTTGCCAGCTAAT
GATCTTACTCCAGAACAAAAGGTTGCTGCCGCAGCTTTGCTTGCCCCTGCCGCAG
CTGCCATCGTTGCAGGTCCTAAACAGAATTGCGAGCCAGACTTGATGCCTTACGC
CAGACCATTTGCCGTTGGAAAGAGAACTTGCTCAGGAATCGTTACCCCTTAAGCG
GCCGCCTCAAGAGGATGTCAGAATGCCATTTGCCTGAGAGATGCAGGCTTCATTT
TTGATACTTTTTTATTTGTAACCTATATAGTATAGGATTTTTTTTGTCATTTTGTTT
CTTCTCGTACGAGCTTGCTCCTGATCAGCCTATCTCGCAGCTGATGAATATCTTGT
GGTAGGGGTTTGGGAAAATCATTCGAGTTTGATGTTTTTCTTGGTATTTCCCACTC
CTCTTCAGAGTACAGAAGATTAAGTGAGACGTTCGTTTGTGCATTTAAAT

9.2.2 P*CTA1*-14B+26B+48B

TAATCGAACTCCGAATGCGGTTCTCCTGTAACCTTAATTGTAGCATAGATCACTTA
AATAAACTCATGGCCTGACATCTGTACACGTTCTTATTGGTCTTTTAGCAATCTTG
AAGTCTTTCTATTGTTCCGATAACCGTGACCTAATAAATTCGAATCGAGATTGCTA
GTACCTGATATCATATGAAGTAATCATCACATGCAAGTTCCATGATACCCTCTAC
TAATGGAATTGAACAAAGTTTAAGCTTGATAACCGTGACCGAATCCATACTATGC
ACCCCTCAAAGTTGGGATTAGTCAGGAAAGCTGAGCAATTAACTTCCCTCGATTG
GCCTGGACTTTTCGCTTAGCCTGCCGCAATCGGTAAGTTTCATTATCCCAGCGGG
GTGATAGCCTCTGTTGCTCATCAGGCCAAAATCATATATAAGCTGTAGACCCAGC
ACTTCAATTACTTGAAATTCACCATAAGATAACCGTGAGTCAAGACTTACAATTA
AA

9.2.3 P_{CTA1}-1A+14B+26B

ATCCTTTTAGCCGAATGCGGTTCTCCTGTAACCTTAATTGTAGCATAGATCACTTA
AATAAACTCATGGCCTGACATCTGTACACGTTCTTATTGGTCTTTTAGCAATCTTG
AAGTCTTTCTATTGTTCCGATAACCGTGACCTAATAAATTCGAATCGAGATTGCTA
GTACCTGATATCATATGAAGTAATCATCACATGCAAGTTCCATGATACCCTCTAC
TAATGGAATTGAACAAAGTTTAAGCTTGATAACCGTGACCGAATCCATACTATGC
ACCCCTCAAAGTTGGGATTAGTCAGGAAAGCTGAGCAATTAACTTCCCTCGATTG
GCCTGGACTTTTCGCTTAGCCTGCCGCAATCGGTAAGTTTCATTATCCCAGCGGG
GTGATAGCCTCTGTTGCTCATCAGGCCAAAATCATATATAAGCTGTAGACCCAGC
ACTTCAATTACTTGAAATTCACCATAACACTTGCTCTAGTCAAGACTTACAATTAA
A

9.2.4 P_{CTA1}

TAATCGAACTCCGAATGCGGTTCTCCTGTAACCTTAATTGTAGCATAGATCACTTA
AATAAACTCATGGCCTGACATCTGTACACGTTCTTATTGGTCTTTTAGCAATCTTG
AAGTCTTTCTATTGTTCCGGTCGGCATTACCTAATAAATTCGAATCGAGATTGCTA
GTACCTGATATCATATGAAGTAATCATCACATGCAAGTTCCATGATACCCTCTAC
TAATGGAATTGAACAAAGTTTAAGCTTCTCGCACGAGACCGAATCCATACTATGC
ACCCCTCAAAGTTGGGATTAGTCAGGAAAGCTGAGCAATTAACTTCCCTCGATTG
GCCTGGACTTTTCGCTTAGCCTGCCGCAATCGGTAAGTTTCATTATCCCAGCGGG
GTGATAGCCTCTGTTGCTCATCAGGCCAAAATCATATATAAGCTGTAGACCCAGC
ACTTCAATTACTTGAAATTCACCATAACACTTGCTCTAGTCAAGACTTACAATTAA
A

9.2.5 P_{DF}

AATGTATCTAAACGCAAACTCCGAGCTGGAAAAATGTTACCGGCGATGCGCGGA
CAATTTAGAGGCGGCGATCAAGAAACACCTGCTGGGCGAGCAGTCTGGAGCACA
GTCTTCGATGGGCCCGAGATCCCACCGCGTTCCTGGGTACCGGGACGTGAGGCAG
CGCGACATCCATCAAATATACCAGGCGCCAACCGAGTGTCTCGGAAAACAGCTTC
TGGATATCTTCCGCTGGCGGCGCAACGACGAATAATAGTCCCTGGAGGTGACGG

AATATATATGTGTGGAGGGTAAATCTGACAGGGTGTAGCAAAGGTAATATTTTCC
TAAAACATGCAATCGGCTGCCCCGCAACGGGAAAAAGAATGACTTTGGCACTCTT
CACCAGAGTGGGGTGTCCCGCTCGTGTGTGCAAATAGGCTCCCACTGGTCACCCC
GGATTTTGCAGAAAAACAGCAAGTTCCGGGGTGTCTCACTGGTGTCCGCCAATAA
GAGGAGCCGGCAGGCACGGAGTTTACATCAAGCTGTCTCCGATACACTCGACTAC
CATCCGGGTCTCTCAGAGAGGGGAATGGCACTATAAATACCGCCTCCTTGCGCTC
TCTGCCTTCATCAATCAAATC

9.3 Validation of HKGs *ARG4* and *ACT1* with target *P. pastoris* strains

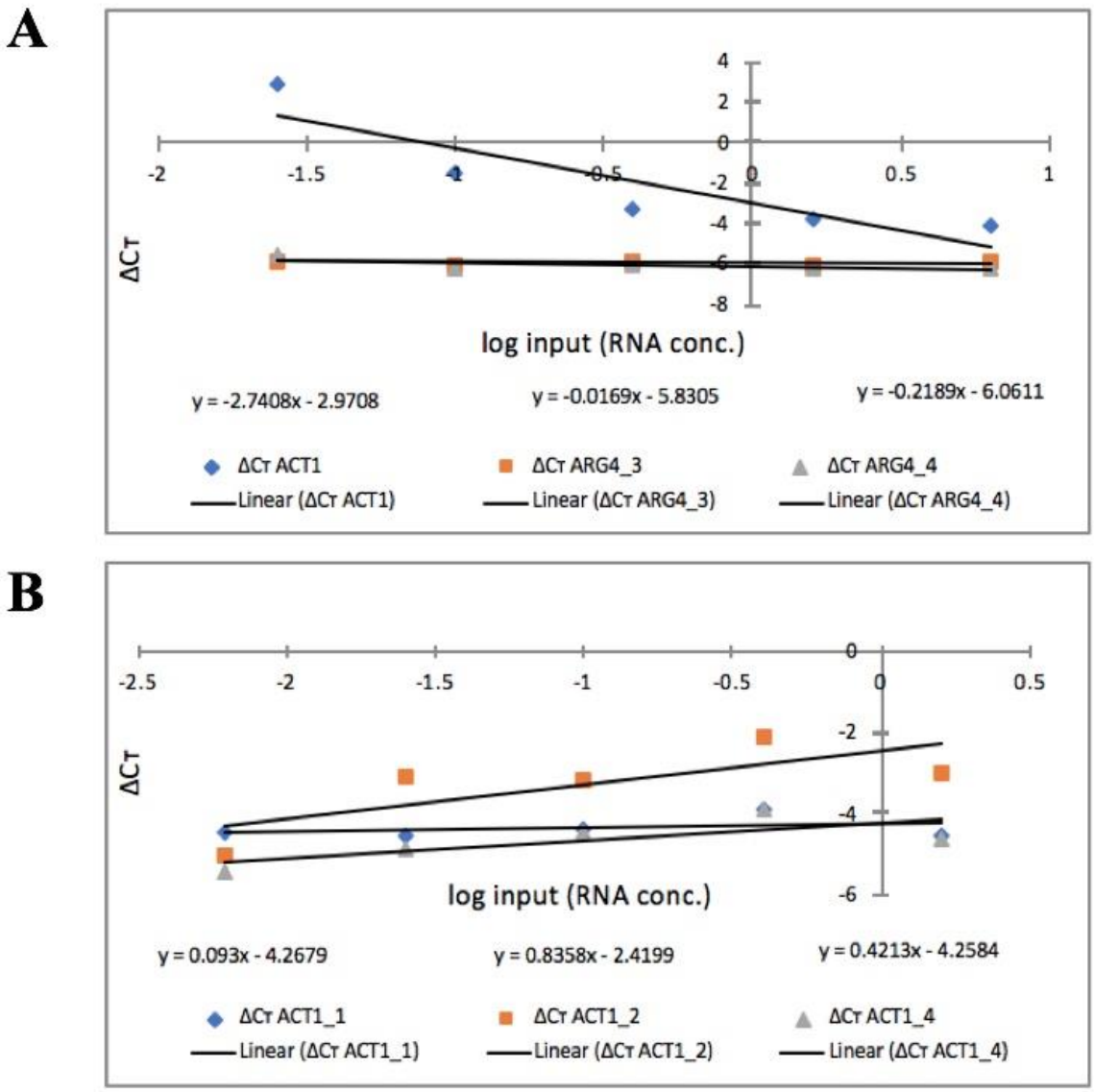

Figure 9.4 Total plot of the validation experiments performed in the course of cultivation round 1 (*A*) and 2 (*B*).

In this graph, all C_T values (y-axis) are plotted against the log input amounts (x-axis) of the specific target gene (CalB). For *A*, Variant 26 at timepoint t4 cultivated in BMG was used to execute the validation experiment. For *B*, WT strain at timepoint t3 cultivated in BMD was used. By using the

respective strain, dilution series with varying target input materials were prepared to obtain a standard curve encompassing the following amounts of absolute RNAs: for *A*, 6.4 ng, 1.6 ng, 0.4 ng, 0.1 ng and 0.025 ng of absolute RNA per reaction; for *B*, 1.6 ng, 0.4 ng, 0.1 ng, 0.025 ng and 0.00625 ng of absolute RNA per reaction. The PCR efficiency E of each reaction was calculated as described in *section 3.5.5.2*.

A

BMD																							
		t1			t2			t3			t4			t5			t6			t7			
		a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	
cultivation round 1	19	0.645	0.588	-	9.3	8.9	-	10.5	10.4	-	12.88	13.1	-	14.6	16.4	-	-	-	-	-	-	-	
	26	0.609	0.705	-	7.7	8.82	-	10.8	10.8	-	12.8	11.5	-	15.7	14	-	-	-	-	-	-	-	
	NC	0.514	0.521	-	8.6	8.5	-	11.3	11.3	-	13.2	13.6	-	15.3	16.4	-	-	-	-	-	-	-	
	WT	0.704	0.578	-	8.12	9.56	-	10.2	11.4	-	12.2	12.4	-	14.9	17.1	-	-	-	-	-	-	-	
	PDF	0.51	0.439	-	11.1	11.04	-	11.1	11.7	-	13.4	12.9	-	16.4	15.2	-	-	-	-	-	-	-	
		t1			t2			t3			t4			t5			t6			t7			
		a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	
cultivation round 2	19	0.486	0.617	0.497	10.5	10.28	11.02	10.8	8.9	10.5	9.5	9.8	10.1	8.4	10.4	11.6	11.3	11.9	8.3	10.6	12	12.4	
	26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	NC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	WT	0.437	0.463	0.673	10.54	10.28	10.3	15.52	13	10.6	11.4	8.9	12.3	10.6	11.8	10.6	11.6	12.2	11.4	12	12.4	11.8	
	PDF	0.435	0.449	0.639	10.24	9.78	9.88	11.6	14.7	12.5	10.8	10.4	9.9	10.4	11	10.5	11.6	11.4	12	12	12	11.8	

B

BMG																							
		t1			t2			t3			t4			t5			t6			t7			
		a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	
cultivation round 1	19	0.58	0.57	-	14.92	17.9	-	16.7	16.6	-	19.4	23.2	-	23.5	25.7	-	-	-	-	-	-	-	
	26	0.509	0.53	-	14.8	14.66	-	16.9	18.5	-	20.7	22.8	-	26.9	27.8	-	-	-	-	-	-	-	
	NC	0.51	0.567	-	14.99	14.9	-	17.9	17.7	-	20.9	23.8	-	26.1	31.3	-	-	-	-	-	-	-	
	WT	0.619	0.64	-	14.52	15	-	18.4	16.04	-	20.5	20.4	-	27.6	29.3	-	-	-	-	-	-	-	
	PDF	0.533	0.53	-	13.32	12.7	-	16.3	16.4	-	21.2	21.1	-	27.5	27.1	-	-	-	-	-	-	-	
		t1			t2			t3			t4			t5			t6			t7			
		a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	
cultivation round 2	19	0.561	0.572	0.575	9.14	10.9	11.06	15.4	14.4	15.1	14.7	15.1	16.3	19.8	19.2	18.3	19.7	20.7	20	22.5	22.4	26.2	
	26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	NC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	WT	0.463	0.482	0.448	8.84	9.74	8.5	14.6	14.8	6.3	16.1	16.7	16.9	20	19.7	20.3	21.9	20.8	21.8	22.5	23	23	
	PDF	0.561	0.56	0.506	11.7	10.64	9.76	13.6	15.2	15.1	8.9	15.9	15.5	20.6	21	22.3	18.4	21.4	20.2	24.1	21.7	21.9	

Figure 9.5 Raw values of cell density measurements obtained during both cultivation processes.

During both cultivation processes (round 1 and 2), samples were taken from each culture to measure the optical cell density by using a spectrophotometer at 600 nm. To do so, clean cuvettes were used and filled up to 1 mL with the culture being analyzed. Depending on the values, the culture samples had to be diluted in order to stay in the linear range (from approximately 0.1-0.6). If a dilution was necessary, the results were re-calculated in order to get the density of cells per 1 mL. Both tables show the absolute D_{600} values – *A* represents the values obtained for cells cultivated in BMD, *B* for cells cultivated in BMG. According to the data obtained, 10 D_{600} unit samples were harvested and further processed.