TRANSCRIPTIONAL ENGINEERING OF PICHIA PASTORIS ALCOHOL DEHYDROGENASE 2 AND ALCOHOL OXIDASE 1 PROMOTERS FOR RECOMBINANT PROTEIN PRODUCTION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

BURCU GÜNDÜZ ERGÜN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
BIOTECHNOLOGY

SEPTEMBER 2018

Approval of the thesis:

TRANSCRIPTIONAL ENGINEERING OF *Pichia pastoris* ALCOHOL DEHYDROGENASE 2 AND ALCOHOL OXIDASE 1 PROMOTERS FOR RECOMBINANT PROTEIN PRODUCTION

submitted by **BURCU GÜNDÜZ ERGÜN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Biotechnology Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Assoc. Prof. Dr. Can Özen
Head of Department, **Biotechnology** _____

Prof. Dr. Pınar Çalık
Supervisor, **Chemical Engineering Dept., METU** _____

Prof. Dr. Diethard Mattanovich
Co-Supervisor, **Biotechnology Dept., BOKU** _____

Examining Committee Members:

Prof. Dr. H. Tunçer Özdamar
Chemical Engineering Dept., Ankara University _____

Prof. Dr. Pınar Çalık
Chemical Engineering Dept., METU _____

Prof. Dr. Haluk Hamamcı
Food Engineering Dept., METU _____

Prof. Dr. A. Elif Erson Bensan
Biology Dept., METU _____

Assist. Prof. Dr. Eda Çelik Akdur
Chemical Engineering Dept., Hacettepe University _____

Date: 03/09/2018

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Burcu Gündüz Ergün

Signature :

ABSTRACT

TRANSCRIPTIONAL ENGINEERING OF *Pichia pastoris* ALCOHOL DEHYDROGENASE 2 AND ALCOHOL OXIDASE 1 PROMOTERS FOR RECOMBINANT PROTEIN PRODUCTION

Gündüz Ergün, Burcu
Ph.D., Department of Biotechnology
Supervisor: Prof. Dr. Pınar Çalık
Co-Supervisor: Prof. Dr. Diethard Mattanovich

September 2018, 310 pages

The aim of the Ph.D. thesis was designing novel engineered promoter variants (NEPVs) to obtain strong ethanol regulated promoters for improved recombinant protein (r-protein) production, and to clarify the role of certain transcription factors (TFs) on the regulation of *P. pastoris* P_{ADH2} and P_{AOX1}. Promoter architecture of P_{ADH2} was redesigned by nucleosome optimization and modification of transcription factor binding sites and, the induction mechanism of P_{AOX1} was changed to ethanol as sole carbon and energy source. Engineered P_{ADH2} allowed a 476% increase in enhanced green fluorescent protein (eGFP) synthesis compared to that of $P_{ADH2\text{-}wt}$. NEPVs of P_{AOX1} spanning an activity range between 74% and 130% of the $P_{AOX1\text{-}wt}$ in fermentations with ethanol; and allow maximum 197% increase in eGFP synthesis in

fermentations with methanol. Subsequent measurements of the eGFP transcript levels confirmed stronger transcriptional capacities of the NEPVs. The second model protein extracellular human serum albumin (HSA) production was also studied with the promising NEPVs. Average extracellular HSA yield per gram of wet-cell-weight ($Y_{P/X}$) was 0.26, 0.38, 1.13, and 0.48 mg/g with $P_{ADH2\text{-}wt}$, $P_{ADH2\text{-}NucOpt}$, $P_{ADH2\text{-}OptCat}$, and $P_{AOX1\text{-}mod,}$ respectively under ethanol induction. In methanol-based defined medium with $P_{AOX1\text{-}wt}$ and $P_{AOX1\text{-}mod}$ $Y_{P/X}$ was calculated as 0.85 and 1.34 mg/g, respectively. In order to clarify the role of the TFs in the regulation of P_{ADH2} and P_{AOX1}, and to create novel r-protein expression systems, overexpression and knock-out *P. pastoris* strains of Adr1(Mxr1), Aca1, Cat8-1, and Cat8-2 were designed and constructed. Cat8-2 overexpression created a derepressible r-protein production system under limited glucose condition with P_{ADH2} reached 37% of the expression reached with ethanol induction. A very potent activation by derepression was observed with $P_{AOX1\text{-}mod}$ in *adr1*Δ strain with 206% of the expression reached with methanol-induced $P_{AOX1\text{-}wt}$ with limited glucose. Adr1 was found to be the main regulator of P_{AOX1} but not P_{ADH2}. Cat8-1 and Cat8-2 were both the activators of P_{ADH2} and $P_{AOX1\text{-}v}$; but in fact Cat8-1 is more effective. The *cat8-1cat8-2Δ* strain lost its ethanol utilization ability. The strongest expression potential by $P_{ADH2\text{-}wt}$ and $P_{AOX1\text{-}mod}$ was obtained in *adr1*Δ strains under methanol induction, although the cells lost their ability to grow. It is conclusively demonstrated that the NEPVs and transcriptionally engineered r-protein expression platforms are promising candidates for industrial r-protein production processes.

Keywords: *Pichia pastoris, ADH2, AOX1,* promoter library, transcriptional engineering

ÖZ

REKOMBİNANT PROTEİN ÜRETİMİ İÇİN *Pichia pastoris* ALKOL DEHİDROJENAZ 2 VE ALKOL OKSİDAZ 1 PROMOTORLARININ TRANSKRİPSİYON MÜHENDİSLİĞİYLE GELİŞTİRİLMESİ

Gündüz Ergün, Burcu
Doktora, Biyoteknoloji Bölümü
Tez Yöneticisi : Prof. Dr. Pınar Çalık
Ortak Tez Yöneticisi: Prof. Dr. Diethard Mattanovich

Eylül 2018, 310 sayfa

Bu doktora tezinin amacı rekombinant protein (r-protein) üretiminin artırılması için güçlü, etanol ile regüle edilen promotor varyantlarının tasarlanması ve geliştirilmesi, ve belirlenen transkripsiyon faktörlerinin (TF) *P. pastoris* P_{ADH2} ve P_{AOX1} promotorlarının regülasyonunda işlevlerinin araştırılmasıdır. P_{ADH2} promotor mimarisi nükleozom optimizasyonu ve transkripsiyon faktörü bağlanma bölgelerinin modifikasyonu ile değiştirilmiş, etanolün tek karbon ve enerji kaynağı olarak kullanılacağı P_{AOX1} indüksiyon mekanizması yeniden tasarlanmıştır. P_{ADH2} varyantları $P_{ADH2\text{-}wt}$ ile kıyaslandığında, %476 daha fazla arttırılmış yeşil floresan protein (eGFP) sentezi gerçekleştirmiştir. P_{AOX1} varyantları etanol indüksiyonunda doğal promotorun

%74'ü ile %130'u arasında aktivite göstermiş; metanol indüksiyonunda $P_{AOX1\text{-}wt}$'in %197'si kadar eGFP üretimi gerçekleştirmiştir. eGFP transkript düzeylerinin ölçümü ile geliştilen promotor varyantlarının (NEPV) güçlendirilmiş transkripsiyon kapasiteleri doğrulanmıştır. NEPV'nin potansiyellerinin değerlendirilmesi için öne çıkan üç promotor varyantı ile hücre dışı insan serum albumin (HSA) üretimi gerçekleştirilmiştir. Etanol indüksiyonu ile, $P_{ADH2\text{-}wt}$, $P_{ADH2\text{-}NucOpt}$, $P_{ADH2\text{-}OptCat}$, ve $P_{AOX1\text{-}mod}$ promotor varyantları ile sırasıyla birim yaş hücre ağırlığı başına üretilen ortalama hücredışı HSA verim katsayıları ($Y_{P/X}$) 0.26, 0.38, 1.13, ve 0.48 mg/g olarak hesaplanmıştır. Metanol içeren tanımlı üretim-ortamlarında $P_{AOX1\text{-}wt}$ ve $P_{AOX1\text{-}mod}$ ile verim katsayısı $Y_{P/X}$ 0.85 ve 1.34 mg/g bulunmuştur. P_{ADH2} ve P_{AOX1} regülasyonunda transkripsiyon faktörlerinin rollerinin netleştirilmesi ve yeni r-protein üretim sistemlerinin oluşturulması için Adr1(Mxr1), Aca1, Cat8-1, ve Cat8-2 TF aşırı üretim (OE) ve TF geni silinmiş (KO) suşlar geliştirilmiştir. Cat8-2 OE, baskısı kaldırılan r-protein üretim sistemi yaratmış ve bu sistemi taşıyan hücre sınırlı-glukoz üretim ortamında, etanol ile indüklenerek üretilenin %37'sine ulaşmıştır. Baskı kaldırılması ile sınırlı-glukoz üretim ortamında *adr1Δ* suşunda $P_{AOX1\text{-}mod}$ ile, metanolle indüklenmiş $P_{AOX1\text{-}wt}$'ın %206'sı kadar r-protein üretimi sağlayan güçlü uyarılma bulunmuştur. P_{AOX1}'in ana aktivatörünün Adr1 olduğu, ancak P_{ADH2}'nin ana düzenleyicisi olmadığı bulunmuştur. Cat8-1 ve Cat 8-2'nin her ikisinin de P_{ADH2} ve $P_{AOX1\text{-}v}$'nin aktivatörleri olduğu ve Cat8-1'in daha etkili olduğu bulunmuştur. *cat8-1cat8-2Δ* suşunun etanol kullanabilme yetisini kaybettiği gözlemlenmiştir. Hücreler çoğalma yetilerini kaybetmiş olsalar da en güçlü üretim potansiyeli metanol uyarımı altında $P_{ADH2\text{-}wt}$ ve P_{AOX1}-mod ile *adr1*Δ suşlarında görülmüştür. Sonuç olarak, NEPV'ler ve transkripsiyon mühendisliği ile tasarlanan r-protein üretim platformları endüstriyel r-protein üretim prosesleri için umut vericidir.

Anahtar kelimeler: *Pichia pastoris, ADH2, AOX1,* promotor kütüphanesi, transkripsiyon mühendisliği

To My Family

ACKNOWLEDGMENTS

I am grateful to my supervisor Prof. Dr. Pınar Çalık for her never-ending support, guidance, encouragement, and help throughout my graduate studies in which I spent seven informative and productive years in her research laboratory in METU. This period not only contributed to my academic development but also to my intellectual maturation.

I am thankful to my co-supervisor Prof. Dr. Diethard Mattanovich for his support, guidance, and help during my research in BOKU, Vienna. His hospitability made me comfortable during my stay. I also want to thank to Assoc. Prof. Dr. Brigitte Gasser for her guidance, comments, and positive attitudes during my studies in BOKU.

I would like to express my sincere gratitudes to Prof. Dr. Tunçer H. Özdamar for his valuable comments and guidance throughout my Ph.D thesis study. His advices contributed not only my academic improvement but also personal life.

I want to thank Prof. Dr. A. Elif Erson Bensan for her valuable comments during the thesis monitoring committees.

A big thank you goes to my friends in my research group: Sibel Öztürk, Aslan Massahi, Erdem Boy, Damla Hüccetoğulları, Özge Kalender, Yiğit Akgün, Onur Ersoy, İrem Demir, Ömer Masri, and Abdullah Keskin for their support, friendship, and making the lab work more enjoyable. I would also like to thank to all my friends from Yeast Molecular Biotechnology Research Group in BOKU for their friendship, support, and making me feel like I am home.

I would like to thank to all academic, administrative, and technical staff of Department of Chemical Engineering (METU) and Departments of Biotechnology in METU and at BOKU, for their help and support throughout my education.

The Scientific and Technological Research Council of Turkey's: *i)* national and international scholarships, respectively, BIDEB 2210 and BIDEB 2214; and, *ii)* the research fund TÜBİTAK-116Z215; and, Middle East Technical University Research Fund Project, are greatly acknowledged.

Last but not least, I would like to deeply thank to my big family for always being there for me and making me stronger. I would never be successful without their support. A special thanks goes to Recebiye & Safettin Gündüz and Nurdan & Akif Ergün who have always supported and encouraged me throughout my life. My special thanks go to Onur Ergün for his loving, caring and support. I want to thank to you for all the wonderful ways you make me happy.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

NOMENCLATURE

WCW	Wet cell weight	g/L
Y	Yield	mg/g

Subscripts

ADH2-v	Refers to ADH2 promoter variant
AOX1-v	Refers to AOX1 promoter variant
P	Refers to product
X	Refers to cell
$(P_{ADH2})_E$	Refers to ethanol induced ADH2 promoter activity
$(P_{AOX1})_M$	Refers to methanol induced AOX1 promoter activity

Abbreviations

ADH2	Alcohol dehydrogenase 2
AOX1	Alcohol oxidase 1
ELISA	The enzyme-linked immunosorbent assay
F$	Fungi specific TF matrix
GGA	Golden Gate Assembly
HSA	Human serum albumin
KO	Knock-out
mTR	mRNA transcription ratio

NEPV	Novel engineered promoter variant
OE	Over-expression
P	Promoter
PCR	Polymerase chain reaction
rHSA	Recombinant human serum albumin
TEC	Transcriptionally engineered cell
TF	Transcription factor
TFBS	Transcription factor binding site
UAS	Upstream activator site
URS	Upstream repressor site

CHAPTER 1

INTRODUCTION

Organization for Economic Co-operation and Development (OECD) (2001) defines biotechnology as the application of science and technology to living organisms, as well as parts, products, and models thereof, to alter living or non-living materials for the production of knowledge, goods, and services. Industrial biotechnology specifically focuses on the development of biomolecule production processes using metabolic engineering, strain engineering, and bioprocess engineering.

A key component of the industrial biotechnological processes is the host organism. Yeast has been the subject of many studies due to its ability to grow rapidly, reach very high cell densities in defined minimal media and provide high yields. The genetic manipulations applied to yeast are rather straightforward, and the eukaryotic post-translational modifications of yeast increase the biopharmaceutical potential of its products (Çelik and Çalık, 2012). Among industrially relevant yeast species, the *Pichia pastoris (Komagetalla pastoris)* has been very successful as an efficient heterologous host that can be attributed to a number additional advantages. Namely, it secretes recombinant proteins (r-protein) with fewer amounts of endogenous proteins, which leads to easier downstream processes, it does not hyperglycosylate r-proteins as much as *Saccharomyces cerevisiae*, and it prefers aerobic respiration rather than fermentation, therefore produces less by-products (Gündüz-Ergün and Çalık, 2016).

Transcription is the first step in protein synthesis and a crucial step in the cell in fermentation bioprocesses. Therefore, strong and controllable promoters are essential for efficient expression systems (Porro et al., 2005). In bioprocess operations, regulatable promoters help to divide cell growth and production phases and enable high cell densities without r-protein production hampering the cellular activities in the growth phase. One of the most important features of methylotrophic yeasts is their strong and tightly regulated endogenous promoters obtained from the methanol utilization pathway. Alcohol oxidase 1 (AOX1) is the first enzyme in the methanol utilization pathway of *P. pastoris* and expressed at high levels; however, P_{AOX1} is strongly repressed when *P. pastoris* is grown on glycerol, ethanol, or glucose (Inan and Meagher, 2001). Upon depletion of these carbon sources, the promoter is derepressed but is fully induced only after the addition of methanol. Due to its strength and tight regulation, P_{AOX1} has become one of the most extensively used yeast promoter for recombinant protein production (Çelik and Çalık, 2012). Despite its high r-protein expression potential, its induction agent, toxic methanol, still limits large-scale industrial applications with P_{AOX1}.

P. pastoris alcohol dehydrogenase 2 promoter (P_{ADH2}) was reported to be an inducible promoter by ethanol and glycerol (Cregg and Tolstorukov, 2012). Functional analysis of *ADH2* revealed that it is the only gene responsible for ethanol utilization in *P. pastoris* while *adh2* knock-out strains can still produce ethanol (Karaoglan et al., 2016a). Expression potential with P_{ADH2} was investigated by pilot-scale bioreactor experiments; and, recombinant xylanase production with P_{ADH2} outcompeted the commonly used strong and inducible P_{AOX1} and constitutive P_{GAP} expression levels (Karaoglan et al., 2016b).

Transcriptional engineering is a novel metabolic engineering tool since it allows modification of the regulation of an intracellular reaction pathway by engineering the multiple components such as promoters or regulators including transcription factors (Ata et al., 2017; Bajhaiya et al., 2017). Various endogenous promoters were identified to satisfy the requirements of different bioprocesses. However, native

promoters usually provide only limited opportunities due to their intricate regulation mechanisms and highly-dependent working principles which hampers their production efficiency and increases the metabolic burden over the host cell.

Promoter engineering aims to modulate transcriptional capacity of promoters by mutating promoter DNA sequences to obtain a promoter library with a diverse dynamic range. The basis of promoter engineering mostly depends on promoter architecture. Promoter strength and regulation are the cumulative effects of short and distinct DNA motifs that facilitate binding of cellular transcriptional machinery. Depending on this idea, promoter engineering includes many approaches such as chimeric promoter design, random mutagenesis, modification of transcription factor binding sites, and synthetic promoter designs (Ata et al., 2017; Blazeck et al., 2012; Hartner et al., 2008; Portela et al., 2017; Vogl et al., 2014). Many promoter engineering methods have been applied to understand the regulation and improve the expression strength of P_{AOX1} (Hartner et al., 2008; Portela et al., 2017; Vogl et al., 2014; Xuan et al., 2009). However, there is no study in the literature mentioning about changing the induction mechanism of P_{AOX1} by modifying its *cis*-acting elements so far. Similarly, despite some promising results with P_{ADH2} recently, there is no study evaluating the *trans*- or *cis*-acting elements of the P_{ADH2}.

The ultimate aim of this Ph.D. thesis is designing and engineering novel promoter variants to obtain strong ethanol regulated promoter variant(s) for improved r-protein production, and to clarify the role of certain transcription factors on the regulation of *P. pastoris* ADH2 (P_{ADH2}) and AOX1 (P_{AOX1}) promoters.

CHAPTER 2

LITERATURE REVIEW

2.1 *Pichia pastoris*

P. pastoris is an ascosporous yeast that was first isolated from exudate of chestnut tree in France in 1919 by Guilliermond, and around mid-20th century its methanol utilizing ability as sole carbon source was discovered (Ogata et al., 1969). Philips Petroleum Company employed *P. pastoris* to produce single cell protein (SCP) as an animal feed additive; however dramatic increase in methanol prices because of the oil crisis in 1973 made SCP processes uneconomical. In the 1980s, *P. pastoris* was transformed into a heterologous host system to produce r-proteins under control of the strong and tightly regulated alcohol oxidase 1 promoter (P_{AOX1}) (Cregg et al., 1985). *P. pastoris* has attracted attention of the scientific community and currently has become the most prevalent yeast system in r-protein production research. More than ten years ago, phylogenetic analyses led to reclassification of the species *P. pastoris* to the genus *Komagataella*, which was further split into *Komagataella pastoris and Komagataella phaffi* (Kurtzman, 2009). As both *K. pastoris* and *K. phaffii* strains have been used in research and industry since the beginning, for the sake of simplicity, the *Pichia* community continues to call the yeast as *P. pastoris.* The success of *P. pastoris* as an efficient heterologous host can be attributed to a number of advantages including its ability to grow rapidly to very high cell densities in defined minimal medium, rather straightforward genetic manipulations, and – like other yeast hosts – perform eukaryotic post-translational

modifications such as glycosylation, disulphide bond formation, proteolytic modification and targeting to subcellular compartments (Gündüz-Ergün and Çalık, 2016). Specifically, *P. pastoris* is able to reach high product yields and secretes r-proteins to the extracellular medium relatively pure of contaminations due to low-level secretion of its endogenous proteins. The N-glycans in *P. pastoris* are of the mannose-rich yeast type and usually contain 9-11 mannoses with terminal α-1,2-linkages. Unlike *S. cerevisiae*, *P. pastoris* does not hyper-glycosylate therapeutic proteins and does not contain potentially immunogenic terminal α-1,3-linked mannoses. During the last ten years several humanized *P. pastoris* strains were developed and applied for r-protein production to allow for more homogenous and proper glycosylation patterns (Choi et al., 2003; Hamilton et al., 2003; Vervecken et al., 2004)

In *P. pastoris*, high-level expression of exogenous genes has been predominantly controlled by methanol-induced P_{AOX1}. Regarding methanol utilization, there are three possible phenotypes of *P. pastoris*, namely Mut^+ (wt methanol consumption; both of the alcohol oxidase enzymes, Aox1 and Aox2, are functional), Mut^S (slow methanol consumption; disrupted Aox1, functional Aox2) and Mut^- (no methanol consumption; both Aox1 and Aox2 are disrupted). The constitutive promoter of glyceraldehyde-3-phosphate dehydrogenase P_{GAP} provides comparable expression level as P_{AOX1} and also facilitates the process management by eliminating the use of potentially hazardous inducer methanol (Çalık et al., 2015).

Moreover, *P. pastoris* is a model organism to study peroxisome biology as it possesses the peculiar methanol utilization metabolism; and for the secretory pathway as the structure of its ER exit sites and Golgi apparatus are similar to higher eukaryotes in contrast to *S. cerevisiae* (Gasser et al., 2013). A major breakthrough was the availability of the genome sequence data of the commercial *P. pastoris* strain GS115 in 2009, which has provided researchers a deeper insight into yeast physiology and genetics, and speeded up the developments of the genetic toolbox or engineered

strains (De Schutter et al., 2009). Recently, the sequence data of *P. pastoris DSMZ* 70382 (Love et al., 2016; Mattanovich et al., 2009) and original SCP production strain CBS7435 (Kuberl et al., 2011; Sturmberger et al., 2016; Valli et al., 2016) were published. Food and Drug Administration (FDA) gave GRAS (generally regarded as safe) status to phospholipase C (Ciofalo et al., 2006) as a feed additive and also in 2009, the first biopharmaceutical protein produced by a non-*Saccharomyces* yeast, a kallikrein inhibitor named Kalbitor® produced by *P. pastoris* was approved by the FDA (Thompson, 2010).

2.1.1 *P. pastoris* Promoters

Transcription, the first step of protein synthesis, is crucial in r-protein production; therefore, strong and controllable promoters are needed for efficient expression systems (Porro et al., 2005). One of the most important features of methylotrophic yeasts is their strong and tightly regulated promoters in the methanol utilization (MUT) pathway. Alcohol oxidase (AOX) catalyses the first step of the methanol utilization, by conversion of methanol to formaldehyde. In *P. pastoris*, under methanol induction, AOX constitutes 5% of total mRNA and 33% of total cellular protein (Couderc and Baratti, 1980). But the growth on repressing carbon sources, *i.e.* glucose, ethanol and glycerol, lead to complete lack of AOX activity (Inan and Meagher, 2001). Tight regulation of P_{AOX1} provides the ability to divide the cell growth into biphasic fermentation processes, which enables high cell densities in the first phase without r-protein production hampering the cellular activities in the growth phase on glycerol or glucose. *P. pastoris* has two *AOX* genes, *AOX1* and *AOX2* (Cregg et al., 1989). As P_{AOX1} is significantly stronger and tightly regulated with carbon sources, it has become the most prevalently used promoter for r-protein production in *P. pastoris*. Up to 22 g/L intracellular (Hasslacher et al., 1997) and 10.5 g/L secreted (Schotte et al., 2016) heterologous protein production was achieved by P_{AOX1}. P_{AOX2} has a similar regulation profile, however, it provides around 10–20 times

lower expression levels than P_{AOX1} (Macauley-Patrick et al., 2005). Although P_{AOX1} is the pioneer promoter in r-protein production research by *P. pastoris*, its inducer methanol is a toxic and flammable compound that can cause problems and safety concerns in industrial applications. Other important promoters of the methanol utilization (MUT) pathway are formaldehyde dehydrogenase (P_{FLD1}) and dihydroxyacetone synthase (P_{DAS1}) promoters. P_{DAS} shows similar regulatory pattern and production levels as P_{AOX1} (Tschopp et al., 1987), while P_{FLD1} has the opportunity to be induced by either methanol or in a methanol-free system with methylamine (Shen et al., 1998). Most of the commercial vectors for *P. pastoris* employ P_{AOX1}, but there are also P_{FLD1} based vectors. Among alternative inducible promoters, P_{G1} is promising as it is induced by limiting the safe carbon source glucose and shows higher expression levels than P_{GAP} (Prielhofer et al., 2013). Other regulated promoters induced under phosphate limitation are the acid phosphatase promoter P_{PHO1} (Payne et al., 1995) and the promoter of the putative sodium-coupled phosphate symporter P_{PHO89} (also named as P_{NPS}) (Ahn et al., 2009). To develop methanol-free inducible systems, P_{PHO89} is also an alternative as it has similar strength as P_{GAP} (Ahn et al., 2009). Some novel promoters, P_{ICL1} (derepressed by glucose depletion and induced by ethanol), P_{ADH1} (induced by glycerol and ethanol) and P_{GUT1} (induced by glucose, glycerol and ethanol) were patented (Cregg and Tolstorukov, 2012). Another methanol free alternative, P_{THI11} that is derived from a gene involved in thiamine biosynthesis, reached 70% of P_{GAP} activity in the absence of thiamine (Stadlmayr et al., 2010). P_{AOD} (alternative oxidase) is a moderate promoter induced by glucose and repressed by methanol or upon glucose depletion, but with limited applicability as it ceased its functioning at about half of the maximal activity of P_{GAP} (Kern et al., 2007).

For production of toxic proteins, regulated expression is favourable but it requires at least two production phases, the first is to obtain high cell yield under repressing conditions and the second is to perform expression of heterologous proteins, resulting in more handling efforts and longer process times. For non-toxic

proteins, constitutive expression simplifies bioprocess operations and provides higher space-time yield. Since the glyceraldehyde 3-phosphate dehydrogenase promoter (P_{GAP}) is strong and constitutive (Çalık et al., 2015), it can reach similar expression levels as P_{AOX1} (Waterham et al., 1997) and performs better under low to limited oxygen transfer as reported by Güneş and Çalık (2016), where hypoxic conditions were claimed as superior by Baumann et al. (2010) without oxygen transfer or uptake rates measurements. The highest expression levels are reached when the cells are grown on glucose but also comparable product yields are obtained with glycerol. The sorbitol dehydrogenase promoter P_{SDH} was asserted as an alternative to P_{GAP} as its strength was reported to be similar in heterologous protein production (Periyasamy et al., 2013). Phosphoglycerate kinase promoter (P_{PGK1}), involved in the glycolysis and gluconeogenesis is a relatively weak promoter showing 10% activity of P_{GAP} (Stadlmayr et al., 2010). Translation elongation factor 1 alpha is an essential module of the eukaryotic translation machinery, thus its promoter P_{TEF1} exhibits growth associated regulatory pattern with high activity during exponential phase and reduced performance in the stationary phase (Ahn et al., 2007). Under carbon limited conditions, almost twofold higher expression levels were obtained with P_{TEF1} compared to P_{GAP} (Ahn et al., 2007); however, with different reporter genes and at different cultivation times, similar or less expression levels were observed with P_{TEF1} compared to P_{GAP} (Stadlmayr et al., 2010). Other constitutive promoters obtained from the glycolytic pathway are P_{ENO1}, P_{GPM1} and P_{TPI1}, and those obtained from stress response metabolism are P_{HSP82}, P_{KAR2} and P_{SSA4} (Table 2.1). P_{PET9} exhibited different expression levels with different reporter genes, in comparison with P_{GAP} (Stadlmayr et al., 2010). Pyruvate kinase (P_{PYK}) and pyruvate decarboxylase (P_{PDC}) promoters were determined around pyruvate node and in high cell density fermentation operations they performed better recombinant human growth hormone synthesis than P_{GAP} (Massahi and Çalık, 2018). Promoter engineering studies and the design of synthetic core promoters have also widened the promoter toolbox for *P. pastoris* as they created novel engineered or synthetic promoters that cover a wide range of

strength and have different regulation patterns (Ata et al., 2017; Hartner et al., 2008; Portela et al., 2017; Vogl et al., 2014).

For the co-expression of multiple proteins simultaneously, it is important to identify alternative promoters covering a wide range of strength and showing different regulatory patterns. Mattanovich et al. (2004) suggested that if correct folding or secretion steps are rate limiting in expression of the active products, weaker expression levels can be favourable. P_{PEX8} (peroxin 8) (Liu et al., 1995) and P_{YPT1} (a GTPase function in secretion) (Sears et al., 1998) are moderate inducible and constitutive alternative promoters, when low expression levels are needed (Prielhofer et al., 2017; Vogl et al., 2016).

Few promoters from other microorganisms have been also employed in r-protein production in *P. pastoris* successfully. P_{ADH2} from *P. stipitis* (Chien and Lee, 2005; Passoth and Hahn–Hägerdal, 2000), and P_{PDC1} from *K. lactis* (Camattari et al., 2007) were induced in *P. pastoris* by decreasing dissolved oxygen concentration. *S. cerevisiae* P_{CUP1} was also utilized in *P. pastoris* expression system, where expression was induced by copper (Koller et al., 2000).

Table 2.1 Naturally occuring promoters commonly used in *P. pastoris*

Promoter	Protein	Regulation	Reference
Constitutive			
GAP	Glyceraldehyde 3-phosphate Dehydrogenase		(Waterham et al., 1997)
GCW14	Potential glycosyl phosphatidyl inositol (GPI)-anchored protein		(Liang et al., 2013)
GPM1	Phosphoglycerate mutase		(Stadlmayr et al., 2010)
HSP82	Cytoplasmic chaperone (Hsp90 family)		(Stadlmayr et al., 2010)
ILV5	Acetohydroxy acid isomeroreductase		(Naatsaari et al., 2012)
KAR2	ER resident hsp70 chaperone (binding protein BiP)		(Stadlmayr et al., 2010)
TEF1	Translation elongation factor 1 α		(Ahn et al., 2007)
PDC	Pyruvate decarboxylase		(Massahi and Çalık, 2018)
PET9	Major ADP/ATP carrier of the mitochondrial inner membrane		(Stadlmayr et al., 2010)
PGK1	3-phosphoglycerate kinase		(de Almeida et al., 2005)
PYK	Pyruvate kinase		(Massahi and Çalık, 2018)
SDH	Sorbitol dehydrogenase		(Periyasamy et al., 2013)
SSA4	Heat shock protein		(Stadlmayr et al., 2010)
TPI1	Triose phosphate isomerase		(Stadlmayr et al., 2010)
YPT1	GTPase involved in secretion		(Sears et al., 1998)

Table 2.1 (continued)

Promoter	Protein	Regulation	Reference
Inducible			
ADH1	Alcohol dehydrogenase 1	Ethanol, glycerol	(Cregg and Tolstorukov, 2012)
AOD	Alternative oxidase	Glucose, antimycin A	(Kern et al., 2007)
AOX1	Alcohol oxidase 1	Methanol	(Tschopp et al., 1987)
AOX2	Alcohol oxidase 2	Methanol	(Cregg et al., 1989; Koutz et al., 1989)
DAS	Dihydroxyacetone synthase	Methanol	(Ellis et al., 1985; Tschopp et al., 1987)
ENO1	Enolase	Glucose, glycerol, ethanol	(Cregg and Tolstorukov, 2012)
FLD1	Formaldehyde dehydrogenase	Methanol, methylamine, choline	(Shen et al., 1998)
G1	High affinity glucose transporter	Glucose limitation	(Prielhofer et al., 2013)
G6	Putative aldehyde dehydrogenase	Glucose limitation	(Prielhofer et al., 2013)
GUT1	Glycerol kinase	Glycerol	(Cregg and Tolstorukov, 2012)
ICL1	Isocitrate lyase	Ethanol, derepressed with glucose depletion	(Menendez et al., 2003)
MET3	ATP sulfurylase	L-methionine (repressible)	(Delic et al., 2013)
PEX8	Peroxisomal matrix protein	Methanol, oleate	(Cereghino and Cregg, 2000; Liu et al., 1995)
PIS1	Phosphatidylinositol synthase	Zinc (repressible)	(Delic et al., 2013)
PHO89 **or** ***NSP***	Putative Na+/phosphate symporter	Phosphate limitation	(Ahn et al., 2009)

Table 2.1 (continued)

Promoter	Protein	Regulation	Reference
SER1	3-phosphoserine aminotransferase	L-serine (repressible)	(Delic et al., 2013)
THI11	THI11	Thiamine (repressible)	(Stadlmayr et al., 2010)
THR1	Homoserine kinase	L-threonine, L-valine, L-leucine and L-isoleucine (repressible)	(Delic et al., 2013)

2.1.1.1 *P. pastoris* Alcohol Dehydrogenase 2 Promoter

P. pastoris alcohol dehydrogenase 2 promoter (P_{ADH2}) was reported as an inducible promoter by ethanol and glycerol (Cregg and Tolstorukov, 2012). Functional analysis of *ADH2* showed that it is the only gene responsible for ethanol utilization in *P. pastoris* while *adh2* knock-out strains can still produce ethanol (Karaoglan et al., 2016a). Recombinant-protein expression potential of P_{ADH2} was investigated with pilot-scale bioreactor experiments, and recombinant xylanase production by P_{ADH2} outcompeted commonly used strong inducible P_{AOX1} and constitutive P_{GAP} expression levels (Karaoglan et al., 2016b).

2.1.1.2 *P. pastoris* Alcohol oxidase 1 (AOX1) Promoter

Alcohol oxidase 1 (AOX1) promoter (P_{AOX1}) is the most widely used *P. pastoris* promoter for r-protein production. AOX1 is the first enzyme in the methanol utilization pathway of *P. pastoris* and expressed at high levels; however, P_{AOX1} is strongly repressed when *P. pastoris* is grown on glycerol, ethanol, or glucose (Inan and Meagher, 2001). However, P_{AOX1} based expression platforms have several drawbacks due to toxicity and flammability of inducer carbon and energy source methanol. Methanol is mainly obtained from petrochemical sources that makes the r-protein unsuitable to be used for production of specific foods or as food additives.

Furthermore, methanol utilization generates the toxic by-product formaldehyde, besides hydrogen peroxide (H_2O_2) that leads oxidative stress and proteolytic degradation of some r-proteins (Sinha et al., 2005; Zhang et al., 2007). In order to clarify transcriptional regulation of outstanding P_{AOX1} and the MUT pathway genes, *cis*-acting elements and *trans*-acting factors were investigated. Deletion of *P. pastoris HXT1* gene coding for a hexose transporter led to glucose and fructose induced P_{AOX1} expression; however, induction levels were sharply dropped during later growth phase (Zhang et al., 2010). *P. pastoris* Gss1 (designated from GlucoSe Sensor) was found to be involved in glucose but not in ethanol catabolite repression of P_{AOX1} and alcohol oxidase activity was detected in *gss1Δ* cells grown on glucose (Polupanov et al., 2012).

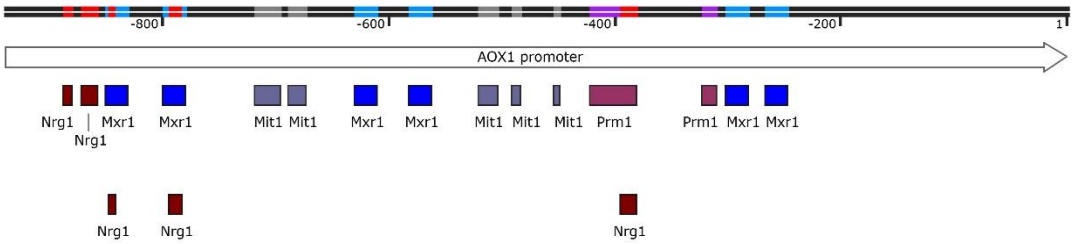


Figure 2.1 Identified *cis*-acting elements on P_{AOX1} represented with their corresponding TFs

P. pastoris Mxr1 (methanol expression regulator 1) was identified as the global regulator of genes involved in the MUT pathway (Lin-Cereghino et al., 2006). *P. pastoris mxr1Δ* mutants were not grow on methanol and formed functional peroxisome, also the level of the MUT pathway enzymes were severely affected (Lin-Cereghino et al., 2006). Six different Mxr1 binding sites were identified on P_{AOX1} using electrophoretic mobility shift assay (EMSA) and DNAse I footprinting and these regions were named as P_{AOXI} Mxr1 response elements (Mxres) (Kranthi et al.,

2009) (Figure.4.16). 5′CYCC3′ (5′GGRG3′ in the complementary strand) motif was found to be conserved in all the six P_{AOX1} Mxres (Kranthi et al., 2009).

Evaluation of P_{AOX1} regulation revealed a Cys_2His_2 zinc finger transcriptional repressor, Nrg1 involves in glucose and glycerol repressions. Based on electrophoretic mobility shift assay (EMSA) and DNAse I footprinting assay, five different DNA binding regions of Ngr1 were identified on P_{AOX1}, the two of which are overlapping with Mxre1 and Mxre2 motifs (Wang et al., 2016a). Nrg1 DNA binding motif contains CYCCNY consensus sequence which is also the binding site of Mxr1 (Adr1). Quantitative real-time PCR results demonstrated that *P. pastoris* Nrg1 involves repression of many methanol utilization and peroxisome biogenesis genes in 0.02 % glucose and 1 % (v/v) glycerol. The *Δnrg1* cells exhibited growth defect on glycerol, methanol, and more severely on glucose; and showed slight AOX expression in 0.02 % glucose and 1 % (v/v) glycerol, but not with 1% (w/v) glucose (Wang et al., 2016a).

Mig1 and Mig2 are catabolite repressors, double deletions of *MIG1* and *MIG2* genes lead to activation of P_{AOX1} in glycerol but not glucose containing media while it damaged the cell growth in both conditions (Wang et al., 2017). RNA-seq data with Δ*mig1*Δ*mig2* mutant grown on glycerol as sole carbon source revealed that the MUT pathway was markedly improved with the AOX1 mRNA level increased by more than 30 times compared to wild-type strains (Shi et al., 2018). Transcription levels of the genes *MIT1* and *Cat8-1* were upregulated 2.17-fold and 2.38-fold, respectively (Shi et al., 2018). Since there is no Mig1 or Mig2 binding motif on P_{AOX1}, Shi et al. (2018) suggested that repression of P_{AOX1} can be controlled via regulation of Mit1 level.

Rop (repressor of phosphoenolpyruvate carboxykinase, PEPCK) is a biotin starvation and methanol inducible zinc finger transcription factor acts as a negative regulator on PEPCK in *P. pastoris* cells grown in biotin-deficient glucose-ammonium medium (Kumar and Rangarajan, 2011). Rop was identified as a negative transcriptional regulator of P_{AOX1} that binds the same consensus motifs recognized by Mxr1 (Adr1);

furthermore, Rop and Mxr1 compete for the same TFBSs and antagonistically regulate the MUT pathway (Kumar and Rangarajan, 2012). *P. pastoris* cells lacking *ROP* gene performed increased expression levels of AOX1 and improved the growth when the cells cultured in a nutrient-rich medium containing yeast extract, peptone, and methanol (YPM). However, no difference on the AOX1 expression and the cell growth were observed compared to the wild-type cells when *ropΔ* mutants were grown in minimal medium containing yeast nitrogen base and methanol (YNBM) (Kumar and Rangarajan, 2012). Subcellular localization studies demonstrated that Rop translocates from cytosol to nucleus in the YPM grown cells but not that of YNBM (Kumar and Rangarajan, 2012).

P. pastoris Prm1 (Trm1) is a zinc finger transcriptional activator of AOX1 and the MUT pathway genes and Prm1 translocolizes to nucleus when the cells are grown on YPM and YNBM (Sahu et al., 2014). *Δprm1* mutants showed growth defects on methanol but not on other non-fermentable carbon sources such as glycerol, oleic acid, and acetate (Sahu et al., 2014). The effect of *PRM1* gene deletion on the expression of AOX1 is more pronounced when the cells were cultured in YNBM than those cultured in YPM; however *mxr1Δ* mutants have pronounced effect on the expression levels of AOX1 in both conditions (Sahu et al., 2014).

Another Zn(II)2Cys6-type zinc cluster transcriptional activator of P_{AOX1} is methanol induced transcription factor 1 (Mit1) that involves in regulation of MUT pathway genes but doesn't participate peroxisome proliferation and transportation of peroxisomal proteins during methanol metabolism (Wang et al., 2016b). Mit1, Mxr1, and Prm1 are bound to P_{AOX1} at different sites and cooperatively activate P_{AOX1} through a cascade response to methanol. Mxr1 is mainly active and essential for derepression of P_{AOX1}, while Mit1 and Prm1 mainly respond to methanol induction (Wang et al., 2016b). Methanol induction signal was blocked in *mxr1Δ* cells due to a failure in derepression, accordingly overexpression of Mit1 and Prm1 couldn't rescue P_{AOX1} activity (Wang et al., 2016b). During methanol induction Prm1 activates

expression of Mit1 and itself, while Mit1 represses Prm1 expression probably as a feed-back loop control.

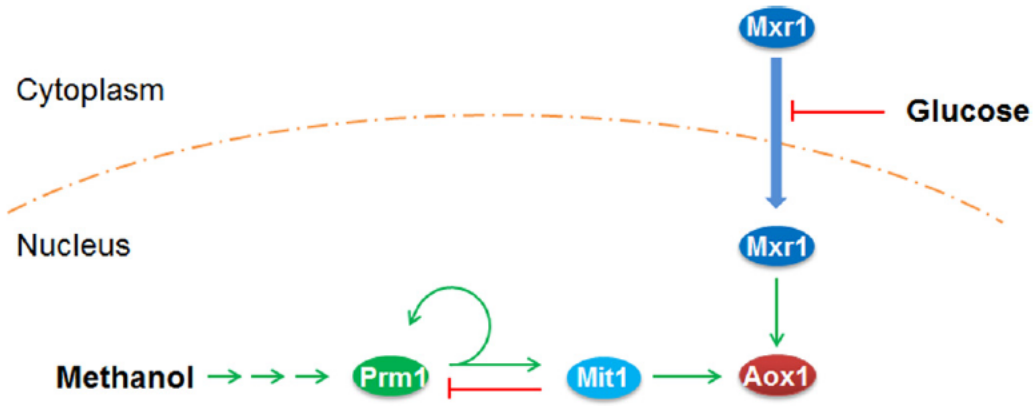


Figure 2.2 Regulatory model of P_{AOX1} activation by Mxr1, Prm1, and Mit1. In the presence of glucose, P_{AOX1} is repressed because Mxr1 located to cytoplasm and failed to perform derepression activity. In response to sole carbon source methanol, Mxr1 is relocated to the nucleus and activates P_{AOX1}. The methanol induction signal transmitted from Prm1 to Mit1 and Mit1 activates P_{AOX1}. Throughout the process, Prm1 induces expression of itself and Mit1, but Mit1 represses the expression of Prm1 (Wang et al., 2016b)

Due to problems caused by methanol, recent research has focused on development of methanol-free expression systems by P_{AOX1}. Kinases are well characterized and crucial elements of cell signaling since phosphorylation and de-phosphorylation processes are essential for many biological activities. Engineered *P. pastoris* strains were developed on the basis of two kinase genes, namely *GUT1* and *DAK;* and, in those cells P_{AOX1} could be activated by glycerol and dihydroxyacetone (DHA), respectively (Shen et al., 2016). *Δdak* mutant induced by DHA reached to 50-60% of methanol induced traditional system (Shen et al., 2016).

P. pastoris glycerol transporter Gt1 functions as an active transmembrane transporter for glycerol uptake into the cells. Gt1 involves in glycerol repression of P_{AOX1}, in *gt1Δ* mutants P_{AOX1} becomes constitutively active in glycerol containing media (Zhan et al., 2016). A transcriptional control between Mxr1(Adr1) and Gt1 was also identified; Mxr1 regulates transcriptional activity of P_{GT1} by direct binding whereas overexpression of Gt1 represses the expression of Mxr1 and Aox1, and thus the methanol metabolism (Zhan et al., 2017).

Methanol-free expression system by P_{AOX1} was developed based on transcriptional engineering principles; identified three transcription repressors of P_{AOX1} namely Mig1, Mig2 and Nrg1 genes were deleted and, one transcription activator Mit1 gene was overexpressed. Repression of P_{AOX1} by glycerol was abolished in designed *P. pastoris* Δ*mig1*Δ*mig2*Δ*nrg1* MIT1 mutant cell, and in glycerol containing medium the mutant strain reached to 77% eGFP expression levels of the methanol induced wild-type strain (Wang et al., 2017). Another transcriptional engineering strategy relied on derepressed overexpression of certain acitvators namely, Mit1, Mxr1 and Prm1 that led to activation of P_{AOX1} under derepression conditions (Vogl et al., 2018). In deep well plate screening conditions, under derepression conditions P_{AOX1} performed up to 44% and 68% of methanol induced levels based on dTomato and eGFP reporter protein levels, repectively, in *P. pastoris* strain overexpressing *MIT1* and *MXR1* (Vogl et al., 2018). When this system was tested with controlled pilot-scale biorecator experiments, without methanol induction Mit1-overexpression strain exhibited a higher yield than the methanol induced parental strain (Vogl et al., 2018).

2.2 Eukaryotic Transcription by RNA Polymerase II

In eukaryotes, all protein coding genes are transcribed by RNA Polymerase II (Pol II). The general model for eukaryotic transcription activation is an ordered recruitment of several transcription activators and repressors (collectively known as specific transcription factors) and general transcription factors to promoter that leads the formation of preinitiation complex (Hahn and Young, 2011). In the early phases

of activation, promoter-specific factors bind presumably with the aid of general chromatin remodeling factors (Li et al., 2007). The activation domains (ADs) and DNA-binding domains of activators allow site-specific coactivator contact and promoter recognition, respectively. For activated transcription, transcriptional activators that bound to sequence-specific DNA positions, require additional proteins, named as coactivators. Coactivators are not required for basal transcription and do not bind to specific DNA sequences by themselves (Malik and Roeder, 2000). Coactivators function to counteract the transcriptionally repressive effect of chromatin and/or participate to basal transcriptional machinery (Featherstone, 2002). Coactivators that are associated with basal transcriptional machinery encompass components of Mediator complex, TATA-binding protein- (TBP-) associated factors (TAFs) and coactivators that modify the structure of chromatin include complexes with histone acetyltransferase (HAT) activity and ATP-dependent chromatin-remodeling complexes such as SWI/SNF, that alter the twist and writhe of nucleosomal DNA to affect its accessibility to DNA-binding proteins (Havas et al., 2000).

Mediator, mediator of RNA Polymerase II transcription, is a large multisubunit protein complex (comprising 25 subunits in *S. cerevisiae*) with modular organization that is necessary for transcription by RNA Polymerase II (Petrenko et al., 2017). The main function of the Mediator complex is to transduce signals from transcription activators to transcription machinery that is assembled as the preinitiation complex (PIC) to regulate transcription initiation (Soutourina, 2017). The mediator complex mediates interaction between activators and carboxy-terminal domain (CTD) of RNA Polymerase II (Malik and Roeder, 2000).

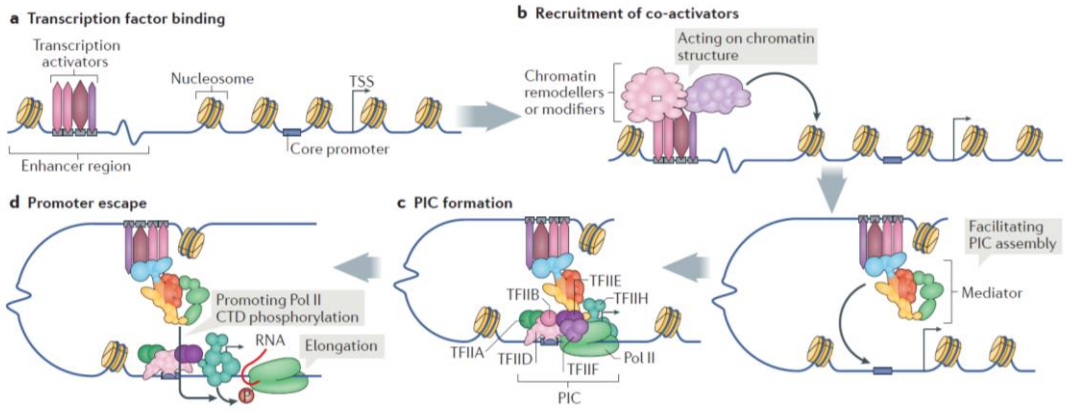


Figure 2.3 Eukaryotic transcription activation by RNA Polymerase II
a) Transcription activation starts with binding of transcription factors (in this case transcription activators) on upstream activation sequences (enhancers in metazoan). Upstream activation sequences in yeast can be located hundreds of bases apart from the core promoter. TSS: Transcription start site **b)** Chromatin remodelers or modifiers are then recruited by activators to increase the accessibility of chromatin for other factors. Other transcriptional co-activators are then recruited that directly influence the basal transcriptional machinery, the so-called preinitiation complex (PIC). In general, transcriptional co-regulators including Mediators transmit the regulatory signals from certain transcription factors to the PIC. **c)** The PIC is assembled at the core promoter including 12 subunits of RNA Polymerase II (Rpb1-12) and general transcription factors (GTFs; including TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH). Mediators recruit different PIC component and/or facilitate their stability. **d)** Mediator stimulates cyclin-dependent kinase 7 (CDK7, also known as Serine/threonine-protein kinase Kin28), to phosphorylate the CTD of Pol II at Ser5 that is required for Pol II escape from the promoter and for the transition from the initiation to the elongation phase (Soutourina, 2017).

2.2.1 General Transcription Factors

General transcription factors (GTFs) are proteins or protein complexes which are required for RNA Pol II promoter recognition, its function, and recruitment. Transcription preinitiation complex (PIC) that assembles on core promoter, is a large protein complex includes GTFs, RNA Pol II and Mediator. GTFs include TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH and they constitute basal transcriptional machinery with RNA Pol II. For the assembly of the PIC, TFIID is the first GTF to associate with promoter DNA. TFIID consists of TATA-binding protein (TBP) and 14 TBP-associated factors (TAFs). Next, TFIIA and TFIIB interacts with the DNA-TBP complex directly and stabilize the DNA-TBP complex and recruit Pol II that is accompanied by TFIIF (Hampsey, 1998). TFIIF contributes to stabilization of PolII in the PIC, setting transcription start site, transcription elongation and stabilization of RNA-DNA hybrid in the early elongation complexes (Thomas and Chiang, 2006). Then, PIC assembly is completed by TFIIE and TFIIH binding in order. TFIIE and TFIIH function in separation and stabilization of promoter DNA strands during transition into the transcriptional active open complex state. PIC is assembled in a stepwise manner with protein-DNA and protein-protein interactions. After that, the PIC undergoes a conformational change into the open complex formation, scans for a correct transcription start site (TSS) and then initiates active elongation (Kostrewa et al., 2009).

2.3 Transcription Factors Involved in Regulation of Genes by Ethanol

Yeast gene expression regulation involves coordinated action of multiple proteins. Promoters mostly have binding sites for more than one transcription factor (TF) and respond to environmental or intracellular signals by employing different combinations of non-redundant TFs, a process called as combinatorial transcriptional control. Combinatorial transcriptional control is accomplished at multiple steps encompassing cooperative activator/repressor binding and cooperative activation/repression by specific set of TFs including activators and repressors. In this

PhD thesis study, three different TFs namely Adr1 (Mxr1), Cat8, and Aca1/2 were investigated; a detailed information about the TFs are provided, as follows.

2.3.1 Adr1

S. cerevisiae Adr1, alcohol dehydrogenase 2 (ADH2) synthesis regulator, is a 1323 amino acids protein that composed from different functional domains including nuclear localization signal which encompasses amino acids 1 to 16 (Thukral et al., 1989); a zinc finger DNA binding domain that binds to P_{ADH2} upstream activating sequence 1 (UAS1) encompassing amino acids 75 to 161 (Bowers et al., 1999); and four transcriptional activation domains (TADs): TADI, amino acids 76 to 172, TADII, amino acids 263 to 359; TADIII, amino acids 359 to 571; and, TADIV, amino acids 642 to 704 (Chiang et al., 1996; Cook et al., 1994).

Adr1 activates catabolic pathways that are essential for growth on non-fermentable carbon sources (derepression). Adr1 is necessary for baker's yeast growth on ethanol, glycerol, and oleate; and, peroxisome function and biogenesis (Cook et al., 1994). Adr1 has a pivotal role in the activation of glucose repressible genes especially *S. cerevisiae* alcohol dehydrogenease 2 promoter, $P_{ADH2.}$

The level of Adr1 in the cell is controlled with several mechanisms: *i)* through transcriptional control of Adr1 by cyclic-AMP dependent protein kinase (cAPK) (Dombek and Young, 1997), and *ii)* post-translational modification of Adr1 through its DNA-binding domain (Kacherovsky et al., 2008; Sloan et al., 1999). Sucrose non-fermenting-1 (Snf1) protein kinase, yeast homologue of the mammalian adenosine monophosphate (AMP)-activated protein kinase, is not necessary for Adr1 expression (Dombek and Young, 1997), but essential for Adr1 regulation under glucose depleted conditions. Under derepressed conditions, Snf1 is requirred for promoter binding of Adr1 (Young et al., 2002), binding is regulated via phosphorylation of Ser-98 in DNA binding domain (Kacherovsky et al., 2008) and by Snf-1 dependent histone H3 hyperacetylation of promoter nucleosome (Verdone, 2002).

Negative control by glucose is exerted on Adr1 activity by several mechanisms, including control of Adr1 expression level (Blumberg et al., 1988), DNA binding of Adr1 (Kacherovsky et al., 2008; Sloan et al., 1999), and transcriptional activation step by Adr1 (Ratnakumar et al., 2009; Tachibana et al., 2007). cAPK phosphorylates Ser-230 position of Adr1 in vitro but cAPK activity is not necessary for in vivo glucose inhibition or phosphorylation (Ratnakumar et al., 2009). Deletion of histone deacetylase genes leads to hyperacetylation of the histone tails and constitutive promoter binding of the Adr1 and Cat8 activators which normally inhibited by glucose; chromatin remodelling and preinitiation complex formation, but not transcription (Tachibana et al., 2007; Verdone, 2002). Activation of Snf1 and activated Adr1stimulate the poised preinitiation complex and allow to complete escape from glucose repression (Tachibana et al., 2007).

Yeast 14-3-3 (Bmh) proteins form a family of highly conserved proteins which can interact with more than 200 different mostly phosphorylated proteins, are regulatory components of multiple signaling pathways and have an effect on protein localization or activity via binding to phosphorylated domains (van Heusden and Steensma, 2006). There are two functionally redundant Bmh isoforms: Bmh1 and Bmh2. Bmh inhibits Adr1 activation function by direct binding to its Ser-230 phosphorylated regulatory domain at a post-promoter binding step (Braun et al., 2013; Parua et al., 2010). Bmh inhibits the assembly and function of PIC via inhibiting a step between Meidator recruitment and PIC activation (Parua et al., 2014). Constitutively active $ADR1^{C}$ alleles, such as Adr1-S230A prevents Bmh binding via reducing phosphorylation at Ser-230 or by eliminating another Bmh-Adr1 interaction (Parua et al., 2014) and constitutively activates *S. cerevisiae* ADH2 promoter in the presence of repressing carbon source glucose (Denis et al., 1992). Adr1 phosphorylation by S230 kinase and dephosphorylation by Snf1 might control its activity on other nutrient-signaling pathways, as well (Ratnakumar et al., 2009).

Adr1 binds and activates transcription of *S. cerevisiae* P_{ADH2} from two sites in a perfect 22-bp inverted repeat, called UAS1 (Upstream activation sequence 1). Two C_2H_2 zinc fingers and a region amino terminal to the fingers involve in promoter binding of Adr1 (Thukral et al., 1991). Amino terminal to the fingers region of Adr1 is crucial for phosphate contact in the core region of UAS1. When the core 6 bp were deleted, in vitro only one Adr1 molecule can bind and in vivo activation was severely damaged. In *S. cerevisiae*, based on the constraints on the orientation and spacing of binding motifs that affect activation in vivo, the preferred Adr1 binding consensus sequence, TTGG(A/G)GA, was found in both halves of the inverted repeat (Cheng et al., 1994).

P. pastoris Adr1, generally known as Mxr1 (methanol expression regulator 1) is a homologue of *S. cerevisiae* Adr1 (alcohol dehydrogenase 2 synthesis regulator), that has acquired new functions and lost some others through evolution due to changes in the range of genes under its control (Lin-Cereghino et al., 2006). *P. pastoris* Adr1 (Mxr1) is a positive regulator *trans*-acting factor of methanol utilization pathway and peroxisome biogenesis (PEX) genes, *adr1* knock-out strains demonstrated growth defect on methanol and oleic acid (Lin-Cereghino et al., 2006). Upon growth on methanol or other gluconeogenic substrates, Adr1 localization shifts to nucleus and especially binds to P_{AOX1} and other methanol pathway and PEX genes (Lin-Cereghino et al., 2006). *P. pastoris adr1*Δ strains couldn't show AOX1 expression after 8 h induction by methanol according to the results of Western and Northern Blot analysis (Lin-Cereghino et al., 2006).

2.3.2 Cat8

Cat8 is a zinc cluster transcriptional activator protein that is necessary for growth of *S. cerevisiae* on non-fermentable carbon sources (Hedges et al., 1995). Expression level of Cat8 and gene activation by Cat8 are subject to glucose regulation and functional Snf1 protein kinase is necessary for this regulation (Randez-Gil et al., 1997). Cat8 is an activator on the expression of genes involved in gluconeogenesis,

ethanol utilization, and diauxic shift from fermentation to oxidative metabolism (Haurie et al., 2001; Rahner et al., 1999).

In the yeast *S. cerevisiae*, structural genes for gluconeogenesis and the glyoxylate cycle are regulated by a carbon source responsive element (CSRE) located as an upstream activating site (UAS) on their promoter sequences (Caspary et al., 1997; Proft et al., 1995; Schöler and Schüller, 1994). Based on known gluconeogenic structural genes preliminary 11 bp CSRE consensus motif was derived as CCRTYSRNCCG (Roth and Schüller, 2001). Although, all CSRE variants involve in regulated expression, activation performance of each sequence differs significantly which lead to a distinction between ‘’weak’’ and ‘’strong’’ CSREs (Roth and Schüller, 2001).

Transcriptional activation by CSRE depends on an intact and functional Cat8 TF with a N-terminal binuclear C_6 zinc cluster motif (Rahner et al., 1996). Cat8 and Sip4 TFs are members of Gal4 family of activators for which interaction with symmetrically disposed CGG triplets is a mutual characteristic (Schjerling and Holmberg, 1996). Dimerization of zinc cluster proteins leads to interaction with two CGG motifs in a direct (CGG-CGG), inverted palindrome (CGG-CCG) or everted palindrome (CCG-CGG) found on the target sequence (Hellauer et al., 1996). Actually, both Cat8 and Sip4 can bind to the CSRE, even in the absence of the each other; however, under glucose derepression condition Cat8 and Sip4 unequally contribute to the activation of CSRE-dependent genes, 85% and 15%, respectively (Hiesinger et al., 2001). Consensus Cat8 binding sequence for efficient gene activation was derived as YCCNYTNRKCCG, whereas more specific motif is necessary for activation by Sip4, that is TCCATTSRTCCGR (Roth et al., 2004). Cat8 and Sip4 genes are transcriptionally controlled by carbon source, and gene activation by Cat8 and Sip4 depends on Snf1 protein kinase function (Lesage et al., 1996; Randez-Gil et al., 1997). Transcriptional control of Sip4 is additionally subjected to a CSRE that is bound primarily by Cat8 but only weakly by Sip4 itself; on the contrary, Sip4 is not

necessary for Cat8 biosynthesis and transcriptional activation (Hiesinger et al., 2001). Due to this regulation, biosynthetic derepression of Sip4 under non-fermentable carbon source growth succeeds that of Cat8 by about 90 min (Roth et al., 2004).

In *S. cerevisiae* $P_{ADH2,}$ a second upstream activation site, UAS2 that contains a distantly related variant of a CSRE was identified, deletion of UAS2 binding sequence decreased the strength of P_{ADH2} to 50% wild-type promoter and it was found that UAS2 could only be activated by Cat8 while Sip4 couldn't show any significant influence (Walther and Schüller, 2001). All these findings suggest that Cat8 and Sip4 are not isofunctional activators. P_{ADH2} UAS2 element was fused with basal ICL1 promoter that is devoid of its natural regulatory sequences, under derepressed conditions (0.2% glucose + 2% ethanol) a synthetic promoter variant that have 1 copy insertion of UAS2 showed 2-fold activation; while 2 and 3 copy insertions of UAS2 into synthetic promoter resulted in 80- and 312-fold activation, respectively that suggests significant synergism between UAS2 elements (Walther and Schüller, 2001). Compared to a wild-type reference strain, under derepressed conditions *adr1Δ*, *cat8Δ* and *sip4Δ* mutant strains performed 46.8%, 1.3%, and 157% activation by UAS2 element, which is fused with basal *ICL1* promoter, respectively (Walther and Schüller, 2001).

S. cerevisiae cat8Δ mutants could not utilize non-fermentable carbon sources such as ethanol, lactate and acetate (Rahner et al., 1996) but no obvious growth defect was observed for *sip4Δ* mutants (Lesage et al., 1996). In *Kluyveromyces lactis*, *cat8Δ* mutant showed severely impaired growth on ethanol, lactate and acetate but not on glycerol, which suggests that in *K. lactis* only glyoxylate shunt is regulated by Cat8 (Georis et al., 2000). Contrariwise, *Candida albicans* mutants lacking *CAT8* have similar phenotype to wild-type in terms of the gluneogenesis, glyoxylate shunt, and ethanol utilization pathway enzymes (Ramírez and Lorenz, 2009). *cat8Δ* modification significantly improved glucose alcoholic fermentation in *S. cerevisiae* (Watanabe et al., 2013) and triggered respirofermentative metabolism in Crabtree-negative yeast

Pichia guilliermondii (Qi et al., 2014). In *Ogataea (Hansenula) polymorpha cat8Δ* mutants ethanol accumulation increases from xylose fermentation while glucose alcoholic fermentation remains the same (Ruchala et al., 2017).

2.3.3 Combinatorial Regulation of Non-Fermentable Carbon Source Utilization Genes

Combinatorial control mechanism is exerted on many genes in response to fermentable and non-fermentable carbon sources. Depletion of glucose concentration leads to the activation of AMP-activated protein kinase homolog, Snf1 (Carlson, 1999). Snf1 targets multiple regulatory proteins including transcription factors, co-activators and proteins active in post-transcriptional gene regulation (Braun et al., 2014). Snf1 mainly activates two TFs: Adr1, by inducing dephosphorylation (Ratnakumar et al., 2009); and Cat8, by direct phosphorylation (Rahner et al., 1996; Randez-Gil et al., 1997). Additionally, Snf1 promotes DNA binding of Adr1 via stimulating SAGA-associated histone acetyltransferase Gcn5 (Abate et al., 2012). Adr1 and Cat8 cooperatively regulate genes that are functional for the utilization of non-fermentable carbon sources e.g. ethanol (*ADH2*, *ALD4* and *ACS1*) and lactate (*CBY2* and *JEN1*) (Tachibana et al., 2005). Genes encoding peroxisomal proteins and β-oxidation enzymes are coregulated by Adr1 and Oaf1/Pip2 TFs (Ratnakumar and Young, 2010). The regulation of peroxisomal and β-oxidation genes by Snf1 is mediated by only Adr1 (Ratnakumar and Young, 2010). In the combinatorial activation processes by these two TFs either Adr1 and Cat8 or Adr1 and Oaf1/Pip2, promoters are co-occupied by them and in the absence of either factor 10-fold or greater reduction in the mRNA abundance was observed (Tachibana et al., 2005; Walther and Schüller, 2001). For *S. cerevisiae* ADH2 promoter, Adr1 stimulates Cat8 binding, but this is not always the case for other genes that require both factors (Tachibana et al., 2005).

2.3.4 Aca1/Aca2

S. cerevisiae Aca1 and Aca2 (Cst6) TFs are belong to ATF/CREB family with a basic leucine zipper domain and in vitro, they bind to TGACGTCA motif (Garcia-Gimeno and Struhl, 2000). Aca1 and Aca2 are paralogs that are derived from whole genome duplication. Aca2 activates transcription via binding to DNA as a homodimer or a heterodimer together with Aca1 (Garcia-Gimeno and Struhl, 2000). Phenotypic analyses revealed Aca2 is important for a variety of cellular processes including growth on non-optimal carbon sources and resistance to various drugs. *ACA2* knock-out strains showed retarded or no growth on respiratory carbon sources like ethanol, glycerol and raffinose (Garcia-Gimeno and Struhl, 2000). Aca2 is also found to be a regulator of oleate responsive genes; oleate induces phosphorylation of Aca2, at the same time this phosphorylation dampens Aca2 activity (Saleem et al., 2010). Mutant Aca2 that is constitutively non-phosphorylated led to increased expression of β-oxidation genes (Saleem et al., 2010). A well-studied direct target of *S. cerevisiae* Aca2 and its orthologue *C. albicans* Rca1 is *NCE103* that encodes a carbonic anhydrase catalyzing conversion of CO_2 to HCO_3 which is an essential substrate for the cell's carboxylation reactions to sustain gluconeogenesis and lipogenesis (Cottier et al., 2012). Under low-CO_2 conditions, Aca2 binds and activates P_{NCE103}. In *S. cerevisiae* genome-wide Aca2 binding-site mapping revealed that regulatory function of Aca2 is condition dependent; Aca2 binds promoter regions of 59 genes with various biological functions when cells are cultivated on ethanol but almost no DNA-binding event occurs in glucose grown cells (Liu et al., 2016). Since *aca2Δ* mutant strains exhibit hypersensitivity to ethanol and oxidative stresses; and target genes of Aca2 mainly overlap with other stress-responsive TFs' target genes, Aca2 is assigned as a stress responsive transcriptional regulator (Liu et al., 2016). In *aca2Δ* mutant cells, apart from *NCE103* all other targets showed moderate decrease in their expression level, thus growth defect on ethanol mainly attributed to decrease in Nce103 (Liu et al., 2016).

2.4 Protein-DNA Interactions

Regulatory proteins can arrive at their target genomic sites faster than diffusion-controlled rates, and they also exhibit remarkable affinity for non-specific DNA stretches. This facilitated target search involves two-step binding mechanism; first, three-dimensional bulk diffusion to any site on the DNA and then one-dimensional hopping along the DNA to locate their specific targets (Berg et al., 1981). Four type of protein translocation mechanism between DNA sites were proposed. These are: ***i)*** macroscopic dissociation-reassociation processes (fully random diffusional search), ***ii)*** microscopic dissociation-reassociation processes between closely spaced DNA sites, ***iii)*** intersegment transfer events by ring-closure, and ***iv)*** sliding along the DNA molecule (Berg et al., 1981).

The efficient transfer of genetic information depends on specific binding of regulatory proteins on certain DNA sequences. However, this targeted binding is complicated since there is a large number of other nonspecific binding sites (~10^6-10^9 per DNA) and low concentrations of corresponding regulatory protein molecules. The specific protein-DNA binding process requires the formation of precise geometrical fit between the protein and its consensus DNA motif that accompanied by the formation of specific hydrogen and electrostatic contacts at the protein-DNA binding interface (von Hippel, 2007).

Theoretical views of protein-DNA interactions indicate two main components of binding forces. The first results from electrostatic attractions between oppositely charged DNA and protein molecules which is generally sequence-independent such as interactions of positively charged protein side chains with negatively charged DNA phosphate groups (Berg et al., 1981; von Hippel and Berg, 1986). The second interaction is rooted from specific DNA sequence motifs which strengthen the attraction of transcription factors (Berg et al., 1981; von Hippel and Berg, 1986). The microscopic mechanism of these increased affinity are probably due to the combinations of van Der Waals forces, hydrogen bonds, covalent bond formations,

and steric interactions, as well as electrostatic charge pattern recognitions (Cherstvy, 2011; Cherstvy et al., 2008).

In addition to binding their consensus motifs, TFs can also bind with lower affinity to non-specific DNA regions lacking any consensus motif. Particularly, repeated homo-oligonucleotide tracts, namely poly(dA:dT) and poly(dC:dG), exhibits the strongest nonconsensus binding affinity toward DNA-binding proteins (Sela and Lukatsky, 2011). The origin of this phenomenon is purely entropic. In yeast, genome-wide landscape of nonspecific protein-DNA binding free energy significantly correlates with DNA binding preferences of ~200 TFs (Afek and Lukatsky, 2012). Also, DNA regions with enhanced nonconsensus TF-DNA binding are statistically significantly depleted of nucleosomes (Afek et al., 2011). It was suggested that the competition between TFs with histones for nonconsensus binding to DNA might affect nucleosome occupancy. Poly(dA:dT) and poly(dC:dG) tracts exhibits the strongest attraction for nonconsensus TF-DNA binding, thus allowing TFs to outcompete nucleosome on repeated homooligonucleotide tracts (Afek et al., 2011).

2.5 Nucleosome

In eukaryotes, genomic DNA is organized into chromatin whose basic repeating unit is a nucleosome that is composed of ~147 base pairs of DNA wrapped around a complex of histone octamer (Richmond and Davey, 2003). In vivo, chromatin further folded into compact fibers that reach to 30 to 400 nm thickness. The chromatin organization poses a barrier to transcription as it prevents the direct interaction between transcription machinery and promoter DNA. Based on their DNA binding modes, TFs differ with respect to how much their binding is affected by the nucleosome. A special group ‘pioneer TFs’ can bind to their corresponding DNA sequence in the context of nucleosomal DNA and recruit nucleosomal remodelers make DNA more competent and facilitate binding of other TFs (Zaret and Carroll, 2011). The TBP (TATA-binding protein), and thus Pol II basal transcription machinery requires nucleosome-free region to bind and initiate transcription

(Workman and Kingston, 1998). Chromatin remodeling enzymes alter the fluidity, folding and basic structure of chromatin. There are two-classes of chromatin remodeling enzymes: the first is nuclear histone acetyltransferases (HATs) that modify nucleosomal histone proteins by acetylation, methylation and phosphorylation; and the second is adenosine triphosphate (ATP)-dependent SWI/SNF like complexes that alter chromatin structure at the expense of ATP (Fry and Peterson, 2001).

2.5.1 Nucleosome Optimization

Chromatin is a key element in gene regulation and contributes to determination of promoter strength; and modifications of nucleosome structure leads to alterations in transcriptional capacity (Curran et al., 2014; Sharon et al., 2012). Nucleosomes are assembled into regularly spaced arrays with varying linker length. The genome-wide organization of nucleosome positioning is determined by combinations of DNA sequence, nucleosome remodeling enzymes and transcription factors including activators, components of PIC and elongating PolII; and these determinants influence each other, therefore nucleosome positioning patterns are probably differing among genes and among cells in a population with an eventual effect on gene expression (Struhl and Segal, 2013).

Nucleosome Positioning Prediction (NuPoP) software is a bioinformatics tool based on Hidden Markov Model that is used to *de novo* prediction of nucleosome occupancy along a DNA sequence (Xi et al., 2010)**.** Evaluation of TEF promoter library showed that cumulative sum of predicted nucleosome affinity across the entire promoter is inversely proportional to promoter strength in a predictable and robust way, in spite of the vast amount of differences in sequence and TFBS mutations (Curran et al., 2014). A greedy algorithm for nucleosome affinity minimization was employed NuPoP to redesign native *S. cerevisiae* promoters to minimize their cumulative affinity scores over several cycles of optimization which is restricted with sequence-based requirements to prevent creation or destruction of any well-defined TFBSs; and,

nucleosome optimized *S. cerevisiae* promoters exhibited up to 16-fold increase in their expression capacity (Curran et al., 2014). Nucleosome architecture plays a leading role in determining yeast promoter activity which makes it a design parameter in promoter engineering approaches.

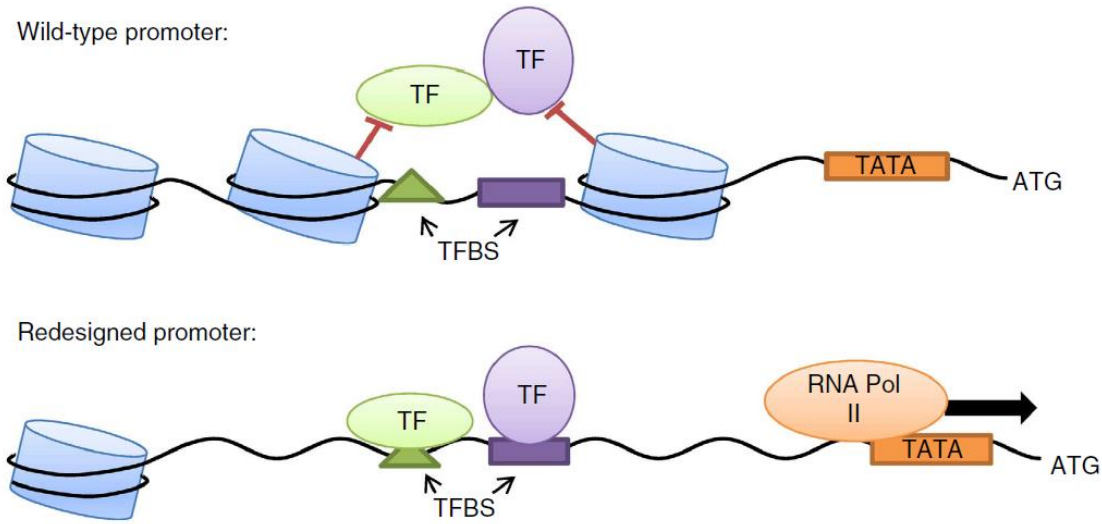


Figure 2.3 A model for promoter nucleosome architecture that depicts wild-type and nucleosome optimized redesigned promoter variant (Curran et al., 2014)

2.6 Yeast Promoter Architecture

A promoter is a DNA sequence that can independently facilitate the binding of transcription factors and enable accurate initiation of transcription by RNA polymerase. Interactions between transcription factors and promoter DNA recruit the transcriptional machinery necessary for activation of transcription (Ptashne and Gann, 1997). In general, yeast promoters contain two distinct elements: 1) a core promoter and 2) an upstream activation sequence (Figure 2.4). Core promoter is the minimal DNA sequence sufficient to direct accurate transcription initiation, in another words

transcriptional direction and TSS are determined by core promoter while upstream activation sequence contributes to promoter strength via determining transcriptional frequency (Smale and Kadonaga, 2003).

Figure 2.4 Yeast promoter architecture

In *S. cerevisiae*, approximately 20% of genes contain a TATA box and remarkably they are highly regulated, associated with stress responses and preferentially recruit SAGA other than TFIID (Basehoar et al., 2004). The distance between TATA-box and TSS was generally 40-120 bp in *S. cerevisiae* (Li et al., 1994).

Upstream activation sequence elements localize transcription factors to control transcriptional frequency and/or impart regulation to the core promoter. Throughout upstream activation sequence element, specific DNA sequences named as transcription factor binding sites (TFBSs) serve as docking platform for certain transcription activators or repressors. Promoter-bound transcription factors interacts with one another and with the basal transcriptional machinery to determine promoter regulation and strength. Upstream activation sites (UAS) that increase transcriptional rate and/or upstream repressive sites (URS) that reduce transcriptional frequency can be located on upstream activation sequence elements (Hahn and Young, 2011). Regulation pattern of promoter, induction or repression depend on varying environmental conditions, is also resulted from interactions of transcription factors located in the upstream elements (West et al., 1984).

2.7 Promoter Engineering

For industrial biotechnology applications, for the overexpression of recombinant proteins, strong and tightly regulated promoters are required to maximize protein production and eliminate toxicity during growth phase. In order to find a proper promoter for those applications, hundreds of them are being screened, however at the end only few of them satisfy the bioprocess requirements. However, in metabolic engineering studies, very high transcriptional capacity of promoter systems might create an excessive metabolic burden on the cell that can lead to decreased product formation and cell viability. In order to fulfill different requirements of diverse cell engineering targets, promoter libraries that span a wide dynamic range of strength should be established.

Promoter engineering approaches aim to modulate transcriptional capacity of promoters by mutating promoter DNA sequence to obtain a promoter library with a varying dynamic range. The basis of promoter engineering mostly depends on promoter architecture. The modular nature of eukaryotic promoters has led to fusion of different combinations of upstream activation sequence elements and core promoters which elicits hybrid promoters. At the most fundamental level, upstream activation sequence and core promoter elements possess TFBSs that govern overall promoter function. Randomized promoter mutagenesis introduces changes in nucleotide sequences of TFBSs, thus alters promoter strength. Since the mutations on TFBSs more likely reduce binding efficiency of specific TFs, mostly weaker promoter variants have been created than original promoters with random mutagenesis approach.

The high complexity of promoters and the cooperativity and competition between TFs for similar cis-regulatory sites and the interactions between TFs often creates diverse expression patterns that are difficult to predict from the effect of individual components of promoter architecture such as TFs (Istrail and Davidson, 2005; Kulkarni and Arnosti, 2005).

2.7.1 Random Mutagenesis

Random mutagenesis approach relies on large scale mutagenesis on promoter consensus motifs and spacer regions and selection to generate promoter library. This method ensures to yield novel promoter variants with sufficient library size; however, it's probable to obtain decreased promoter strength due to mutations occurring on TFBSs and reducing affinity between TF-promoter. *S. cerevisiae TEF1* (translation and elongation factor 1) promoter was subjected to error-prone PCR and a promoter library ranging between 8% and 120% of the activities of native TEF1 promoter was created (Alper et al., 2005; Nevoigt et al., 2006). Regulation of *S. cerevisiae* oxygen responsive *DAN1* promoter was changed to be induced under less-stringent anaerobiosis conditions by random mutagenesis, too (Nevoigt et al., 2007). Regulation of *P. pastoris* AOX1 promoter was modified by randomly mutated synthetic oligonucleotides, random mutagenesis of P_{AOX1} resulted in significantly abolished glucose repression and increased promoter activity among mutated P_{AOX1} promoter library (Berg et al., 2013). Another random mutagenesis approach in *P. pastoris* was applied to glyceraldehyde-3-phosphate dehydrogenase promoter (P_{GAP}) and P_{GAP} library was generated that activity range was spanning from 8% to 218% activity of wild type P_{GAP} promoter (Qin et al., 2011).

Saturation mutagenesis provides more directed promoter engineering that mutates only variable spacer regions while retaining consensus regions intact, therefore, it seems somewhat rational method to modify promoter strength. As yeast saturation mutagenesis studies widely used in synthetic promoter design, its applications mentioned under synthetic promoter section.

2.7.2 Hybrid Promoter Engineering

A hybrid promoter engineering method requires the assembly of yeast upstream activation sequence-core promoter fusions to improve basal core transcriptional capacity or provide novel promoter regulation. Upstream activation sequence

elements encompass TFBSs that maintain native regulation and expression activation or repression independent from core promoter element. Minimal specific regions of promoter are identified as upstream activation site (UAS) that is essential for transcriptional activation or induction control. In chimeric hybrid promoter design, tandem repeats of UAS elements can radically increase the expression capacity of core promoter, thus hybrid promoter engineering enables generation of large coverage promoter libraries and improvements in transcriptional capacity even of the strongest native core promoters.

The oleaginous yeast, *Yarrowia lipolytica*, has been widely used for hybrid promoter engineering applications owing to its 105-bp upstream activation element, UAS1B that was identified on strong and tightly regulated alkaline extracellular protease (XPR2) promoter (Blanchin-Roland et al., 1994). Fusion of one to four tandem UAS1B element to P_{LEU2} core promoter created hybrid promoters with increasing strength, called hp1d through hp4d (Madzak et al., 2000); and, the hybrid promoter P_{hp4d} has been employed in more than half of the studies for r-protein production by *Y. lipolytica* (Madzak, 2015). Hybrid promoter library construction has expanded from 1 to 32 UAS1B tandem copy addition to P_{LEU2} minimal core promoter which created the strongest characterized promoters in *Y. lipolytica* that spanned a 400-fold activity range (Blazeck et al., 2011). Also, fusion of the same 8 to 16 UAS1B tandem repeats with P_{TEF} core promoter demonstrated that UAS element and core promoter act as independent modular parts in the context of hybrid promoter engineering (Blazeck et al., 2011).

In *S. cerevisiae* fusion of discrete UAS elements and core promoters formed a constitutive promoter library that span a 90-fold dynamic range (Blazeck et al., 2012). This approach also increased the strength of strongest constitutive yeast promoter P_{GPD} by 2.5-fold and P_{GAL1} by 1.15-fold (Blazeck et al., 2012). To our knowledge, there is no hybrid promoter engineering study in *P. pastoris*, yet.

2.7.3 Synthetic Promoter Designs

Yeast promoter engineering approaches has mainly focused on native scaffolds; however synthetic promoter designs show a promise to move away from native parts but still widely relies on endogenous elements. Consensus yeast promoter structure still is not available strictly. In *S. cerevisiae* different motifs including two Gcr1 TFBSs, two Rap1 TFBSs and a TATA-box pieced together that were separated by degenerate nucleotide spacer regions; variations in spacer regions resulted in promoter library covering a activity range of three order of magnitude (Jeppsson et al., 2003).

Synthetic yeast core promoters for *P. pastoris* were designed by using consensus sequences of natural core promoters and incorporation of common TFBSs. To test their potential, designed synthetic core promoters were fused with upstream sequence of P_{AOX1}; upon methanol induction expression strength of synthetic P_{AOX1} library ranged from 10% to 117% of wild-type P_{AOX1} (Vogl et al., 2014).

Recently, a methodical workflow to identify robust minimal core elements with exactly modular and context-independent function was established. Synthetic promoters that can maintain high expression to the level of strong native promoters with relatively short DNA size (with up to an 80% reduction in regulatory DNA region) were created (Redden and Alper, 2015).

2.7.4 Modification of TFBSs

Eventually promoter strength and regulation are the cumulative effect of short, distinct DNA motifs that facilitate binding of cellular transcriptional machinery. Hybrid promoter engineering approaches depend on this idea by fusing tandem UAS elements which possess TFBSs to localize specific TFs and accordingly improve or regulate transcriptional capacity. Random mutagenesis implements haphazard mutations around or within TFBSs while saturation mutagenesis only enables modifications around TF binding motifs. Synthetic promoter engineering employs

TFBSs as a modular genetic element to design novel promoter architectures. In the end, all promoter engineering methods depend on addition, abrogation or modification of TFBSs, and their genetic context to modulate promoter transcriptional capacity. Rational construction of promoter libraries with distinct regulatory characteristics will be possible with the progressive increase in the knowledge on TFBSs and their corresponding TFs that regulate transcription directly.

Modification of TFBSs was implemented to *P. pastoris* P_{AOX1} via addition or deletion of putative TFBSs; and transcriptional capacity of constructed P_{AOX1} promoter library spans 6% to 160% activity of native promoter (Hartner et al., 2008). Recently, *P. pastoris* P_{GAP} was modified by means of its putative TFBSs, in this context putative activator binding sites were duplicated while repressor motifs were deleted and the strength of the P_{GAP} library was in the range of 82% to 190% of wild-type P_{GAP} (Ata et al., 2017).

Alternative promoter engineering approaches rely on the systematic deletions of larger gene stretches of promoters to identify regulatory regions. Internal deletions in P_{AOX1} revealed a cis-regulatory sequence named as Region D that is located -638 to -510 bp positions; and, when tandem copies of Region D was inserted into P_{AOX1}, expression capacity of designed promoter was increased to 157% of the native P_{AOX1} (Xuan et al., 2009).

2.8 Transcriptional Engineering

Transcriptional engineering is a tool which allows modification of the regulation of an intracellular reaction pathway by engineering the multiple components such as promoters or regulators including transcription factors (Ata et al., 2017). *Cis*-acting promoter elements such as upstream activation sites and upstream repression sites provide docking platform for the binding of *trans*-acting activators and suppressor. These discrete elements of transcription can form a complete transcriptional unit using rational combinatorial engineering strategies.

Transcription factors have a crucial role in activation of transcription and determining the specificity of transcription; thus, TFs can also be an engineering target besides promoter engineering approaches. The above mentioned P_{GAP} library was further improved via transcriptional engineering by overexpression or deletion of interested TF genes; in that *P. pastoris* cells $P_{GAP\text{-}V}$ expression levels varied in the range of 35% to 310% of wild-type P_{GAP} driven expression in native *P. pastoris* (Ata et al., 2017). A hybrid promoter based on the *ARO9* upstream region that shows high expression in response to exogenous tryptophan and a mutant Aro80p transcription factor capable of high constitutive expression while retaining inducible characteristic were used, by this combinatorial transcriptional engineering approach very strong expression levels around 2-fold higher than P_{TDH3} was achieved (Leavitt et al., 2016). Another *S. cerevisiae* hybrid promoter library was designed by replacing operator region with a sequence recognized by bacterial LexA DNA binding protein and that library was activated with a synthetic TF including bacterial LexA DNA binding domain fused with the human estrogen binding domain and the viral activator domain, Vp16 (Dossani et al., 2018). This synthetic transcriptional unit can be activated by estradiol which is not have any detectable impact on *S. cerevisiae* physiology and also bacterial DNA binding domain provides orthogonal activation that avoids the transcription of native *S. cerevisiae* genes (Dossani et al., 2018). Some methods for engineering synthetic transcriptional units based on synthetic biology principles were recently reviewed by Mehrotra et al (Mehrotra et al., 2017).

CHAPTER 3

MATERIALS AND METHODS

3.1 Sequence Analysis of *P. pastoris* ADH2 and AOX1 Promoters

The P_{ADH2} sequence of *P. pastoris* (1048 bp) (Appendix A) was retrieved from the U.S. Patent No. 8,222,386 B2 (Cregg and Tolstorukov, 2012) and putative TFBSs were analyzed using MatInspector release 8.4.1 (Cartharius et al., 2005) within the Genomatix Suite via Matrix Family Library for fungi and core elements (www.genomatix.de). *Saccharomyces cerevisiae* TFBS database was used to find putative TFBSs. Matrix similarity scores higher than 0.75 were considered as potentially true matches. Additional TFBSs were manually curated using YEASTRACT (http://www.yeastract.com/) (Teixeira et al., 2018) and Yeast Promoter Atlas (http://ypa.csbb.ntu.edu.tw/) (Chang et al., 2011) databases. In addition to these analysis, TFBS optimizations that were already described in *Saccharomyces cerevisiae* literature (Cheng et al., 1994; Roth et al., 2004) were tested within P_{ADH2} context.

The P_{AOX1} sequence of *P. pastoris* (940 bp) (Appendix B) was obtained from pPICZαA vector map (downloaded from https://www.thermofisher.com/order/catalog/product/V19520) and putative TFBSs were searched with MatInspector release 8.2 (Cartharius et al., 2005) within the Genomatix Suite via Matrix Family Library for fungi and core elements, *S. cerevisiae*

TFBS database was used to find putative TFBSs. Matrix similarity scores higher than 0.75 were considered as potentially true matches.

3.2 Nucleosome Optimization of ADH2 Promoter

Computational redesign of nucleosome optimized P_{ADH2} was performed using NuPOP, a software tool for Nucleosome Positioning Prediction (Wang et al., 2008; Xi et al., 2010), with minor modifications as explained by Curran et al. (2014). Nucleotides outside of the identified TFBSs were systematically perturbed via custom MATLAB scripts, that employs NuPOP Fortran implementation to predict cumulative sum of predicted nucleosome affinity score across the entire P_{ADH2} gene through optimization process (Curran et al., 2014). A greedy algortihm was utilized to minimize cumulative affinity score over several cycles of optimization; in each cycle, potential candidates with a single nucleotide difference were computationally created and the candidate with the smallest cumulative affinity score was saved as the product of that optimization cycle. Each successive iteration, generated lower nucleosome occupancy across promoter sequence. The optimization method was also restricted from creation or destruction of known transcription factor binding sites that was obtained from YEASTRACT database (Teixeira et al., 2018).

MATLAB with Bioinformatics toolbox was installed and Fortran code for NuPOP (obtained from http://nucleosome.stats.northwestern.edu/) was edited to enable the acceptance of comand-line inputs of MATLAB. Modified Fortran code was compiled to create Npred.exe that predicts nucleosome affinity for a given DNA sequence. The directory of Npred.exe was added to computer's system's path to make it available for MATLAB. Input data for P_{ADH2} was designed starting from 200 bp upstream of the promoter untill 100 bp downstream (enhanced green fluorescent protein (eGFP)'s first 100 nucleotides was given as downstream sequence) of the sequence for more reliable prediction of nucleosome affinity.

MATLAB Input Data ADHG: 200 bp upstream of promoter, P_{ADH2} sequence, 100 bp downstream sequence (the first 100 nucleotides of eGFP). Promoter start and end positions are underlined.

‘’AGGTACCTGGAATTCTCAGAGTTTTTCCAGTACATCGCAGCGTTTTCTG
ACGGTACTAGAGGACTCTTAGGGGAAGGTAGAATCAATAAAGATCATAT
TAGGTAAGCAAATTTTGGATGGAATAGGAGACTAGGTGTGGATGCGCGA
TCTCGCCAAATTGCACGACCAGAGTGGATGCCGGATGGTGGTAAACCGT
TTCTTCCTTTTTACCACCCAAGTGCGAGTGAAACACCCCATGGCTGCTCT
CCGATTGCCCCTCTACAGGCATAAGGGTGTGACTTTGTGGGCTTGAATTT
TACACCCCCTCCAACTTTTCTCGCATCAATTGATCCTGTTACCAATATTG
CATGCCCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATC
GCAGGGTGAGACTGTAGAGACCCCACATAGTGACAATGATTATGTAAGA
AGAGGGGGGTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTG
AACAATGCCCAGCAGTCCTCCCCACTCTTTGACAAATCAGTATCACCGAT
TAACACCCCAAATCTTATTCTCAACGGTCCCTCATCCTTGCACCCCTCTTT
GGACAAATGGCAGTTAGCATTGGTGCACTGACTGACTGCCCAACCTTAA
ACCCAAATTTCTTAGAAGGGGCCCATCTAGTTAGCGAGGGGTGAAAAAT
TCCTCCATCGGAGATGTATTGACCGTAAGTTGCTGCTTAAAAAAAATCA
GTTCAGATAGCGAGACTTTTTTGATTTCGCAACGGGAGTGCCTGTTCCAT
TCGATTGCAATTCTCACCCCTTCTGCCCAGTCCTGCCAATTGCCCATGAA
TCTGCTAATTTCGTTGATTCCCACCCCCCTTTCCAACTCCACAAATTGTCC
AATCTCGTTTTCCATTTGGGAGAATCTGCATGTCGACTACATAAAGCGAC
CGGTGTCCGAAAAGATCTGTGTAGTTTTCAACATTTTGTGCTCCCCCCGC
TGTTTGAAAACGGGGGTGAGCGCTCTCCGGGGTGCGAATTCGTGCCCAA
TTCCTTTCACCCTGCCTATTGTAGACGTCAACCCGCATCTGGTGCGAATA
TAGCGCACCCCCAATGATCACACCAACAATTGGTCCACCCCTCCCCAAT
CTCTAATATTCACAATTCACCTCACTATAAATACCCCTGTCCTGCTCCCA
AATTCTTTTTTCCTTCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTA
CGAAAATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCA
TCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTC
CGGCGAGGG’’

Yeast TF consensus list from YEASTRACT database (Teixeira et al., 2018) was also provided as an input file to avoid creation or destruction of known TFBSs. Throughout optimization process, yeast stress responsive elements, carbon source responsive elements, TFBSs related with non-optimal carbon source utilization, core promoter region and originally nucleosome depleted regions were protected from any mutation via defining them as forbidden sites to the optimization algorithm.

ADHGforbidden = [250:264 298:312 331:351 351:355 363:377 377:393 387:401 414:422 441:455 448:451 453:467 459:469 461:465 479:493 484:490 519:539 547:555 582:596 591:605 654:668 674:688 679:685 804:818 830:844 863:877 878:882 911:915 939:952 987:990 1004:1007 1018:1025 1043:1057 1055:1075 1076:1247]

$P_{ADH2\text{-}NucOpt}$ nucleotide sequence

TCCTTTTTACCACCTAAGTGCGAGTGAAACACCCTATGGCTGCTCTCCGA
TTGCCCCTCTACAGGCATAAGGGTGTGATTTTTTTTTTTTTAATTTTACAC
CCCCTCCAACTTTTTTCGCGTAAATTGATCCTGTTACCAATATTGCATGC
CCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATCGCAGG
GTAAAAATATATAGACCCCACAAAAAAAAAATGATTATGTAAGAAGAG
GGGGGTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTGAAAA
ATGCCCAGCTTTTTTCCCTATTCTTTGACAAATCAGTATCACTTATTAACA
CCCCAAATTTTTTTCTCAACGGTCCCTCATCCTTGCACCCCTCTTTGGACA
AATGGCAGTTAGTATTAGTGCACTGACTGACTGCCTAACCTTAAACCCTA
ATTTCTTAGAAGGGGCCCATATAGTTAGCGAGGGGTGAAAAATTCCTCC
ATCGGAGATGTATTAACCGTAATTTTTTTTTTAAAAAAAAAAAATTCAGA
TAGCGAAATTTTTTTGATTTCGCGACGCGCGTTTTTTTTTTTTTTTTTTTT
TTTCTCACCCCTTCTGCCCAGTTCTGCCAATTGCCCATGAATCTACTAATT
TCGTTGATTCCCACCCCCCTTTCCAACTCCAAAAATTTTTTAATTTTTTTT
TTTTTTTGGGAGAATCTGAATGTATATTACATAAAGCGACCGGTGTCCGA
AAAAATTTTTTTTTTTTTTAATTTTTTTTTTTTCCCCCGCTTTTTAAAAACG
GGGGTAAGCGCTCTCCGGGGTGCGAATTCGCGCCCTATTCCTTTCACCCT
GCCTATTGTAGACGTCAACCCGCATCTGGTGCGAATATAGCGCACCCCC
AATGATCACACCAACAATTGGTCCACCCCTCCCCAATCTCTAATATTCAC
AATTCACCTCACTATAAATACCCCTGTCCTGCTCCCAAATTCTTTTTTCCT
TCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTACGAAA

* Protected nucleotide stretches from any mutation are highligted.

Using edited MATLAB commands (Curran et al., 2014) P_{ADH2} was optimized in 1bp steps. A list of nucleosome-optimized promoters starting from the wild-type (or seed) sequence and proceeding in 1bp steps towards a variant with reduced predicted nucleosome affinity; and the corresponding cumulative affinity score and the nucleosome affinity curve of each cycle's optimized promoter variant were created

throughout optimization. Redesigned P_{ADH2} variant via 100 cycle of optimization was defined as nucleosome optimized P_{ADH2}, $P_{ADH2\text{-}NucOpt}$.

3.3 Strains, Plasmids, Primers, and Maintenance

E. coli DH5α (Invitrogen, USA), *E. coli* DH10B (Invitrogen, USA), *P. pastoris* X-33 (for eGFP expression under P_{ADH2} or P_{AOX1} variants, for overexpression or deletion of Adr1, Aca1, Cat8-1 and Cat8-2 transcription factors, and for construction of human serum albumin (HSA) producing strains) and *P. pastoris* CBS7435 (for construction of Flag-Tagged Adr1 TF) were used for transformation experiments. For propogation of the shuttle vectors that carry P_{ADH2} and P_{AOX1} variants with either eGFP or HSA, *E. coli* DH5α cells were used. For propagation of the shuttle vectors that carry TF overexpression and deletion casettes and gRNA and Cas9 genes, *E. coli* DH10B cells were employed. For long term storage, modified microorganisms were kept in 10% glycerol at 80°C. Stock (100 μM) and working (10μM) solutions of primers, plasmid DNA, PCR amplified or restriction enzyme digested gene fragments were stored at -20°C. Designed primers' list for contruction of recombinant plasmids is presented in Table 3.1 Strains and plasmids constructed in this study are presented in Table 3.2

Table 3.1 Primers designed and used in this study. Restriction enzyme recognition sites used for cloning are shown with *Italic* characters.

Primer	Sequence	Used for
Forward P_{ADH2}	GTC*GGATCCCTGCAG*TCCTTTTTAC	Amplification of wt P_{ADH2} and its variants
Reverse P_{ADH2}	CTG*AGCACTGC*ACGCCGTAGGT	Amplification of wt P_{ADH2} and its variants
Forward P_{AOX1}	CTC*AGATCT*AACATCCAAAGACGAAAGG	Amplification of wt P_{AOX1} and its variants
Reverse P_{AOX1}	CTG*AGCACTGC*ACGCCGTAGGT	Amplification of wt P_{AOX1} and its variants
P_{ADH2}-optHSA Forward	TTAGGATCCGGCGCGCCTTCCTTTTACCACCCAAG	Amplification of P_{ADH2} to clone with optHSA
P_{ADH2}_GGA Reverse	GTCATT*GAAGAC*TGCATTTTCGTAAAGTAAATAAGATAAAAGCTAGTAG	Amplification of P_{ADH2} to clone with optHSA
P_{AOX1}-optHSA Forward	AGATCTAACATCCAAAGACGAAAG	Amplification of P_{AOX1} to clone with optHSA
P_{AOX1}_GGA Reverse	TGAATGGGTCTCACATTCGTTTCGAATAATTAGTTGTTTTTTG	Amplification of P_{AOX1} to clone with optHSA
OptHSA_GGA Forward	GATCTA*GGTCTC*AAATGAAGTGGGTTACTTTCATCT	Amplification of optHSA
OptHSA reverse	CA*GGTACC*CTATTACAGACCCAAAGCAGC	Amplification of optHSA
Primers used to construct P_{ADH2} variants		
Forward_TATA	TGCCTATTGTAGACGTCAACCCGCAACAGCAATATATAAACAGAACCCCCAATGATCACACCAACAATT	Integration of TATA binding sequence
Reverse_TATA	AATTGTTGGTGTGATCATTGGGGGTTCTGTTTATATATTGCTGTTGCGGGTTGACGTCTACAATAGGCA	Integration of TATA binding sequence
Forward_opt_Adr1_1	ACTGTAGAGACCCCAAATAGTGACAATGATTATGTAAGAATTGGGGAGTGATTCGGCCGGCT	Optimization of 1st Adr1 TFBS
Reverse_opt_Adr1_1	AGCCGGCCGAATCACTCCCCAATTCTTACATAATCATTGTCACTATTTGGGGTCTCTACAGT	Optimization of 1st Adr1 TFBS
Forward_opt_Adr1_2	CCACCCCCCTTTCCATCTCCAAAAATTGTCCAATCTCGTTTTCCATTTTGGAGAATCTGCATGT	Optimization of 2nd Adr1 TFBS

Table 3.1 (continued)

Primer	Sequence	Used for
Reverse_opt_Adr1_2	ACATGCAGATTCTCCAAAATGGAAAACGAGATTGGACAATTTTTGGAGATGGAAAGGGGGGTGG	Optimization of 2nd Adr1 TFBS
Forward_opt_Adr1_3	GTGTAGTTTTCAACATTTTGTGCTCTCCCCAATGTTTGAAATTGGGGATGAGCGCTCTCCGGGGTGCGAATTC	Optimization of 3rd Adr1 TFBS
Reverse_opt_Adr1_3	GAATTCGCACCCCGGAGAGCGCTCATCCCCAATTTCAAACATTGGGGAGAGCACAAAATGTTGAAAACTACAC	Optimization of 3rd Adr1 TFBS
Forward_add_Adr1	TTGATTTCGCAACGGGAGTGCCTGTCCCCAACGATTGCATTGGGGACCCCTTCTGCCCAGTCCTGCCAAT	Integration of Adr1 binding sequence
Reverse_add_Adr1	ATTGGCAGGACTGGGCAGAAGGGGTCCCCAATGCAATCGTTGGGGACAGGCACTCCCGTTGCGAAATCAA	Integration of Adr1 binding sequence
Forward_opt_Cat8	GTCGACTACATAAAGTTCCGTTCGTCCGAAAAGATCTG	Optimization of Cat8 TFBS
Reverse_opt_Cat8	CAGATCTTTTCGGACGAACGGAACTTTATGTAGTCGAC	Optimization of Cat8 TFBS
Forward_add_Cat8	CCTGCCAATTGCCCATGAATCTGCTTCCGTTCGTCCGACCCACCCCCCTTTCCAACTCCACAA	Integration of Cat8 binding sequence
Reverse_add_Cat8	TTGTGGAGTTGGAAAGGGGGGTGGGTCGGACGAACGGAAGCAGATTCATGGGCAATTGGCAGG	Integration of Cat8 binding sequence
Forward_add_Aca2	TACATAAAGCGACCGGTGTCCGAAAGCCTATTGTAGACGTCAACCCTTTTGTGCTCCCCCCGCTGTTTGAA	Integration of Aca2 TFBS
Reverse_add_Aca2	TTCAAACAGCGGGGGGAGCACAAAAGGGTTGACGTCTACAATAGGCTTTCGGACACCGGTCGCTTTATGTA	Integration of Aca2 TFBS
Forward-Nucopt-OptCat8	GTATATTACATAAAGTTCCGTTCGTCCGAAAAAATTTT	Combinatorial modification of NucOpt and OptCat8 TFBS
Reverse-NucOpt-OptCat8	AAAATTTTTTCGGACGAACGGAACTTTATGTAATATAC	Combinatorial modification of NucOpt and OptCat8 TFBS

Table 3.1 (continued)

Primer	Sequence	Used for
Primers used to construct P_{AOX1} variants		
Forward mAOX-AddCat8-1	ATTCCGTTCGTCCGATTAGCAGACCGTTGCAAACG	Integration of Cat8_1st TFBS
Reverse mAOX AddCat8-1	AATCGGACGAACGGAATTGTTGCGTTTGGCACTTATG	Integration of Cat8_1st TFBS
Forward mAOX-AddAdr-1	ACCCCAATATTATTTGGGGTACTTTTGCCATCGAAAAAC	Integration of Adr1_1st TFBS
Reverse mAOX-AddAdr1-1	ACCCCAAATAATATTGGGGTGTGGAGGTCCTGCGTTTG	Integration of Adr1_1st TFBS
Forward mAOX-AddCat8-2	CCTCTCGTCCGGGCTTTTTCCGAACATCACTCCAG	Integration of Cat8_2nd TFBS
Reverse mAOX-AddCat2	GAAAAAGCCCGGACGAGAGGGCATTCGGAAATAAACAAAC	Integration of Cat8_2nd TFBS
Forward mAOX-AddAdr1-2	GACCCCACATTTTTTTTTTGACCCCACATGTTCCCCAAATGGCC	Integration of Adr1_2nd TFBS
Reverse mAOX-AddAdr1-2	TGGGGTCAAAAAAAAAAATGTGGGGTCGCCCTCATCTGGAGTGATG	Integration of Adr1_2nd TFBS
Forward mAOX-AddAca2	GCCTATTGTAGACGTCAACCCAAGTCGGCATACCGTTTGTC	Integration of Aca2 TFBS
Reverse mAOX-AddAca2	GGGTTGACGTCTACAATAGGCACTGGCCGTTAGCATTTC	Integration of Aca2 TFBS
Forward mAOX-AddCat8-3	CATATTCCGTTCGTCCGAATCTTTTTGGATGATTATGC	Integration of Cat8_3rd TFBS
Reverse mAOX-AddCat8-3	ATTCGGACGAACGGAATATGTTTCGGCACAGGTGCACCG	Integration of Cat8_3rd TFBS
Forward mAOX-AddAdr1-3	ACCCCAATACATTTTGGGGTTGCTTCCAAGATTCTGGTGG	Integration of Adr1_3rd TFBS
Reverse mAOX-AddAdr1-3	ACCCCAAAATGTATTGGGGTTAATCATCCAAAAAGCGGGTG	Integration of Adr1_3rd TFBS

Table 3.1 (continued)

Primer	Sequence	Used for
Primers used to construct Transcription factor overexpression and knock-out casettes		
Adr1_Forward	GATCTA*GGTCTC*ACATGAGCAATCT ACCCCC	Amplification of *Adr1* gene
Adr1-1_Reverse	GTCATT*GGTCTC*AGAGGACCGACG AGTT	Elimination of internal *Bsa*I or *Bpi*I site
Adr1_1-2_Forward	GATCTA*GGTCTC*ACCTCTAATGCCG CG	Elimination of internal *Bsa*I or *Bpi*I site
Adr1_1-2_Reverse	GTCATT*GGTCTC*AGGACTCAAGTTC TTGAATGG	Elimination of internal *Bsa*I or *Bpi*I site
Adr1_1-3_Forward	GATCTA*GGTCTC*AGTCCTCTTTGAA TGCATCC	Elimination of internal *Bsa*I or *Bpi*I site
Adr1_1-3_Reverse	GTCATT*GGTCTC*AGCAAACAAACA ATTGATATTCGTT	Elimination of internal *Bsa*I or *Bpi*I site
Adr1_4_Forward	GATCTA*GGTCTC*TTTGCCTCTTTTG GTGGCTACG	Elimination of internal *Bsa*I or *Bpi*I site
Adr1_4_Reverse	GTCATT*GGTCTC*GAAGCCTACTAGA CACCACCATCTAGTCG	Amplification of *Adr1*
ribozyme_gRNA_rib1_Adr1KO_Forward	GATA*GGTCTC*CCATGTGTTCTCTGA TGAGTCCGTGAGGACGAAACGAGT AAGCTCGTC	gRNA design for *Adr1*-KO
ribozyme_gRNA_rib2_Adr1KO_Forward	AAACGAGTAAGCTCGTCAGAACAA TTAATAGGAGTGGGTTTTAGAGCT AGAAATAGCAAG	gRNA design for *Adr1*-KO
Adr1_KO_HA1_forward	GATCTA*GAAGAC*GCCATGAAGGTG AAAATTCAG	Amplification of homologous region 1 for *Adr1*-KO
Adr1_KO_HA1_reverse	GATCTA*GAAGAC*TCAACCTTCGGTC CGAGTACT	Amplification of homologous region 1 for *Adr1*-KO
Adr1_KO_HA2_forward	GATCTA*GAAGAC*ACGGTTAATGAA TTATGATTTTGTTTGACTATAGATT AG	Amplification of homologous region 2 for *Adr1*-KO
Adr1_KO_HA2_reverse	CATGTA*GAAGAC*GCAAGCCCTTAC GTCTCATATCACAAGA	Amplification of homologous region 2 for *Adr1*-KO

Table 3.1 (continued)

Primer	Sequence	Used for
Adr1_KO_HA_final_forward	CATGAAGGTGAAAATTCAGAC	Amplification of homologous region for *Adr1*-KO
Adr1_KO_HA_final_reverse	CCTTACGTCTCATATCACAAG	Amplification of homologous region for *Adr1*-KO
ACA1 Forward	GATCTA*GGTCTC*ACATGGCTCATAATAACCCAC	Amplification of *Aca1*
ACA1 Reverse	GTCATT*GGTCTC*GAAGCCTATTATTCTGAATCTTCATCTTGAGTTAG	Amplification of *Aca1*
ACA1 SOE Forward	GAGCTTGTGAAAACGTCTGAAC	Elimination of internal *Bsa*I or *Bpi*I site
ACA1 SOE Reverse	GTTCAGACGTTTTCACAAGCTC	Elimination of internal *Bsa*I or *Bpi*I site
Aca1_KO_BGE_rib_2_Aca1_KO_Forward	AAACGAGTAAGCTCGTCTTTCCGAGAGATTTAGGATAGTTTTAGAGCTAGAAATAGCAAG	gRNA design for *Aca1*-KO
Aca1_KO_BGE_rib_1_Aca1_KO_Forward	GATA*GGTCTC*CCATGCGGAAACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA design for *Aca1*-KO
ACA1_KO_gRNA Rib 1 Forward 2nd	GATA*GGTCTC*CCATGAACTTACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	2nd gRNA design for *Aca1*-KO
ACA1_KO_gRNA Rib 2 Forward 2nd	AAACGAGTAAGCTCGTCTAAGTTGAAATAGGTCGCCAGTTTTAGAGCTAGAAATAGCAAG	2nd gRNA design for *Aca1*-KO
ACA1_KO_HA1 Forward	AATAAT*GGTCTC*ACATGAGATACCTATAACGATATGGATACC	Amplification of homologous region 1 for *Aca1*-KO
ACA1_KO_HA1 Reverse	GATCAT*GGTCTC*ATAGGGTGAATCTATTTATTTTCAATTCCAAT	Amplification of homologous region 1 for *Aca1*-KO
ACA1_KO_HA2 Forward 1	GAGTTT*GGTCTC*ACCTAAGAAAACCTTTTTGACGGATAAC	Amplification of homologous region 2 for *Aca1*-KO
ACA1_KO_HA2 Reverse 1	ACATCT*GGTCTC*CTCCCTTATAACATGCTCTACCTCATTGG	Elimination of internal *Bsa*I or *Bpi*I site
ACA1_KO_HA2 Forward 2	AGCACA*GGTCTC*AGGGACCAGAGTGCCATCC	Elimination of internal *Bsa*I or *Bpi*I site

Table 3.1 (continued)

Primer	Sequence	Used for
ACA1_KO_HA2 Reverse 2	ATTATT*GGTCTC*AAAGCGTTCCCCCCGTTAAACC	Amplification of homologous region 2 for *Aca1*-KO
ACA1_KO_HA Final Forward	ATGAGATACCTATAACGATATGGATAC	Amplification of homologous region for *Aca1*-KO
ACA1_KO_HA Final Reverse	GTTCCCCCCGTTAAACC	Amplification of homologous region for Aca1-KO
Cat8-1 HA1 Forward	TGCTCGTCGTGTCCTAAC	Amplification of homologous region for *Cat8-1* KO
Cat8-1 HA2 Reverse	AGAAAGATGCTGAGTGACAATCA	Amplification of homologous region for *Cat8-1* KO
Cat8-2 HA1 Forward	GGGCAGAAAGAAAGTACTAATATGGA	Amplification of homologous region for *Cat8-2* KO
Cat8-2 HA2 Reverse	AGGATGGTCCAAGCGTC	Amplification of homologous region for *Cat8-2* KO
Primers used to add FLAG-TAG to Adr1 transcription factor		
ribozyme_gRNA_rib2_Adr1_FLAG-Tag	AAACGAGTAAGCTCGTCCTAGAAAACCGACTAGATGGGTTTTAGAGCTAGAAATAGCAAG	gRNA design for Adr1-FLAG-Tag
ribozyme_gRNA_rib1_Adr1_FLAG-Tag	GATA*GGTCTC*CCATGTTCTAGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA design for Adr1-FLAG-Tag
Adr1_HA1_Forward	GATTTCGAATTCTGATTTGTTTCC	Amplification of homologous region 1 for Adr1- FLAG-Tag
Adr1_HA1_Reverse	GTCATT*GAAGAC*AACCGACACCTCCGTCTAGTCG	Amplification of homologous region 1 for Adr1- FLAG-Tag
Adr1_FLAG_Forward	GTCATT*GAAGAC*AATCGGGGGAGGTGGATC	Amplification of FLAG-Tag
Adr1_FLAG_Reverse	GTCATT*GAAGAC*GAAATTACTTATCGTCGTCATCCTTG	Amplification of FLAG-Tag
Adr1_HA2_Forward	GTCATT*GAAGAC*GCAATTAATGAATTATGATTTTGTTTGACTATAGATTAG	Amplification of homologous region 2 for Adr1- FLAG-Tag
Adr1_HA2_Reverse	CAATGAATTCTACATCGCAGG	Amplification of homologous region 2 for Adr1- FLAG-Tag

Table 3.2 Strains and plasmids constructed in this study.

Microorganism	Plasmid	Comment	Source
E. coli DH5α	pUC57::wtADH2	Plasmid carrying synthesized $P_{ADH2\text{-}wt}$ with the first 213 nucleotides of eGFP	Plasmid was synthesized and stored in *E. coli* DH5α.
E. coli DH5α	pUC57::NucOptADH2	Plasmid carrying synthesized $P_{ADH2\text{-}NucOpt}$ with the first 213 nucleotides of eGFP	Plasmid was synthesized and stored in *E. coli* DH5α.
E. coli DH5α	pUC57::modAOX	Plasmid carrying synthesized $P_{AOX1\text{-}mod}$ with the first 213 nucleotides of eGFP	Plasmid was synthesized and stored in *E. coli* DH5α.
E. coli DH5α	pPICZα-A::eGFP	Backbone plasmid	Ata et al., (2017)
E. coli DH5α	BB1aK_xx_pTHI11	GGA Level 1 plasmid carrying pTHI11	Prielhofer et al., (2017)
E. coli DH5α	BB1aK_34_RPS3TT	Level 1 plasmid carrying RPS3TT terminator	Prielhofer et al., (2017)
E. coli DH5α	BB3aK_14	Base plasmid for construction of overexpression plasmids	Prielhofer et al., (2017)
E. coli DH5α	Cas9-25	Backbone CRISPR/Cas9 Plasmid	Prielhofer et al., (2017)
E. coli DH10B	BB3aK_14_pTHI11_Cat8-1_RPS3TT	GGA Level 2 plasmid, used for overexpression of Adr1 TF in *P. pastoris*	Barbay et al. (Unpublished study)
E. coli DH10B	BB3aK_14_pTHI11_Cat8-2_RPS3TT	GGA Level 2 plasmid, used for overexpression of Aca1 TF in *P. pastoris*	Barbay et al. (Unpublished study)
E. coli DH10B	HR-Cat8-1_BB1	Used for amplification of homologous region to knock-out Cat8-1	Barbay et al. (Unpublished study)
E. coli DH10B	HR-Cat8-2_BB1	Used for amplification of homologous region to knock-out Cat8-2	Barbay et al. (Unpublished study)
E. coli DH10B	Cas9_gRNA_Cat8-1_BB3	CRISPR/Cas9 plasmid to knock-out Cat8-1	Barbay et al. (Unpublished study)

Table 3.2 (continued)

Microorganism	Plasmid	Comment	Source
E. coli DH10B	Cas9_gRNA_Cat8-1_BB3	CRISPR/Cas9 plasmid to knock-out Cat8-2	Barbay et al. (Unpublished study)
E. coli DH5α	pADH2-wt::eGFP	Plasmid carrying native P_{ADH2} with *eGFP*	Constructed in this study
E. coli DH5α	pADH2-NucOpt::eGFP	Plasmid carrying $P_{ADH2\text{-}NucOpt}$ with *eGFP*	Constructed in this study
E. coli DH5α	pADH2-AddCat::eGFP	Plasmid carrying $P_{ADH2\text{-}AddCat}$ with *eGFP*	Constructed in this study
E. coli DH5α	pADH2-OptCat::eGFP	Plasmid carrying $P_{ADH2\text{-}OptCat}$ with *eGFP*	Constructed in this study
E. coli DH5α	pADH2-AddAdr::eGFP	Plasmid carrying $P_{ADH2\text{-}AddAdr}$ with *eGFP*	Constructed in this study
E. coli DH5α	pADH2-OptAdr1-1::eGFP	Plasmid carrying $P_{ADH2\text{-}OptAdr1\text{-}1}$ with *eGFP*	Constructed in this study
E. coli DH5α	pADH2-OptAdr1-2::eGFP	Plasmid carrying $P_{ADH2\text{-}OptAdr1\text{-}2}$ with *eGFP*	Constructed in this study
E. coli DH5α	pADH2-OptAdr1-3::eGFP	Plasmid carrying $P_{ADH2\text{-}OptAdr1\text{-}3}$ with *eGFP*	Constructed in this study
E. coli DH5α	pADH2-AddAca::eGFP	Plasmid carrying $P_{ADH2\text{-}AddAca}$ with *eGFP*	Constructed in this study
E. coli DH5α	pADH2-AddTATA::eGFP	Plasmid carrying $P_{ADH2\text{-}AddTATA}$ with *eGFP*	Constructed in this study
E. coli DH10B	pADH2-NucOpt-OptCat::eGFP	Plasmid carrying $P_{ADH2\text{-}NucOpt\text{-}OptCat}$ with *eGFP*	Constructed in this study
E. coli DH10B	pADH2-wt::optHSA	Plasmid carrying P_{ADH2wt} with *optHSA* with its native secretion signal	Constructed in this study
E. coli DH10B	pADH2-NucOp::optHSA	Plasmid carrying $P_{ADH2\text{-}NucOpt}$ with *optHSA* with its native secretion signal	Constructed in this study
E. coli DH10B	pADH2-OptCat::optHSA	Plasmid carrying $P_{ADH2\text{-}OptCat}$ with *optHSA* with its native secretion signal	Constructed in this study
E. coli DH5α	pAOX1-wt::eGFP	Plasmid carrying native P_{AOX1} with *eGFP*	Constructed in this study

Table 3.2 (continued)

Microorganism	Plasmid	Comment	Source
E. coli DH5α	pAOX1-Adr1-1::eGFP	Plasmid carrying $P_{AOX1\text{-}Adr1\text{-}1}$ with *eGFP*	Constructed in this study
E. coli DH5α	pAOX1-Adr1-2::eGFP	Plasmid carrying $P_{AOX1\text{-}Adr1\text{-}2}$ with *eGFP*	Constructed in this study
E. coli DH5α	pAOX1-Adr1-3::eGFP	Plasmid carrying $P_{AOX1\text{-}Adr1\text{-}3}$ with *eGFP*	Constructed in this study
E. coli DH5α	pAOX1-Cat8-1::eGFP	Plasmid carrying $P_{AOX1\text{-}Cat8\text{-}1}$ with *eGFP*	Constructed in this study
E. coli DH5α	pAOX1-Cat8-2::eGFP	Plasmid carrying $P_{AOX1\text{-}Cat8\text{-}2}$ with *eGFP*	Constructed in this study
E. coli DH5α	pAOX1-Cat8-3::eGFP	Plasmid carrying $P_{AOX1\text{-}Cat8\text{-}3}$ with *eGFP*	Constructed in this study
E. coli DH5α	pAOX1-Aca2::eGFP	Plasmid carrying $P_{AOX1\text{-}Aca2}$ with *eGFP*	Constructed in this study
E. coli DH5α	pAOX1-Cat3Adr3::eGFP	Plasmid carrying $P_{AOX1\text{-}Cat3Adr3}$ with *eGFP*	Constructed in this study
E. coli DH5α	pAOX1-AcaCat3Adr3::eGFP	Plasmid carrying $P_{AOX1\text{-}AcaCAt3Adr3}$ with *eGFP*	Constructed in this study
E. coli DH5α	pAOX1-mod::eGFP	Plasmid carrying $P_{AOX1\text{-}Mod}$ with *eGFP*	Constructed in this study
E. coli DH10B	pAOX1-wt::optHSA	Plasmid carrying P_{AOX1wt} with *optHSA* with its native secretion signal	Constructed in this study
E. coli DH10B	pAOX1-mod::optHSA	Plasmid carrying $P_{AOX1\text{-}mod}$ with *optHSA* with its native secretion signal	Constructed in this study
E. coli DH10B	BB1_23_Adr1_KanMX	GGA Level 1 plasmid, construcred for CDS of *Adr1*	Constructed in this study
E. coli DH10B	BB1_23_Aca1_KanMX	GGA Level 1 plasmid, construcred for CDS of *Aca1*	Constructed in this study
E. coli DH10B	BB3aK_14_pTHI11_Adr1_RPS3TT	GGA Level 2 plasmid, used for overexpression of Adr1 TF in *P. pastoris*	Constructed in this study
E. coli DH10B	BB3aK_14_pTHI11_Aca1_RPS3TT	GGA Level 2 plasmid, used for overexpression of Aca1 TF in *P. pastoris*	Constructed in this study

Table 3.2 (continued)

Microorganism	Plasmid	Comment	Source
E. coli DH10B	HR-Adr1_BB1	Used for amplification of homologous region to knock-out Adr1	Constructed in this study
E. coli DH10B	HR-Aca1_BB1	Used for amplification of homologous region to knock-out Aca1	Constructed in this study
E. coli DH10B	Cas9_gRNA_Adr1_BB3	CRISPR/Cas9 plasmid to knock-out Cat8-1	Constructed in this study
E. coli DH10B	Cas9_gRNA_Aca1_BB3	CRISPR/Cas9 plasmid to knock-out Cat8-2	Constructed in this study
E. coli DH10B	HR-Adr1-Flag_BB1	Used for amplification of homologous region to add Flag-Tag on Adr1	Constructed in this study
E. coli DH10B	Cas9_gRNA_Adr1-Flag_BB3	CRISPR/Cas9 plasmid to	Constructed in this study
P. pastoris X-33	pADH2-wt::eGFP	*P. pastoris* strain used for screening $P_{ADH2\text{-}wt}$	Constructed in this study
P. pastoris X-33	pADH2-NucOpt::eGFP	*P. pastoris* strain used for screening $P_{ADH2\text{-}NucOpt}$	Constructed in this study
P. pastoris X-33	pADH2-AddCat::eGFP	*P. pastoris* strain used for screening $P_{ADH2\text{-}AddCat}$	Constructed in this study
P. pastoris X-33	pADH2-OptCat::eGFP	*P. pastoris* strain used for screening $P_{ADH2\text{-}OptCat}$	Constructed in this study
P. pastoris X-33	pADH2-AddAdr::eGFP	*P. pastoris* strain used for screening $P_{ADH2\text{-}Add\text{-}Adr}$	Constructed in this study
P. pastoris X-33	pADH2-OptAdr1-1::eGFP	*P. pastoris* strain used for screening $P_{ADH2\text{-}OptAdr1\text{-}1}$	Constructed in this study
P. pastoris X-33	pADH2-OptAdr1-2::eGFP	*P. pastoris* strain used for screening $P_{ADH2\text{-}OptAdr1\text{-}2}$	Constructed in this study
P. pastoris X-33	pADH2-OptAdr1-3::eGFP	*P. pastoris* strain used for screening $P_{ADH2\text{-}OptAdr1\text{-}3}$	Constructed in this study
P. pastoris X-33	pADH2-AddAca::eGFP	*P. pastoris* strain used for screening $P_{ADH2\text{-}AddAca}$	Constructed in this study
P. pastoris X-33	pADH2-AddTATA::eGFP	*P. pastoris* strain used for screening $P_{ADH2\text{-}AddTATA}$	Constructed in this study
P. pastoris X-33	pADH2-NucOpt-OptCat::eGFP	*P. pastoris* strain used for screening $P_{ADH2\text{-}NucOpt\text{-}OptCat}$	Constructed in this study

Table 3.2 (continued)

Microorganism	Plasmid	Comment	Source
P. pastoris X-33	pADH2-wt::optHSA	*P. pastoris* strain used for screening $P_{ADH2\text{-}wt}$ for hSA production	Constructed in this study
P. pastoris X-33	pADH2-NucOp::optHSA	*P. pastoris* strain used for screening $P_{ADH2\text{-}wt}$ for HSA production	Constructed in this study
P. pastoris X-33	pADH2-OptCat::optHSA	*P. pastoris* strain used for screening $P_{ADH2\text{-}wt}$ for HSA production	Constructed in this study
P. pastoris X-33	pAOX1-wt::eGFP	*P. pastoris* strain used for screening $P_{AOX1\text{-}wt}$	Constructed in this study
P. pastoris X-33	pAOX1-Adr1-1::eGFP	*P. pastoris* strain used for screening $P_{AOX1\text{-}Adr1}$	Constructed in this study
P. pastoris X-33	pAOX1-Adr1-2::eGFP	*P. pastoris* strain used for screening $P_{AOX1\text{-}Adr2}$	Constructed in this study
P. pastoris X-33	pAOX1-Adr1-3::eGFP	*P. pastoris* strain used for screening $P_{AOX1\text{-}Adr3}$	Constructed in this study
P. pastoris X-33	pAOX1-Cat8-1::eGFP	*P. pastoris* strain used for screening $P_{AOX1\text{-}Cat1}$	Constructed in this study
P. pastoris X-33	pAOX1-Cat8-2::eGFP	*P. pastoris* strain used for screening $P_{AOX1\text{-}Cat2}$	Constructed in this study
P. pastoris X-33	pAOX1-Cat8-3::eGFP	*P. pastoris* strain used for screening $P_{AOX1\text{-}Cat3}$	Constructed in this study
P. pastoris X-33	pAOX1-Aca2::eGFP	*P. pastoris* strain used for screening $P_{AOX1\text{-}Aca}$	Constructed in this study
P. pastoris X-33	pAOX1-Cat3Adr3::eGFP	*P. pastoris* strain used for screening $P_{AOX1\text{-}Cat3Adr3}$	Constructed in this study
P. pastoris X-33	pAOX1-AcaCat3Adr3::eGFP	*P. pastoris* strain used for screening $P_{AOX1\text{-}AcaCat3Adr3}$	Constructed in this study
P. pastoris X-33	pAOX1-mod::eGFP	*P. pastoris* strain used for screening $P_{AOX1\text{-}mod}$	Constructed in this study
P. pastoris X-33	pAOX1-wt::optHSA	*P. pastoris* strain used for screening $P_{AOX1\text{-}wt}$ for HSA production	Constructed in this study

Table 3.2 (continued)

Microorganism	Plasmid	Comment	Source
P. pastoris X-33	pAOX1-mod::optHSA	*P. pastoris* strain used for screening $P_{AOX1\text{-}mod}$ for HSA production	Constructed in this study
P. pastoris X-33	pADH2-wt::eGFP BB3aK_14_pTHI11_Aca1_RPS3TT	*P. pastoris* strain that over-expresses Aca1	Constructed in this study
P. pastoris X-33	pADH2-NucOpt::eGFP BB3aK_14_pTHI11_Aca1_RPS3TT	*P. pastoris* strain that over-expresses Aca1	Constructed in this study
P. pastoris X-33	pADH2-AddAca::eGFP BB3aK_14_pTHI11_Aca1_RPS3TT	*P. pastoris* strain that over-expresses Aca1	Constructed in this study
P. pastoris X-33	pAOX1-wt::eGFP BB3aK_14_pTHI11_Aca1_RPS3TT	*P. pastoris* strain that over-expresses Aca1	Constructed in this study
P. pastoris X-33	pAOX1-mod::eGFP BB3aK_14_pTHI11_Aca1_RPS3TT	*P. pastoris* strain that over-expresses Aca1	Constructed in this study
P. pastoris X-33	pADH2-wt::eGFP BB3aK_14_pTHI11_Adr1_RPS3TT	*P. pastoris* strain that over-expresses Adr1	Constructed in this study
P. pastoris X-33	pADH2-NucOpt::eGFP BB3aK_14_pTHI11_Adr1_RPS3TT	*P. pastoris* strain that over-expresses Adr1	Constructed in this study
P. pastoris X-33	pADH2-AddAdr::eGFP BB3aK_14_pTHI11_Adr1_RPS3TT	*P. pastoris* strain that over-expresses Adr1	Constructed in this study
P. pastoris X-33	pADH2-OptAdr1-3::eGFP BB3aK_14_pTHI11_Adr1_RPS3TT	*P. pastoris* strain that over-expresses Adr1	Constructed in this study

Table 3.2 (continued)

Microorganism	Plasmid	Comment	Source
P. pastoris X-33	pAOX1-wt::eGFP BB3aK_14_pTHI11_Adr1_RPS3TT	*P. pastoris* strain that over-expresses Adr1	Constructed in this study
P. pastoris X-33	pAOX1-mod::eGFP BB3aK_14_pTHI11_Adr1_RPS3TT	*P. pastoris* strain that over-expresses Adr1	Constructed in this study
P. pastoris X-33	pADH2-wt::eGFP pADH2-OptCat::eGFP BB3aK_14_pTHI11_Cat8-1_RPS3TT	*P. pastoris* strain that over-expresses Cat8-1	Constructed in this study
P. pastoris X-33	pADH2-NucOpt::eGFP pADH2-OptCat::eGFP BB3aK_14_pTHI11_Cat8-1_RPS3TT	*P. pastoris* strain that over-expresses Cat8-1	Constructed in this study
P. pastoris X-33	pADH2-OptCat::eGFP pADH2-OptCat::eGFP BB3aK_14_pTHI11_Cat8-1_RPS3TT	*P. pastoris* strain that over-expresses Cat8-1	Constructed in this study
P. pastoris X-33	pAOX1-wt::eGFP pADH2-OptCat::eGFP BB3aK_14_pTHI11_Cat8-1_RPS3TT	*P. pastoris* strain that over-expresses Cat8-1	Constructed in this study
P. pastoris X-33	pAOX1-mod::eGFP pADH2-OptCat::eGFP BB3aK_14_pTHI11_Cat8-1_RPS3TT	*P. pastoris* strain that over-expresses Cat8-1	Constructed in this study

Table 3.2 (continued)

Microorganism	Plasmid	Comment	Source
P. pastoris X-33	pADH2-wt::eGFP BB3aK_14_pTHI11_Cat8-2_RPS3TT	*P. pastoris* strain that over-expresses Cat8-2	Constructed in this study
P. pastoris X-33	pADH2-NucOpt::eGFP BB3aK_14_pTHI11_Cat8-2_RPS3TT	*P. pastoris* strain that over-expresses Cat8-2	Constructed in this study
P. pastoris X-33	pADH2-OptCat::eGFP	*P. pastoris* strain that over-expresses Cat8-2	Constructed in this study
P. pastoris X-33	pAOX1-wt::eGFP BB3aK_14_pTHI11_Cat8-2_RPS3TT	*P. pastoris* strain that over-expresses Cat8-2	Constructed in this study
P. pastoris X-33	pAOX1-mod::eGFP BB3aK_14_pTHI11_Cat8-2_RPS3TT	*P. pastoris* strain that over-expresses Cat8-2	Constructed in this study
P. pastoris X-33	pADH2-wt::eGFP HR-Aca1_BB1	*P. pastoris Δaca1*	Constructed in this study
P. pastoris X-33	pAOX1-wt::eGFP HR-Aca1_BB1	*P. pastoris Δaca1*	Constructed in this study
P. pastoris X-33	pAOX1-mod::eGFP HR-Aca1_BB1	*P. pastoris Δaca1*	Constructed in this study
P. pastoris X-33	pADH2-wt::eGFP HR-Adr1_BB1	*P. pastoris Δadr1*	Constructed in this study
P. pastoris X-33	pAOX1-wt::eGFP HR-Adr1_BB1	*P. pastoris Δadr1*	Constructed in this study
P. pastoris X-33	pAOX1-mod::eGFP HR-Adr1_BB1	*P. pastoris Δadr1*	Constructed in this study
P. pastoris X-33	pADH2-wt::eGFP HR-Cat8-1_BB1	*P. pastoris Δcat8-1*	Constructed in this study
P. pastoris X-33	pADH2-OptCat::eGFP HR-Cat8-1_BB1	*P. pastoris Δcat8-1*	Constructed in this study

Table 3.2 (continued)

Microorganism	Plasmid	Comment	Source
P. pastoris X-33	pAOX1-wt::eGFP HR-Cat8-1_BB1	*P. pastoris Δcat8-1*	Constructed in this study
P. pastoris X-33	pAOX1-mod::eGFP HR-Cat8-1_BB1	*P. pastoris Δcat8-1*	Constructed in this study
P. pastoris X-33	pADH2-wt::eGFP HR-Cat8-2_BB1	*P. pastoris Δcat8-2*	Constructed in this study
P. pastoris X-33	pADH2-OptCat::eGFP HR-Cat8-2_BB1	*P. pastoris Δcat8-2*	Constructed in this study
P. pastoris X-33	pAOX1-wt::eGFP HR-Cat8-2_BB1	*P. pastoris Δcat8-2*	Constructed in this study
P. pastoris X-33	pAOX1-mod::eGFP HR-Cat8-2_BB1	*P. pastoris Δcat8-2*	Constructed in this study
P. pastoris X-33	pADH2-wt::eGFP HR-Cat8-1_BB1 HR-Cat8-2_BB1	*P. pastoris Δcat8-1Δcat8-2*	Constructed in this study
P. pastoris CBS7435	HR-Adr1-Flag_BB1	*P. pastoris* that have Flag-Tagged Adr1	Constructed in this study

3.4 Genetic Engineering Methods

3.4.1 Plasmid Isolation

Plasmid isolation from *E. coli* cells was performed by GeneJet Plasmid Purification Kit (ThermoFisher, USA) according to the manufacturer's protocol. Prior to plasmid isolation, cells were grown in LB medium with appropriate antibiotics (Zeocin or Kanamycin) at 37°C for 16 h with shaking at 180 rpm. 2 mL from the culture were harvested by centrifugation at 12000 g for 5 mins and pellets were used for plasmid isolation. Isolated plasmid was stored at -20°C until further use.

3.4.2 Genomic DNA Isolation

Genomic DNA isolation from *P. pastoris* cells were performed using Wizard Genomic DNA Purification Kit (Promega, USA) according to the manufacturer's instructions. *P. pastoris* cells were grown in YPD medium including appropriate antibiotics (Zeocin or Geneticin) at 25°C for 20 h with shaking at 225 rpm. 2 mL of the culture were harvested by centrifugation cells at 13000 g for 2 mins and pellets were used for genomic DNA isolation. Isolated DNA was stored at -20°C until further use.

3.4.3 RNA Isolation and cDNA Synthesis

For total RNA isolation form *P. pastoris* cells, pellets from the samples collected during the exponential growth phase were resuspended in 1 mL TRI Reagent (Sigma, USA). Cells were disrupted by glass beads using a ribolyser (MP Biomedicals, USA) for 40 seconds at 5.5 ms^{-1}. RNA was extracted by chloroform and isoproponal used for precipitation. After washing twice with 75% ethanol, RNA was dissolved in nuclease-free water. After DNase treatment (Invitrogen, USA), concentration of RNA was measured by Nanodrop. Isolated RNA was stored at -80°C until further use. cDNA synthesis was performed with Biozym cDNA synthesis kit (Biozym,

Germany) using oligo d(T)23 VN primer (NEB, USA). cDNA was stored at -20°C until further use.

3.4.4 Gel Elution

Gel elution was performed by GeneJet Gel Extraction Kit (ThermoFisher, USA) or Wizard SV Gel and PCR Cleanup System (Promega, USA) to purify a single DNA product. DNA was purified according to the manufacturer's instruction after agarose gel electroporation. Purified DNA fragment was eluted in distilled water and stored at -20°C until further use.

3.4.5 PCR Purification

Purification of PCR amplified DNA fragments was carried out by GeneJet PCR Purification Kit (ThermoFisher, USA) or Wizard SV Gel and PCR Cleanup System (Promega, USA) according to the manufacturer's instructions.

3.4.6 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used for visualization, separation, and extraction of DNA fragments. 1% agarose gel was prepared in TAE buffer including ethidium bromide or Midori Green. DNA was visualized after running at 120 V for 30-60 min depending on the size of the fragments. DNA Ladders were used for size determination.

3.4.7 DNA Sequencing

Sequences of the amplified DNA and purified plasmids were verified by METU Central Laboratory, Ankara or Microsynth AG, Balgach, Switzerland.

3.4.8 Transformation of *E. coli*

The constructed plasmids (given in Section 3.3) were used for transformation of *E. coli* DH10B or DH5α. Competent *E. coli* cells were prepared by RbCl or $CaCl_2$ as described, as follows.

3.4.8.1 RbCl Method

E. coli cells from LB plate was inoculated into 10 mL of LB medium and incubated overnight at 37˚C at 180 rpm. 2 mL of overnight culture was transferred into 200 mL LB medium and incubated for approximately 2 hours until it reaches OD_{600} 0.6. The cells were then cooled down on ice for 10 min and centrifuged for 5 min at 4000 g at 4˚C. Pelleted cells were resuspended in 80 mL of (0.4 V of the culture) ice-cold TFB1 solution (Table 3.3). After repeating the centrifugation step, cells were resuspended in 8 mL of (0.04 V of the culture) ice-cold TFB2 solution (Table 3.4). Competent cells were then immediately used or stored at -80˚C for further use in 100 µL aliquots. Ligation mixture was mixed with 100 µL of competent *E. coli* cells and incubated on ice for 30 min. Thereafter, cells and DNA mix were heat shocked at 42˚C for 90 seconds. Following the incubation of the cells on ice for 5 min, 1 mL of LB medium was added to the mixture and cells were incubated at 37˚C with shaking for 1 h. 25-100 µL of the cells were streaked on LB plates with appropriate antibiotics. Plates were incubated at 37˚C for 16-20 h until the colonies appeared.

Table 3.3 Composition of TFB1 solution.

Reagent	**Concentration**
Potassium acetate	30 mM
$CaCl_2$	10 mM
$MnCl_2$	50 mM
RbCl	100 mM
Glycerol	15%

*Adjust pH to 5.8 with 1 M acetic acid, filter sterilize and store at 4˚C.

Table 3.4 Composition of TFB2 solution.

Reagent	Concentration
MOPS (or PIPES)	100 mM
CaCl2	75 mM
RbCl	10 mM
Glycerol	15%

*Adjust pH 6.5 with 1 M KOH, filter sterilize and store at 4˚C.

3.4.8.2 $CaCl_2$ Method

A single cell from LB plate was inoculated into 50 mL of LB medium and incubated at 37˚C at 180 rpm for 3.5-4 h until the OD_{600} reaches approximately 0.35-0.4. The cells were transferred to a 50 mL ice-cold falcon, stored on ice for 10 minutes and centrifuged for 10 min at 2700 g at 4 ˚C. Pellet was resuspended in 30 mL ice-cold 80 mM $MgCl_2$-20 mM $CaCl_2$ solution by swirling and pelleted again by centrifuging the broth at 2700 g for 10 minutes at 4°C. Finally, pellet was resuspended in 2 mL ice cold 0.1 M $CaCl_2$ by vortexing. 50 µL of this solution was transferred to a fresh Eppendorf tube and mixed with ligation mixture. After the incubation on ice for 30 min, the mixture was heat shocked at 42˚C for 90 seconds and cooled down on ice for 3-5 min. 900 µL of sterile LB was added to the mixture and incubated at 37°C for 1 h. 25-100 µL of the cells were streaked on LB plates with appropriate antibiotics. Plates were incubated at 37˚C for 16-20 h until the colonies appeared.

3.4.9 Transformation of *P. pastoris*

For the integration of the expression cassettes, plasmids were linearized and purified by gel elution. Transformation of *P. pastoris* X-33 or CBS7435 cells was either carried out by LiCl method or electroporation.

3.4.9.1 LiCl method

The solutions required for the transformation, *i.e.* 50% PEG (3350) and 1 M LiCl, were prepared by filter-sterilizing and used within a month. *P. pastoris X-33* from glycerol stock was streaked onto YPD plate and incubated for 48 hours at 30°C. A single colony was inoculated to 50 mL YPD medium and incubated to an OD_{600} of 0.8 to 1.0 at 30°C at 200 rpm for approximately 12-14 hours. The cells were then harvested by centrifugation at 4000 g for 5 minutes at room temperature. Cells were washed with 25 mL sterile water and centrifuged at 1500 g for 10 minutes at room temperature. After removing supernatant, pellet was resuspended in 1 mL 100 mM filter-sterilized LiCl, transferred to a 1.5 mL Eppendorf tube and centrifuged at maximum speed for 15 seconds. LiCl was removed with a pipette and cells were resuspended in 400 µL 100 mM filter-sterilized LiCl. For each transformation, 50 µL of cell suspension was dispensed into 1.5 mL Eppendorf tubes and immediately centrifuged at maximum speed for 15 seconds. LiCl was removed with a pipette. For transformation of the cells 240 µL of %50 PEG, 36 µL of 1 M LiCl, 25 µL of 2 mg/mL single-stranded DNA (denatured prior to transformation and used immediately) and 0.5-1 µg linearized plasmid DNA in 50 µL sterile water were added to each tube in the given order and vortexed vigorously until the cell pellet was completely mixed. The tube was incubated at 30°C for 30 min without shaking, then heat shocked in a water bath at 42°C for 20-25 min. Cells were pelleted by centrifuging at 6000-8000 rpm for 15 seconds, transformation solution was removed and pellet was gently resuspended in 1 mL YPD and incubated at 30°C with shaking. After 2 hours of incubation, 25-100 µL of the medium was spread on YPD plates containing appropriate antibiotics and incubated for 2-3 days at 30°C.

3.4.9.2 Electroporation

For electroporation of *P. pastoris* X33 or CBS7435, cells from glycerol stock was streaked onto YPD plate and incubated for 48 hours at 25°C. Single cell from the plate was transferred into 10 mL YPD (including antibiotics if the parental strain already

carrying an antibiotic resistance) and incubated at 25°C with shaking at 180 rpm overnight. The grown cells were then used for inoculation to main culture including 100 mL YPD according to the formula shown below:

$$V_{inoc}\ (\mu L) = \frac{OD_m \times V_m}{e^{\mu \times t}} \times \frac{1000}{OD_{pre}}$$

- OD_m: OD_{600} of main culture after time t (use OD_{600} of 3 for the calculation – it should be between 1.8-3.0.)
- V_m: Volume of the main culture (mL)
- t_m: Incubation time for main culture (h)
- μ: Growth rate of the cells (use 0.25 h^{-1} for *P. pastoris*-wt in YPD at 25˚C)
- OD_{pre}: OD_{600} of the pre-culture

Grown cells were then centrifuged at 1500 g, 4˚C for 5 min. 10 mL of pre-warmed YPD including 20 mM HEPES (pH 8) and 25 mM DTT was added to the pelleted cells and incubated for 30 min by shaking at 25˚C. After addition of 40 mL of ice cold water, the cells were centrifuged at 1500 g, 4˚C for 5 min and the supernatant is discarded. 45 mL of ice old HEPES (1mM, pH 8) was added to the pelleted cells and centrifuged at 1500 g, 4˚C for 5 min. The pellet was then resuspended in 45 mL ice cold 1 M sorbitol and centrifuged at 1500 g, 4˚C for 5 min. Finally, the cells were resuspended in 500 µL of ice cold 1M sorbitol. 80 µL of these competent cells were used for transformation and mixed very gently with 0.5-1 µg linearized plasmid dissolved in 20 µL water. The mixture was incubated on ice for 5 min and electroporation was performed at 2000 V, 25 µF and 200 Ω. Immediately after transformation, the cells were mixed with 1 mL ice cold YPD and transformed cells were then regenerated by incubating at 30ºC for 1.5 h in YPD medium and then

streaked on YPD plates including appropriate antibiotics and incubated at 25˚C for 2-3 days.

3.5 Construction of the strains and the plasmids

3.5.1 Construction of Recombinant Plasmids $P_{ADH2\text{-}v}$::eGFP and $P_{AOX1\text{-}v}$::eGFP

The base plasmid pPICZα-A*::egfp* (Ata et al., 2017) was used in the construction of recombinant expression systems (Sequence is given in Appendix C). For the intracellular eGFP synthesis, P_{AOX1} and α-factor signal sequence were replaced with the designed promoter variants.

First, noncutters of $P_{ADH2\text{-}v}$, $P_{AOX1\text{-}v}$, *eGFP* and pPICZα-A*::egfp* were determined with Restriction mapper 3 (available online: http://www.restrictionmapper.org/). In the pPICZα-A*::egfp* plasmid, BglII restriction enzyme (RE) recognition site is found 5' of the P_{AOX1}; however, it cuts P_{ADH2} at position 755. On the other hand, *BamH*I is a noncutter of $P_{ADH2\text{-}v}$ and BamHI and BglII could form compatible cohesive ends with each other. In the primer design at the 5' of the $P_{ADH2\text{-}v}$ a *BamH*I recognition sites for cloning and a *Pst*I recognition site for lineerization of the recombinant plasmid was included. To achieve seamless cloning with $P_{ADH2\text{-}v}$ and *eGFP*, a Type IIS RE, *Bts*$^{\alpha}$I was used. Type IIS restriction enzyme recognize asymmetric DNA sequences and cleave outside of their recognition sequence. Bts$^{\alpha}$I RE cuts eGFP at position 214. For seamless cloning purpose, the first 216 nucleotides of eGFP were added to 3' of the designed promoter variants namely $P_{ADH2\text{-}wt}$, $P_{ADH2\text{-}NucOpt}$, $P_{AOX1\text{-}mod}$ and synthesized by GenScript, USA. Synthetic genes were cloned into PUC57 by GenScript.

Promoter variants together with the first 216 nucleotides of eGFP were amplified by Q5 High Fidelity DNA Polymerase (NEB, USA). Content of PCR mixture and thermal cycler conditions are given, as follows:

Table 3.5 Composition of PCR mixture

Component	50 µl Reaction	Final Concentration
5X Q5 Reaction Buffer	10 µl	1X
1 mM dNTPs	10 µl	200 µM
10 µM Forward Primer	2.5 µl	0.5 µM
10 µM Reverse Primer	2.5 µl	0.5 µM
Template DNA	2 µl	< 1,000 ng
Q5 High-Fidelity DNA Polymerase (NEB)	0.5 µl	0.02 U/µl
Nuclease-Free Water	to 50 µl	

Table 3.6 Thermocycling Conditions with Q5 Polymerase (NEB)

Step	Temperature	Time
Initial Denaturation	98°C	30 seconds
30 Cycles	98°C	10 seconds
	50-72°C	30 seconds
	72°C	30 seconds/kilobase
Final Extension	72°C	2 minutes
Hold	4–10°C	Forever

*For $P_{ADH2\text{-}v}$ and $P_{AOX1\text{-}v}$ PCR annealing temperature was set to 71 °C and 68 °C, respectively.

Mutations into $P_{ADH2\text{-}v}$ and $P_{AOX1\text{-}v}$ were introduced by two-step overlap extension PCR using Q5 DNA polymerase. In the first step of overlap extension PCR, reverse mutagenic primer was combined with Forward P_{ADH2} or Forward P_{AOX1}, and forward mutagenic primer was combined with Reverse P_{ADH2} or Reverse P_{AOX1}.

Amplified $P_{ADH2\text{-}v}$ were double digested by *BamH*I and *Btsα*I REs (Table 3.7) while amplified $P_{AOX1\text{-}v}$ and base vector plasmid pPICZαA*::eGFP* were double digested by *Bgl*II and *Btsα*I REs (Table 3.8). RE double digestion conditions are given, as follows.

Table 3.7 Double Digestion with *BamH*I-HF and *Btsα*I (NEB)

Component	**50 µl reaction**
DNA	1 µg
10X CutSmart Buffer	5 µl
*BamH*I-HF	1 µl (10 U)
Nuclease-free water	to 50 µl

- Incubate at 37°C for overnight.
- Add 1 µl of *Btsα*I (NEB) for 50 µl reaction.
- Incubate at 55 °C for overnight.
- Stop the reaction by adding 10 µl of 6X loading dye.

Table 3.8 Double Digestion with *Bgl*II and *Btsα*I

Component	**50 µl reaction**
DNA	1 µg
10X CutSmart Buffer	5 µl
*Btsα*I (NEB)	1 µl (10 U)
Nuclease-free water	to 50 µl

- Incubate at 55°C for overnight.
- Add 2 µl 2.6 M NaCl to adjust the salt concentration to 100mM.
- Add 1 µl of *Bgl*II (NEB) for 50 µl reaction.
- Incubate at 37 °C for overnight.
- Stop the reaction by adding 10 µl of 6X loading dye.

After restriction digestion vector and insert DNA was purified by gel elution kit. Purified insert and vector genes were ligated with T4 DNA Ligase (Thermo Scientific) by using a ligation ratio of 3:1. Composition of the sticky-end ligation reaction mixture is given, as follow.

Table 3.9 Composition of the sticky-end ligation reaction mixture

Component	Amount
Linear Vector DNA	50 ng
Insert DNA	3:1 molar ratio
10X T4 DNA Ligase Buffer	2 µl
T4 DNA Ligase	0.2 µl (1 Weiss Unit)
Nuclease-free water	to 20 µl

- Incubate 16 h at 16 °C.
- Heat inactivation of T4 DNA Ligase at 65 °C for 10 min or 70°C for 5 min.

After ligation, transformation of wild type *E.coli* DH5α cells was performed with calcium chloride method. Transformation mixture was inoculated into selective Zeocin containing media. After 16-18 h of incubation at 37 °C, single colonies were selected and inoculated into a fresh Zeocin containing LSLB Agar media. Colony PCR was performed and among positives, four colonies were selected and plasmid isolation was performed with Plasmid isolation kit. Sequence of the insert was evaluated by gene sequencing.

3.5.2 Construction of Recombinant Plasmids $P_{ADH2\text{-}v}$::*HSA* and $P_{AOX1\text{-}v}$:: *HSA*

For the HSA synthesis and secretion, in the base plasmid the promoter P_{AOX1}, α-factor signal sequence, and *eGFP* gene were replaced with the novel engineered promoters (NEPVs) and *HSA* gene. For extracellular HSA production, *HSA* was cloned with its native secretion signal. NEPVs and engineered promoters (NEPVs) and *HSA* gene. For extracellular HSA production, *HSA* were cloned by two-step overlap extension PCR to achieve a semaless cloning without adding any additional nucleotides between promoter and coding sequence. Gene amplifications were performed with Q5 DNA Polymerase as stated in Table 3.5 and Table 3.6.

pPICZα-A*::eGFP* vector was double digested with the REs *Bgl*II and *Kpn*I, and then purified by using gel purification kit. For HSA producing recombinant plasmid construction, $P_{ADH2\text{-}V}$-*HSA* gene fragments were double digested by *BamH*I and *Kpn*I (Table 3.10) whereas $P_{AOX1\text{-}V}$-*HSA* gene fragments were double digested with *Bgl*II and *Kpn*I (Table 3.11) and purified with gel extraction kit.

Table 3.10 Double Digestion with *BamH*I-HF and *Kpn*I-HF

Component	**50 µl reaction**
DNA	1 µg
10X CutSmart Buffer	5 µl
*BamH*I-HF (NEB)	1 µl (10 U)
*Kpn*I-HF (NEB)	1 µl (10 U)
Nuclease-free water	to 50 µl

- Incubate at 37°C for overnight.
- Stop the reaction by adding 10 µl of 6X loading dye.

Table 3.11 Double Digestion with *Bgl*II and *Kpn*I

Component	**50 µl reaction**
DNA	1 µg
10X CutSmart Buffer	5 µl
*Kpn*I-HF	1 µl (10 U)
Nuclease-free water	to 50 µl

- Incubate at 37°C for overnight.
- Add 2 µl 2.6 M NaCl to adjust the salt concentration to 100mM.
- Add 1 µl of *Bgl*II (NEB) for 50 µl reaction.
- Incubate at 37 °C for overnight.
- Stop the reaction by adding 10 µl of 6X loading dye.

After restriction digestion vector and insert DNA was purified by gel elution kit. Purified insert and vector genes were ligated with T4 DNA Ligase (Thermo Scientific) by using a ligation ratio of 3:1 (Table 3.9). After ligation, transformation of wild type *E.coli* DH10B cells was performed with rubidium chloride based chemical method. Transformation mixture was inoculated into selective Zeocin containing media. After 16-18 h of incubation at 37 °C, single colonies were selected and inoculated into a fresh Zeocin containing LSLB Agar media. Colony PCR was performed and among positives, four colonies were selected and plasmid isolation was performed with Plasmid isolation kit. Sequence of the insert was evaluated by gene sequencing.

3.5.3 Construction of *P. pastoris* Strains Carrying $P_{ADH2\text{-}v}$ and $P_{AOX1\text{-}v}$

Verified $P_{ADH2\text{-}v}$*::eGFP*, $P_{AOX1\text{-}v}$*::eGFP*, $P_{ADH2\text{-}v}$*::HSA* and $P_{AOX1\text{-}v}$*::HSA* plasmids were linearized in order to promote the insertion of the plasmid into the genome prior to *P. pastoris* transformation. Linearization was performed by incubation the plasmids at 37°C for 5-6 h according to the amount of the DNA in the reaction mixture. Lineerization of $P_{ADH2\text{-}v}$ were performed by PstI RE digestion whereas lineerization of $P_{AOX1\text{-}v}$ were performed with BglII RE digestion. Linearization reaction compositions are given in Table 3.12 and 3.13. Linearized product was purified as described in Section 3.4.4. Sterile ultrapure water (autoclaved and 0.2 µm filter-sterilized) was used for elution. The concentration of the purified product was measured by Nanodrop. Amount of the linearized and purified plasmid was adjusted to 0.5-1 µg in 50 µL sterile water. Transformation of *P. pastoris* X33 was performed by LiCl transformation for intracellular eGFP producing strains and by electroporation for extracellular eGFP producing strains as described in Section 3.4.9.1 and 3.4.9.2, respectively. For each construct, ca. 40 single colonies were selected and transferred to new plates and incubated at 30°C for 24 h. Integration of the expression casettes to *P. pastoris* X33 was confirmed by PCR. At least 10 positive clones for each construct was selected and used in the screening experiments.

Table 3.12 Linearization of recombinant plasmids with *Pst*I.

Compound	Amount
DNA	1 µg
10X CutSmart Buffer	5 µl
*Pst*I (NEB)	1 µl
Water	Up to 50 µl

- Incubate at 37°C for 1 hour.
- Stop the reaction by adding 10 µl of 6X gel loading dye to 50 µl reaction.

Table 3.13 Linearization of recombinant plasmids with *Bgl*II RE.

Compound	Amount
DNA	1 µg
10X 3.1 Buffer	5 µl
*Bgl*II (NEB)	1 µl
Water	Up to 50 µl

- Incubate at 37°C for overnight.
- Stop the reaction by adding 10 µl of 6X gel loading dye to 50 µl reaction.

3.5.4 Construction of Overexpression and Knockout Plasmids and Strains

In the PhD thesis study, 4 different TFs were selected as the targets for over expression or knock-out: *ACA1*, *ADR1*, *CAT8-1* and *CAT8-2*. The sequences of these genes were retrieved from http://pichiagenome-ext.boku.ac.at:8080/apex/f?p=100:1. Chromosomal regions of the selected transcription factors are given in Table 3.14.

Table 3.14 Chromosomal locations of the coding regions of the selected transcription factors.

Transcription factor	Chromosomal location of the gene
Aca1	*PAS_chr3_0135*
Adr1	*PAS_chr4_0487*
Cat8-1	*PAS_chr2-1_0757*
Cat8-2	*PAS_chr4_0540*

Golden gate assembly (GGA) was used for the construction of the plasmids which are used for over-expression or knocking-out the desired gene. GGA is an efficient and seamless method which allows the scientists to assembly multiple DNA fragments into a backbone vector by using Type IIS restriction enzymes (REases) and T4 DNA Ligase simultaneously (Engler et al., 2008; Prielhofer et al., 2017). REases such as *Bsa*I, *Bpi*I (*Bbs*I) or *Bsm*BI cleave the DNA outside of the recognition site; therefore, once two fragments are overlapped and assembled after their cleavage by Type IIS REases, no scar sequence is left. Moreover, multiple fragments can be assembled simultaneously and as the restriction site is removed from the assembled product, digestion and ligation can be performed at the same time. One drawback of this method is that the occasional presence of one or more internal TypeIIS REases sites found on the gene of interest sequences. This problem, however, can easily be eliminated by designing primers which create single nucleotide mutations but keeps the amino acid as same as the original sequence.

3.5.4.1 Overexpression strains

For over-expression, a repressible *THI11* promoter (annotated as P_{THI11}) was selected since this promoter expression capacity can be controlled by the presence or absence of thiamine. A two-level Golden Gate assembly (GGA) method was used for the construction over-expression cassettes. First, the plasmids that carry the CDS of the TF were constructed. Thereafter, these plasmids were used to construct the plasmid that carry the entire expression module: Promoter, CDS and the terminator. Level 1

plasmids (BB1) are the plasmids that carry the promoter, CDS or the terminator regions. BB1 plasmid that carry the promoter, *i.e.* P_{THI11} or terminator, *i.e. RPS3tt* were constructed by Prielhofer et al., 2017.

The genes of the corresponding CDSs were amplified from the genome by PCR. Internal *Bsa*I or *Bpi*I sites within the CDSs were eliminated by designing primers which enable to overlap the modified regions by GGA or ordering *in-vitro* synthesized gBlocks where these restriction sites were mutated without altering the original codons. The primers used for amplification are given in Table 3.1. All PCR reactions were performed by Q5 DNA Polymerase as explained in Section 3.5.1. After the amplification of DNA fragments, they were purified by gel elution and their concentrations were measured by Nanodrop. The backbone plasmid (BB1) carrying a kanamycin resistance gene was used for the insertion of the TF CDSs. Golden gate assembly reaction composition and condition are given in Table 3.15 and 3.16. A 3:1 molar ratio was used for the integration between PCR products and the plasmids.

Table 3.15 GGA reaction composition for construction of Level 1 plasmids for CDSs.

Reagent	**Amount (µL)**
BsaI (20 U/uL)	1
T4 Ligase (1:10 diluted;40 U/ul)	1
CutSmart buffer (10X)	2
ATP (10 mM)	2
BB1 plasmid(40 nM)	1
DNA Fragments (40 nM)	3 (each)
Water	Up to 20 µL

Table 3. 16 GGA conditions for construction of Level 1 plasmids for CDSs.

Temperature (˚C)	Duration	Number of cycles
37	5 min	50
16	5 min	
37	10 min	1
50	30 min	1
80	10 min	1
23	10 sec	1

Entire GGA mixture was then immediately used for *E. coli* transformation by RbCl transformation method as described in Section 3.4.8.1 or stored at -20˚C. An additional 2 h digestion with *Bsa*I was performed prior to transformation to eliminate the fragments with *Bsa*I recognition site to discard the false positive products as much as possible. Transformed cells were streaked onto LB plates including 50 μg/mL Kanamycin and incubated at 37°C for 16-20 h until the colonies appear. Single colonies were transferred to fresh LB plates and their plasmids were isolated as described in Section 3.4.1. Plasmids were checked with PCR and then confirmed by gene sequencing.

Level 2 plasmids are the plasmids that carry the entire expression cassette including the promoter (P_{THI11}), CDS (*ACA1*, *ADR1*, *CAT8-1* and *CAT8-2*) and the terminator (*RPS3tt*). Level 2 plasmids (BB3) are the plasmids which are used for *P. pastoris* cloning in the next steps.

After selecting and verifying the clones that carry the true BB1 plasmids, they were isolated and their concentration was measured by Nanodrop. BB3 that has a Kanamycin/G418 resistance gene and a homologous integration site to AOX2 transcription terminator region are used as backbone plasmids. *Bpi*I was used as the TypeIIS RE instead of *Bsa*I as the BB3 plasmid was designed with *Bpi*I recognition

sites. GGA procedure and the reaction conditions are given in Table 3.17 and Table 3.18.

Table 3.17 GGA reaction composition for construction of Level 2 plasmids.

Reagent	**Amount (uL)**
*Bpi*I 10 U/uL)	1
T4 Ligase (1:10 diluted; 40 U/ul)	1
CutSmart buffer (10X)	2
ATP (10 mM)	2
BB3 plasmid(40 nM)	1
BB1 vector with promoter, pCS1 (40 nM)	1
BB1 vector with CDS (40 nM)	1
BB1 vector with terminator, RPS3TT (40 nM)	1
Water	Up to 20 uL

Table 3.18 GGA conditions for construction of Level 2 plasmids.

Temperature (˚C)	**Duration**	**Number of cycles**
37	5 min	8
16	5 min	
37	10 min	1
50	30 min	1
80	10 min	1
23	10 sec	1

GGA mixture was then immediately used for *E. coli* transformation or stored at -20˚C. An additional digestion with *Bpi*I for 2 h was performed prior to transformation to eliminate the fragments with *Bpi*I recognition site to discard the false positive products as much as possible. Transformed cells were streaked onto LB plates including 50 µg/mL Kanamycin and incubated at 37°C for 16-20 h until the colonies

appear. Single colonies were transferred to fresh LB plates and integraion of gene was confirmed with colony PCR, then their plasmids were isolated as described in Section 3.4.1.

After confirmation of the Level 2 plasmid, linearization with *Asc*I and subsequent gel purification (3.4.4) of the plasmids were carried out to integrate the over-expression cassettes to the genome of *P. pastoris* carrying $P_{ADH2\text{-}v}$*::eGFP* or $P_{AOX1\text{-}v}$*::eGFP*. *P. pastoris* transformation was performed by electroporation as described in Section 3.4.9.2 by using 0.5-1 µg of each linearized and purified plasmid. Transformed cells were then regenerated by incubating at 30ºC for 3 h in YPD medium and then streaked on YPD plates including 500 µg/mL G418 and 25 µg/mL Zeocin. Cells were incubated at 25ºC for 2-3 days and single colonies were transferred to fresh YPD plates. For each construct, lest 10 different true clones were selected for further screening experiments to investigate the effect of overexpressed TFs on *egfp* production by $P_{ADH2\text{-}v}$ and $P_{AOX1\text{-}v}$.

3.5.4.2 Knock-out Strains

Pichia pastoris TF knock-out strains were constructed using *P. pastoris* Crispr/Cas9 system (Prielhofer et al., 2017). A one-level GGA was performed to obtain the plasmids that carry the knock-out cassettes. Prior to the GGA reaction, the homologous sites were amplified by PCR from the gDNA of *P. pastoris* CBS7435. These homologous sites were selected from upstream (5') and downstream (3') of the target gene with an approximate 1000 bp length. The primers used for amplification of these homologous fragments and elimination of internal *Bsa*I sites are given in Table 3.1. The flanking homologous regions of the target gene approximately 1000 bp upstream and 1000 bp downstream were fused with each other by two-step overlap extension PCR as explained in Section 3.5.1. Internal *Bsa*I sites found on these homologous regions were eliminated by designing primers which enable to overlap the modified regions by GGA. Golden gate assembly reaction composition and conditions were applied as given in Table 3.15 and 3.16.

A BB1 vector containing humanized Cas9 that was cloned with *P. pastoris* P_{PFK300} and terminator Sc*CYCtt* was used for construction of TF-knockout Crispr/Cas9 plasmids. In order to find Protospacer adjacent motif (PAM) sequence, Chop chop (available online http://chopchop.cbu.uib.no/) with *P. pastoris* genome was used. Single guide RNA was designed and amplified based on the identified PAM sequence. In the designed Carisp/Cas9 plasmid, the designed single guide RNA was inserted between P_{GAP} and terminator *RPS25Att* by GGA experiments as explained in Table 3.15 and 3.16.

Designed homologous DNA region plasmid and Crispr/Cas9 plasmids transformed into *E. coli* by RbCl method as stated in 3.4.8.1. Transformed cells were streaked onto LB plates including 50 µg/mL Kanamycin and incubated at 37°C for 16-20 h until the colonies appear. Single colonies were transferred to fresh LB plates and integraion of gene was confirmed with colony PCR, then their plasmids were isolated as described in Section 3.4.1 and confirmed by gene sequencing.

For the integration of the knock-out cassettes, the plasmids carrying the homologous DNA for knock-out were used as the templates to amplify the fragments. 3-5µg of amplified homologous DNA for TF knock-out and 0.5-1µg of circular Crispr/Cas9 plasmid DNA was simultaneously used for *P. pastoris* transformation by electroporation as described in Section 3.4.9.2. Transformed cells were then regenerated by incubating at 30ºC for 3 h in YPD medium and then streaked on YPD plates including 500 µg/mL G418 and 25 µg/mL Zeocin. The cells were incubated at 30°C for 2-3 days and single colonies were transferred to fresh YPD plates. Knock-out strains were controlled by PCR using designed primers at outside of the targeted deletion site. After confirmation of the TF deletions, true transformed inoculated to 25 µg/mL Zeocin without Geneticin to lose Crispr/Cas9 plasmid. Knock-out strains were passaged at least three times and then their resistnace to Geneticin was controlled to be sure about the absence of Crispr/Cas9 plasmid. For each knock-out *P. pastoris* strain, least 10 different true clones were selected for further screening

experiments to investigate the effect of TF knock-out on *egfp* production by $P_{ADH2\text{-}v}$ and $P_{AOX1\text{-}v}$.

3.6 Growth Media and Conditions

Growth media used in this study were either autoclaved at 121°C for 20 min or filter sterilized (22 μM).

3.6.1 Complex Media

3.6.1.1 YPD

YPD was used for growing *P. pastoris* cells either from glycerol stock or for transformation. It was sterilized by autoclave and appropriate amount of the desired antibiotics was added after cooling down to approximately 50-60°C (Table 3.19).

Table 3.19 YPD medium.

Component	Concentration (g/L)
Yeast extract	10
Peptone	20
Glucose	20
Agar (if necessary)	20

*Glucose was filter sterilized and added after autoclaving yeast extract, peptone and agar (if necessary).

3.6.1.2 LB

LB medium was used for growing *E. coli* cells either from glycerol stock or for transformation. It was sterilized by autoclave and appropriate amount of the desired antibiotics was added after cooling down to approximately 50-60°C (Table 3.20).

Table 3.20 LB medium.

Component	Concentration (g/L)
Yeast extract	5
Tryptone	10
NaCl	10
Agar (if necessary)	15

3.6.2 Screening media and conditions

3.6.2.1 eGFP and rHSA production

Screening of the expression of intracellular eGFP and the second model protein extracellular rHSA, and comparison of promoter variants as well as overexpression and knockout strains was carried out using 24-deep well plates (Whatman, UK). Randomly selected clones were grown in 2 mL YP medium (YPD medium without glucose) including 25 μg/mL Zeocin and/or 500 μg/mL G418 (if necessary) for 20 h and then transferred into ASMv6 medium with the addition of the selected carbon source (Table 3.21). For limited glucose condition %25 m2p kit Polysaccharide and %0.7 enzyme was used (m2p-labs GmbH, Germany). Different fermentation conditions used for eGFP production is given Table 3.22.

Table 3.21 ASMv6 medium used for screening eGFP and HSA production.

Component	Concentration (g/L)
$(NH_4)_2HPO_4$	6.3
$MgSO_4*7H_2O$	0.49
$(NH_4)_2SO_4$	0.8
KCl	2.64
$CaCl_2*2H_2O$	0.0535
Citric acid monohydrate	22
NH_4OH (25%)	20 mL
PTM Trace Element Solution	1.47 mL
Biotin (0.2 g/L)	2 mL

Table 3.22 Cultivation conditions for eGFP production.

Name of the condition	Carbon source	Concentration	Start OD_{600}	Cultivation time
Excess Glucose	Glucose	20 g/L	0.1	20 h
Excess Glycerol	Glycerol	20 g/L	0.1	20 h
Limited Glucose	Polysaccharide	%25 (v/v)	1	20 h
	Enzyme	%0.7 (v/v)	1	20 h
Methanol	Methanol	1% (v/v)	1	20 h
Ethanol	Ethanol	2% (v/v)	1	20 h

The cells were harvested at t = 19-20 h of the cultivation, diluted in 1 M PBS to a final OD_{600} of 0.4 and were used for measurement of eGFP fluorescence with flow cytometer immediately.

For extracellular HSA production, the fermentations were initiated at OD_{600} = 1 in ASMv6 minimal medium in 24-well plates. At t = 0 1% (v/v) ethanol was added into each bioreactor; thereafter, at t = 5, t = 17, t = 29, and t = 41 h, 1% (v/v) ethanol was also added to induce r-protein production, and the cells were harvested at t = 49 h. For methanol induction, at t = 0 0.5% (v/v) methanol was added into each bioreactor; thereafter, at t = 5, t = 17, t = 29, and t = 41 h, 1% (v/v) methanol was also added to

induce r-protein production, and the cells were harvested at t = 49 h. Supernatants were collected for rHSA determinantion and cell growth was determined by wet cell weight (WCW) measurement.

3.7 Analyses

3.7.1 Cell concentration

Wet cell concentrations were calculated related to OD_{600} and used for further calculations. Samples were collected in the exponential growth phase. For wet cell weight (WCW), pellets were measured in pre-weighed falcon tubes directly after centrifugation and removal of the excess supernatant.

3.7.2 eGFP Production

Flow cytometer (Guava easy-cyte, Milipore) was used for the determination of eGFP expression as described in Stadlmayr et al., (2010). The cell distribution was determined and the cells that show linear regression in terms of FSC-H and FSC-A values were gated for specific selection of the singlets in each population. The specific eGFP synthesis values of the cells based on the fluorescence intensity related to the cell volume (Hohenblum et al., 2003); and, in eGFP fluorescence calculations, geometric mean of the gated population was used. For each measurement, fluorescence signal from 10000 cells were taken into account.

3.8.3 rHSA Production

Human Albumin ELISA Quantitation Set (Bethyl, USA) was used for the specific quantitative detection of human albumin concentrations in cultivation supernatant. Affinity purified Human Albumin Coating Antibody was pre-coated onto a 96-well microtiter plate. Standards and samples were pipetted into the wells and any Albumin present was bound by the immobilized antibody. An enzyme-linked antibody specific for HSA (HRP Conjugated Human Albumin Detection Antibody) was added to the

wells. Finally Ultra TMB-ELISA Substrate Solution (Thermo Fisher scientific, USA)was added and color developed in proportion to the amount of HSA bound in the initial step. The color development is stopped and the intensity of the color was measured.

3.8.4 eGFP Transcript Levels

For total RNA isolation from *P. pastoris* strains TRI Reagent (Ambion) was used. DNAse treatment of isolated RNA samples were performed with DNA-free kit (Applied Biosystems) and cDNA synthesis was performed with Biozym cDNA synthesis kit according to directions of manufacturer (Biozym, Germany). Real time PCR was used for the quantification of mRNA levels of modified *P. pastoris* cells and Blue S'Green qPCR Mix (Biozym, Germany) was used according to manufacturer's instructions. For relative quantification of mRNA levels *Actin* gene of *P. pastoris* was used as a calibrator. Used primers for qPCR analysis is provided in Table 3.1.

CHAPTER 4

RESULTS AND DISCUSSION

The ultimate aim of this PhD thesis is designing and engineering novel promoter variants to obtain strong ethanol regulated promoter variant(s) for improved r-protein production, and to clarify the role of certain transcription factors (TFs) on regulation of *P. pastoris* ADH2 (P_{ADH2}) and AOX1 (P_{AOX1}) promoters. Promoter architecture of P_{ADH2} was altered via nucleosome optimization and modification of transcription factor binding sites (TFBSs); and, the regulation of P_{AOX1} was aimed to be changed from methanol to ethanol induction. Adr1 (Mxr1), Aca2, Cat8 and TATA binding motifs were used to design and construct P_{ADH2} and P_{AOX1} promoter libraries. The strength of the novel engineered promoter variants (NEPVs) was investigated initially in the intracellular synthesis of the model protein eGFP in different carbon sources . Thereafter, to demonstrate the potential of the NEPVs, extracellular human serum albumin (HSA) production under the NEPVs, denoted by $P_{ADH2\text{-}V}$ and $P_{AOX1\text{-}V}$, also was studied. In order to identify the role of the TFs on the regulation of P_{ADH2} and P_{AOX1}, overexpression and knock-out strains of Adr1 (Mxr1), Aca1, Cat8-1, and Cat8-2 were developed for the synthesis of eGFP under $P_{ADH2\text{-}V}$ and $P_{AOX1\text{-}V}$. Influences of the deletion and overexpression of the TFs on the strength of the NEPVs $P_{ADH2\text{-}V}$ and $P_{AOX1\text{-}V}$ in different carbon sources, and regulation of the TFs in response to the carbon sources using mRNA quantification, were investigated.

4.1 Construction of Recombinant Plasmids and Strains

In order to demonstrate and evaluate the impact of the novel engineered promoter variants (NEPVs), expression cassettes: ***i)*** first, for the recombinant intracellular-heterologous eGFP synthesis to eliminate the influences of the secretion process, and then, ***ii)*** for the recombinant extracellular-heterologous HSA synthesis and secretion, were constructed under the ENVPs and tested in the fermentations using different carbon sources.

The base plasmid pPICZα-A*::eGFP* (Ata et al., 2017) was used in the construction of recombinant expression systems (Sequence is given in Appendix C). For the intracellular eGFP synthesis, P_{AOX1} and α-factor signal sequence were replaced with the NEPVs. For the extracellular HSA synthesis and secretion, in the base plasmid the promoter P_{AOX1}, α-factor signal sequence, and *eGFP* gene were replaced with the NEPVs and *HSA* gene. Based on noncutter analysis of $P_{ADH1\text{-}V}$, $P_{AOX1\text{-}V}$, *eGFP*, *HSA* and pPICZα-A*::eGFP* forward and reverse primers were designed with appropriate restriction enzyme (RE) recognition sites. In order to construct NEPVs mutagenic primers were designed and modifications were introduced by two-step overlap extension PCR. $P_{ADH2\text{-}V}$ gene fragments, $P_{AOX1\text{-}V}$ gene fragments and *HSA* gene fragments were amplified as explained in Section 3.5.1. pPICZα-A*::eGFP* vector was double digested with the REs *Bgl*II and *Btsα*I or *Bgl*II and *Kpn*I, and then purified by using gel purification kit. For construction of eGFP containing plasmids, $P_{ADH2\text{-}V}$ gene fragments were double digested by *BamH*I and *Btsα*I whereas $P_{AOX1\text{-}V}$ gene fragments were double digested with *Bgl*II and *Btsα*I and purified with gel extraction kit. For hSA producing recombinant plasmid construction, the RE *Kpn*I was used instead of *Btsα*I. Purified insert and vector genes were ligated by using a ligation ratio of 3:1 and recombinant plasmids including *eGFP* as a reporter gene were transformed into *E. coli* DH5α cells, while hSA gene containing recombinant plasmids were transformed into *E. coli* DH10B cells as described in Section 3.4.8 At least 6 clones per transformant were selected and checked with colony PCR. The recombinant

plasmids carrying the designed systems under the NEPVs were isolated from the putative positive cells according to colony PCR results; and, DNA sequences were analyzed (METU Central Laboratory, Ankara; Microsynth AG, Balgach, Switzerland). For long-term storage glycerol stock of the positive clones were prepared and stored at -80°C.

The recombinant plasmids constructed under the NEPVs were isolated from the transformants and sequences were analysed (ABI Prism 310 Genetic Analyzer, Applied Biosystems, USA). The sequence authenticity was checked with NCBI (NIH, USA) nucleotide BLAST tool.

Confirmed plasmids were linearized and used for transformation of *P. pastoris* X33; and transformations were performed as described in Sections 3.4.9, and minimum ten clones were selected for each novel *P. pastoris* strain for screening. For reliable comparison of the strength of promoter variants, *P. pastoris* clones that represent the whole population were selected and eGFP gene copy numbers (GCNs) was determined by qPCR relative to the housekeeping gene *ARG4*, and the results are given in Table 4.1.

Table 4.1 *P. pastoris* strains eGFP gene copy numbers (GCNs) obtained with relative quantification to *ARG4*

***P. pastoris* Strains**	**Gene Copy Number of eGFP**		
	Run$_1$	**Run$_2$**	**Run$_3$**
wtADH #1	0.9/0.8	1.2	
wtADH #2		1.1/1.1	
wtADH #3		1.0/1.0	0.9
wtADH #4		0.9	1.0
Add_Aca #1	0.8/0.7	1.0	
Add_Aca #4			1.012
Add_Cat #5	0.7	1.3	
Add_Cat #6			1.0/1.0
Add_Cat #10			1.0/0.9
Add-Adr #1	0.9	1.2	

Table 4.1 (continued)

***P. pastoris* Strains**	**Gene Copy Number of eGFP**		
	Run_1	**Run_2**	**Run_3**
NucOpt #1	0.8	1.0	1.0
NucOpt #5			1.2
Opt_Cat #2	0.9	1.2	1.0/0.9
Opt_Cat #5			1.0/1.0
Opt_Cat #11			1.0/1.0
AddTATA #1	0.8	1.2	
AddTATA #3			1.1
OptAdr1-1 #1	0.7	1.0	
OptAdr1-2 #1	0.7	1.1	
OptAdr1-2 #4			1.0
OptAdr1-3 #1	0.7/1.0		
NucOpt-OptCat #1	0.9		
NucOpt-OptCat #13	1.1		
wtAOX #1	0.8	0.9	
wtAOX #2		1.1	
wtAOX #4		0.6	
wtAOX #5		0.9	
modAOX #6	0.8	0.6	
modAOX #8	1.1		
mAOX-Adr1 #1	1.2		
mAOX-Adr2 #1	1.2		
mAOX-Adr2 #7	1.2		
mAOX-Adr3 #1	1.3/ 1.1		

* Results given in columns *Run_1*, *Run_2*, *Run_3* obtained in different runs of experiment.

** Results given with slash were technical replicates of the same run.

4.2 Design and Performance of NEPVs of P_{ADH2}

This work on the promoter engineering of *P. pastoris* alcohol dehydrogenase 2 promoter and the results presented here are for the first time in the literature, despite *P. pastoris* alcohol dehydrogenase 2 promoter functional *cis*-acting elements and *trans*-acting factors that involve in P_{ADH2} transcriptional activation are not known yet.

The recruitment and interactions of sequence specific binding regulatory proteins known as transcription factors is a pivotal dimension of transcriptional regulation. Promoters mostly have multiple regulators. In order to design rationally NEPVs, these very critical motifs bound by specific TFs were identified by *in silico* analysis tool MatInspector. The *ADH2* promoter sequence (1047 bp upstream of the gene *PAS_chr2-1_0472*) was searched for transcription factors binding site matrix families belonging to the matrix groups ′fungi′ and ′general core promoter elements′ using the MatInspector from Genomatix (Cartharius et al., 2005). 108 putative transcription factor binding site (TFBS) belonging to 37 different matrix families with matrix similarity score higher than 0.75 were found (data is provided in Appendix D). For detailed characterization of P_{ADH2} cis-acting elements, previously characterized DNA binding sequences of specific TFs that are involved in promoter regulation by ethanol were also manually curated. Based on the regulatory modules identified, the engineered P_{ADH2} variants were designed via addition and/or optimization of Adr1, Cat8, Aca1 and TATA binding motifs. Also, nucleosome optimized P_{ADH2} was redesigned using MATLAB and NuPoP tools.

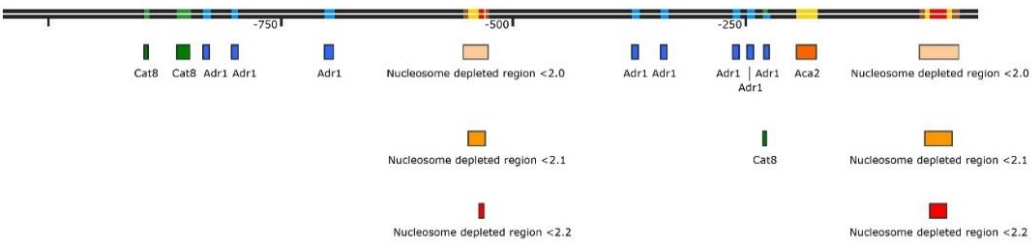


Figure 4.1 Wild-type P_{ADH2} putative TFBSs predicted as important for ethanol regulation and nucleosome-depleted regions.

4.2.1 NEPVs of P_{ADH2} - Design-1: TFBS Modifications of Adr1

S. cerevisiae Adr1, alcohol dehydrogenase 2 synthesis regulator, has a pivotal role in the activation of glucose repressible genes and essential for growth on ethanol, glycerol, and oleate; and, also on peroxisome function and biogenesis (Cook et al., 1994). *P. pastoris* Mxr1, methanol expression regulator 1, is a homologue of *S. cerevisiae* Adr1 (Lin-Cereghino et al., 2006). Adr1 binds to a palindromic sequence, a perfect 22-bp inverted repeat annotated as UAS1, positioned between -271 to -291 bp upstream of *S. cerevisiae* ADH2 translation start site (Eisen et al., 1988), and binds symmetrically as two monomers to activate P_{ADH2} (Thukral et al., 1991). Adr1 recognition motifs according to YEASTRACT database are TTGG(A/G)GN{6,38}C(T/C)CCAA, TTGG(A/G)G, (G/T)(T/N)GGGG(T/N), (G/N)(C/T)GGGG(T/N)(A/N) (Teixeira et al., 2018).

4.2.1.1 Optimization of P_{ADH2} Putative Adr1 TFBSs

In *S. cerevisiae* Adr1 recognition motif was mutated to increase the relative binding affinity of its corresponding TF Adr1, and the results are given in Table 4.2 (Cheng et al., 1994). The putative Adr1 palindromic repeat binding site mutated by Cheng et al. (1994):

1 2 3

5' C(T/C)CC(A/G)(A/C/T) --spacer(4-36bp)--(T/G/A)(T/C)GG(A/G)G*

Table 4.2 Relative binding affinity of Adr1 with respect to mutations applied to position 1, 2 and 3(Cheng et al., 1994).

	Position 1	**Position 2**	**Position 3**
T	1.0	1.0	0.7
C	0.01	0.1	0.6
A	0.1	0.01	1.0
G	0.3	0.01	0.4

Based on the knowledge about constraints on the orientation and spacing of binding motifs that affect activation *in vivo*, the preferred Adr1 binding consensus sequence TTGG(A/G)GA was found in both halves of the inverted repeat (Cheng et al., 1994). Three putative palindromic Adr1 TFBSs were identified by manual curation of *P. pastoris* P_{ADH2} using YEASTRACT database (Teixeira et al., 2018), although these sequences are not recognized as Adr1 TFBS *in silico* analysis by MatInspector. These putative Adr1 TFBSs annotated as Adr1-1, Adr1-2 and Adr1-3 which are located at bp -832 to -797, -370 to -334, and -261 to -241 positions, respectively (Figure 4.2).

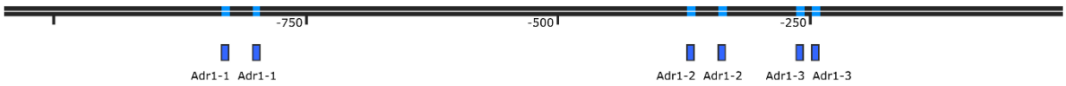


Figure 4.2 Putative Adr1 binding motifs that are optimized for construction of the NEPVs $P_{ADH2\text{-}OptAdr1\text{-}1}$, $P_{ADH2\text{-}OptAdr1\text{-}2}$, $P_{ADH2\text{-}OptAdr1\text{-}3}$

In order to optimize the putative Adr1 binding motifs for increased Adr1 affinity, proximal regions of the three palindromic repeats were mutated by designing $P_{ADH2\text{-}OptAdr1\text{-}1}$, $P_{ADH2\text{-}OptAdr1\text{-}2}$, $P_{ADH2\text{-}OptAdr1\text{-}3}$, based on the findings of Cheng et al. (1994). The mutations designed are represented, as follows:

Design-1 for TFBS optimization of Adr1:

ACCCCACATAGTGACAATGATTATGTAAGAAGAGGGGG was replaced with

ACCCCA**A**ATAGTGACAATGATTATGTAAGAA**TT**GGGG**A**

Design-2 for TFBS optimization of Adr1:

ACTCCACAAATTGTCCAATCTCGTTTTCCATTTGGGAGA was replaced with:

TCTCCA**A**AAATTGTCCAATCTCGTTTTCCATTT**T**GGAGA

Design-3 for TFBS optimization of Adr1:

C<u>CCCC</u>GCTGTTTGAAAAC<u>GGGG</u>G was replaced with:

T<u>CCCC</u>**AA**TGTTTGAAA**TT**<u>GGGG</u>**A**

*Mutations are highlighted by light-grey and presented wit bold-letters;

** Putative Adr1 binding core motifs are underlined.

In order to evaluate the changes in putative TFBSs resulting from modifications, the designed promoter variants were analyzed by MatInspector. Compared to wild-type P_{ADH2} ($P_{ADH2\text{-}wt}$), after mutating three nucleotides on $P_{ADH2\text{-}OptAdr1\text{-}1}$, three different TFBS formations and a motif for deletion were determined. Two zinc finger transcriptional repressors, namely: Mig3 (F$YMIG; matrix: F$MIG3.01), Mig1(F$YMIG; matrix: F$MIG1.02), and a transcriptional activator for genes in multistress response, Msn2 (F$YSTR; matrix: F$MSN2.01) binding motifs were determined for deletion. According to the MatInspector results, an inverted CCG repeat motif (F$ICGG; matrix F$ICGG_N10.01) was created on $P_{ADH2\text{-}OptAdr1\text{-}1}$. A heat shock transcription factor binding motif (F$YHSF; matrix: F$HSF1.01) was found on $P_{ADH2\text{-}OptAdr1\text{-}2}$ different than $P_{ADH2\text{-}wt}$. On $P_{ADH2\text{-}OptAdr1\text{-}3}$ two Mig1 binding site deletions (F$YMIG; matrixes: F$MIG1.01, F$MIG1.02), and an inverted CCG motif deletion (F$ICGG; Matrixes: F$ICGG_N10.01), and formation of a binding site for qa-1F (F$YQA1: matrix: F$QA1F.01) required for quinic acid induction of transcription in the qa gene cluster, were observed. It is indeed noteworthy that the putative TFBSs determined by the MatInspector analyses does not always reveal the actual situation; therefore, functional relevance of identified TFBSs in the context of P_{ADH2} was also considered. According to the results, all the three mutation design on Adr1 binding motif was not identified by the MatInspector, because of the different consensus sequences used in MatInspector.

Among the TFBS designs for Adr1, no one lead to a significant change in promoter activity. Under the promoter variants $P_{ADH2\text{-}OptAdr1\text{-}1}$, $P_{ADH2\text{-}OptAdr1\text{-}2}$, $P_{ADH2\text{-}OptAdr1\text{-}3}$, eGFP synthesis capacity of the *P. pastoris* strains were quite similar to that of under

$P_{ADH2\text{-}wt}$. These results showed that, in order to alter P_{ADH2} activity, these sequences are not challenging for an effective engineering strategy. Since *P. pastoris* Adr1(Mxr1) has been evolved for new functions, possibly its optimal DNA binding sequences might be evolved, as well. It should also be noted that, Adr1 binding may not be the limiting factor for activation of *P. pastoris* P_{ADH2}; thus, even these sequences provide better motifs for Adr1 binding, it could not create any change in the promoter activity unless occurrence of further contributions by the other TFs.

4.2.1.2 Integration of Additional Adr1 TFBS

Promoter regulation and strength of the NEPVs can be controlled by integration of activator and repressor binding sites which are introduced to a rationally determined position to co-operatively function with other transcriptional activation machinery. Therefore, $P_{ADH2\text{-}AddAdr}$ variant was designed with an extra activator binding sequence addition to increase the strength of the NEPV. An Adr1 binding motif integration site was determined at -462 to -441 bp positions from the start codon considering the promoter architecture constraints: ***i)*** avoiding a change on the other essential TFBSs for the activation of the P_{ADH2}, ***ii)*** determining a position for an Adr1 binding site that is in close proximity to the core promoter for a smoothly operating activation process, and ***iii)*** considering the distances between Adr1-2 and Adr1-3 TFBSs. Thus, the ENVP $P_{ADH2\text{-}AddAdr1}$ was designed according to the architecture depicted in Figure 4.3; where, original nucleotides located in determined positions were mutated as represented, as follows:

TTCCATTCGATTGCAATTCTCA was replaced with:
TCCCCAACGATTGCATTGGGGA

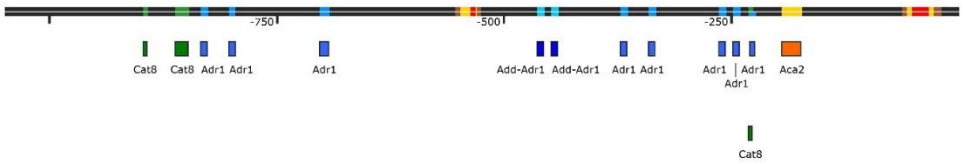


Figure 4.3 Designed promoter architecture of the NEPV $P_{ADH2\text{-}AddAdr1}$ having an additional Adr1 TFBS

$P_{ADH2\text{-}AddAdr1}$ performed similar expression levels with $P_{ADH2\text{-}wt}$. Addition of another Adr1 binding motif as an activator did not increase the promoter activity. According to MatInspector results, by the Add-Adr1 modification a high performance TFBS was not obtained; however, two heat shock TF binding motifs (F$YHSF; matrixes: F$HSF1.01, F$HSF1.02) and an inverted CGG motif (F$ICGG; matrix: F$ICGG_N10.01) were disrupted.

In *S. cerevisiae,* Adr1 is the main activator of ethanol regulated ADH2 promoter and disruption of its binding sequence severely damage the promoter activity (Beier et al., 1985; Cheng et al., 1994). When six bp core DNA region of UAS1 was disrupted, no Adr1-DNA complex was observed (Cheng et al., 1994). Under derepression conditions, deletions of *ADR1* gene lead to significant reduction in gene expression by P_{ADH2} to 19.5% of the wild-type strain (Walther and Schüller, 2001). Due to importance of Adr1 in activation of P_{ADH2} in *S. cerevisiae*, it is one of the engineering targets of *P. pastoris* P_{ADH2}; however, neither addition of an extra Adr1 TF binding palindromic repeat nor optimization of three putative Adr1 binding sites increased the promoter strength. These results reveal that, P_{ADH2} has enough *cis*-acting elements to interact with Adr1 that were evolutionarily optimized for *P. pastoris* Adr1 (Mxr1) binding. Although *P. pastoris* Adr1 (Mxr1) is a homologue of *S. cerevisiae* Adr1, the binding site optimization of which was studied by Cheng et al. (1994), their preferred Adr1 binding sequences can be different because of their evolutionary differences. Adr1 TFBSs on P_{ADH2} might have already been optimized for *P. pastoris***,** and changes

in the surrounding nucleotides of core TFBS region may not increase its strength, even may lead to a decreased affinity between the sequences and Adr1(Mxr1) molecule.

The studies in yeast *S. cerevisiae* and filamentous fungi *Aspergillus niger* showed that increase in the recombinant gene expression cassettes is not provide advantage in terms of r-protein production beyond a certain limit, since the essential TFs become limiting (Archer et al., 2006). In *P. pastoris*, Adr1(Mxr1) is constitutively expressed at low levels and crucial for peroxisome function and methanol utilization (Lin-Cereghino et al., 2006). Meanwhile, r-protein production under P_{AOX1} in *P. pastoris* with multi-copy expression cassettes were reported to be downregulated, and negatively affected the methanol consumption, the cell growth, and r-protein production in the cell; which was considered as a Mxr1 titration effect that is increase in Mxr1(Adr1) molecules in the liquid-phase in the micro-environment of Mxr1-DNA binding regions in the genome (Cámara et al., 2017). Since TFs are regulatory proteins on the cell metabolism, their concentrations need to be within a range in the cell. Conclusively it is possible to assert that, either: ***i)*** Adr1 concentration might be limited in the cell; or, ***ii)*** Adr1 might not be the single limiting factor but the other essential TFs also can have limiting effect which avoids a concerted transcriptional activation of P_{ADH2}. Consequently, addition and/or optimization of an Adr1(Mxr1) binding site on P_{ADH2} did not lead to an increase in the promoter strength.

4.2.2 NEPVs of P_{ADH2} - Design-2: TFBS Modifications of Cat8

S. cerevisiae Cat8 is a zinc cluster transcriptional activator necessary for the gluconeogenesis, glyoxylate shunt enzymes, diauxic shift from fermentation to oxidative metabolism, and ethanol utilization (Haurie et al., 2001; Rahner et al., 1999). The transcriptional regulation by Cat8 is performed by a carbon source responsive element (CSRE) located as an upstream activating site (UAS) on promoters; and, transcriptional activation by CSRE depends on an intact and

functional Cat8 TF with a N-terminal binuclear C_6 zinc cluster motif (Rahner et al., 1996). In *S. cerevisiae*, CSRE is bound by Cat8 or Sip4 TFs to activate the gluconeogenesis, glyoxylate shunt, and ethanol utilization pathway genes (Hiesinger et al., 2001). Whereas, in *P. pastoris* there are two homolog proteins namely Cat8-1 (*PAS_chr2-1_0757*) and Cat8-2 (*PAS_chr4_0540*), and both of them show high homology to both *S. cerevisiae* Cat8 and *S. cerevisiae* Sip4. Cat8 and Sip4 TFs are members of Gal4 family of activators for which interaction with symmetrically disposed CGG triplets is a mutual characteristic (Schjerling and Holmberg, 1996). Dimerization of zinc cluster proteins leads to interaction with two CGG motifs in a direct (CGG-CGG), inverted palindrome (CGG-CCG) or everted palindrome (CCG-CGG) found on the target sequence (Hellauer et al., 1996). Actually, both Cat8 and Sip4 can bind to the CSRE, even in the absence of each other; however, Cat8 and Sip4 unequally contribute to the activation of CSRE-dependent genes, under glucose derepression condition 85% and 15%, respectively (Hiesinger et al., 2001). Consensus Cat8 binding sequence for efficient gene activation was derived as YCCNYTNRKCCG, whereas more specific motif is necessary for activation by Sip4, that is TCCATTSRTCCGR (Roth et al., 2004).

S. cerevisiae P_{ADH2} is transcriptionally controlled by Adr1 that is bound to UAS1; however, even in *adr1Δ* mutants substantial level of derepression was occurred and a CSRE element as a second upstream activating site (UAS2) was identified. P_{ADH2} UAS2-dependent gene activation is specifically fulfilled by Cat8, but not with the related zinc cluster transcriptional activator Sip4 (Walther and Schüller, 2001). Under derepression conditions, deletions of *ADR1* and *CAT8* genes lead to significant reduction in gene expression by P_{ADH2} to 19.5% and 12.5% of the wild-type strain, respectively, even double mutant *adr1Δcat8Δ* lost almost all the activity; thus for the maximal activation of P_{ADH2} synergistic functioning of Adr1 and Cat8 is necessary (Walther and Schüller, 2001). In contrast to additive contribution of Adr1 and Cat8 on the derepression of *ACS1* promoter, synergistic activation of P_{ADH2} by Adr1 and Cat8 can be simply resulted from varying distances between Cat8 and Adr1 binding

sites which are 200bp for P_{ACS1} and 3bp for P_{ADH2} (Walther and Schüller, 2001). Adr1 and Cat8 may cooperatively facilitate recruitment of both transcriptional coactivators/mediators including potentially Ada2, histone acetyltransferase Gcn5, and components of core transcription factors TFIIB and TFIID for which direct interactions with transactivation domains (TADs) of Adr1 were already demonstrated (Chiang et al., 1996; Komarnitsky et al., 1998).

According to YEASTRACT database DNA binding motifs for Cat8 are ''NCCDTYNVNCCGN'', ''CCNYTNRRCCGN'' and ''tCCGG'' (Teixeira et al., 2018). In *S. cerevisiae*, depending on the activation potential of mutated Cat8 binding sequence elements, an optimal Cat8 binding motif was derived as TTCCGTTCGTCCGA (Roth et al., 2004)

In silico analysis of *P. pastoris* P_{ADH2} revealed a putative Cat8 TFBS located at -861 to -847 bp from translation start site. Throughout manual curation of TFBSs based on YEASTRACT database, two putative Cat8 binding site tCCGG were identified at bp -896 to -892 and bp -230 to -226 positions. Additionally, 4 putative Sip4 binding site and a yeast CSRE were identified by *in silico* analysis by MatInspector (Figure 4.4).

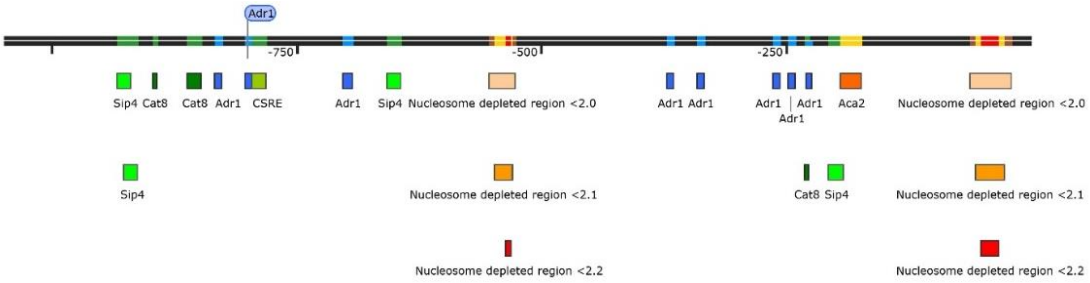


Figure 4.4 Putative Cat8, Sip4, and yeast CSRE binding sites in the P_{ADH2} architecture.

4.2.2.1 Integration of Additional Cat8 Binding Site

In order to design the NEPV denoted by $P_{ADH2\text{-}AddCat8\text{-}1}$ based on the P_{ADH2} promoter architecture, an additional optimized Cat8 binding sequence derived by Roth et al. (2004) was integrated at -401 to -388 bp positions by replacing the original nucleotide stretches (Figure 4.5). At the specified position, the original nucleotide stretch TAATTTCGTTGATT was replaced with TTCCGTTCGTCCGA. Since, Adr1 and Cat8 collaboratively activates P_{ADH2} in *S. cerevisiae*, the integration position for Cat8 TFBS was chosen 16 bp upstream in the middle of Adr1 (Adr1-2) palindromic repeat. After construction, $P_{ADH2\text{-}AddCat8\text{-}1}$ was analyzed by MatInspector. According to MatInspector results, in addition to intentionally inserted Cat8 binding site (F$CSRE; matrix: F$CAT8.01) TFBSs also were generated at the overlapping positions Gal4 (F$YGAL; matrix: F$GAL4.01), Rgt1 (F$MGCM; matrix: F$RGT1.02), and Rap1 (F$YRAP; matrix: F$RAP1.04) (Appendix E).

Amongst the overlapping TF binding sites, in *S. cerevisiae*, Gal4 is a transcriptional activator of GAL genes in response to galactose (Bhat and Murthy, 2001). However, *P. pastoris* cannot utilize galactose; thus, Gal4 should have been evolved for a different function. The second TF, Rgt1is a glucose responsive transcription factor that can act both as an activator or repressor and mainly regulates expression of many glucose transporter (Hxt) genes (Özcan et al., 1996). Lastly the third TF, Rap1 can bind to many gene loci and relocolizes to cytosol under hypoxic conditions; involved in a variety of functions including glycolysis, chromatin silencing/maintenance and telomere length maintenance (Chen et al., 2011). Rgt1 and Gal4 have binuclear zinc cluster or zinc knuckle DNA binding domains that share similar consensus binding sequences, the CCG triplets, therefore a modification targeting to insert a TFBS for a Zinc-cluster transcriptional activator Cat8 can also possibly create coinciding binding sites for other related Zinc-cluster factors. However, each TF is regulated at many steps; for example, depending on carbon sources their expression levels, localization in the cell, and their active or passive states are affected. Although, there are other

overlapping TF binding sites in Add-Cat8-1st position, under derepressed conditions, mainly Cat8 binding and its collaboration with other elements in the context of P_{ADH2} is expected.

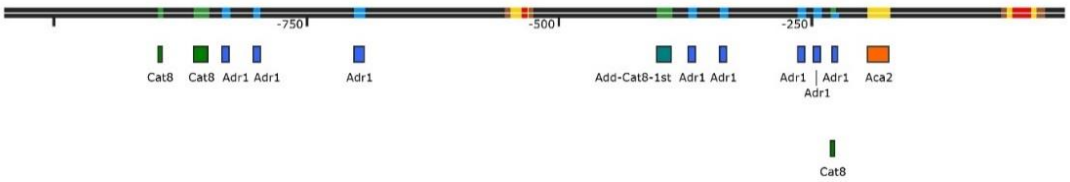


Figure 4.5 Designed promoter architecture of the NEPV $P_{ADH2\text{-}AddCat8\text{-}1}$ having an additional Cat8 TFBS

In ethanol $P_{ADH2\text{-}AddCat8\text{-}1}$ variant performed 2.36-fold higher r-protein production than that of $P_{ADH2\text{-}wt}$. The NEPV $P_{ADH2\text{-}AddCat8\text{-}1}$ increased the transcriptional potential, possibly benefited from increased synergism between Cat8-Cat8, but mainly due to Adr1-Cat8 interactions. In the literature, a synthetic promoter variant of P_{ADH2} UAS2 element, Cat8 TFBS, and basal ICL1 promoter that is devoid of its natural regulatory sequences showed increase in the expression capacity with an increase in the Cat8 TFBSs; 1 copy Cat8 TFBS provided 2-fold activation whereas 2 and 3 copy insertion of Cat8 into the synthetic promoter resulted in 80- and 312-fold activation, respectively (Walther and Schüller, 2001).

4.2.2.1 Design of TFBS in NEPVs Mimicking TFBS of Cat8

The manual curation of P_{ADH2} revealed that there are two CCG triplets within bp -309 to -296 in the nucleotide stretch: CGACCGGTGTCCGA, which are similar to interaction domains of zinc cluster protein Cat8 dimers reported by Roth et al. (2004). MatInspector could not identified the sequence as Cat8 TFBS; however, another zinc cluster protein at the overlapping position, Leu3 binding site (cgaCCGGtgtccgaaaa),

located at -309 to -292 bp to start codon was identified. The TF Leu3 regulates enzymes involved in the synthesis of branched-chain amino acids (Friden and Schimmel, 1988). Since Leu3 is not related with ethanol utilization metabolism and due to the importance of Cat8 for the activation of *S. cerevisiae* P_{ADH2} in collaboration with Adr1, this zinc cluster protein binding motif was modified for another zinc cluster protein, Cat8. Considering the promoter architecture, there are two different palindromic repeats for Adr1 binding in close proximity of this position which are 23 bp and 33 bp apart from the targeted Cat8 TFBS optimization site. Therefore, a promising clue of potential synergism between the two combinations of Adr1-Cat8 was obtained. In order to optimize the DNA fragment for an optimum Cat8 binding, the TFBS derived by Roth et al. (2004) was adopted, where the green-highlighted 7 nucleotides were mutated, as depicted below:

CGACCGGTGTCCGA sequence was replaced with
TTCCGTTCGTCCGA

In this design approach, the first CCG triplet was shifted 1 bp apart from the second CCG triplet which lengthened the spacer region from 4- to 5- nucleotides, and the nucleotides in the spacer region and upstream region of the 1st CCG triplet were changed; consequently, the designed NEPV is denoted as $P_{ADH2\text{-}AddCat8\text{-}2(OptCat)}$.

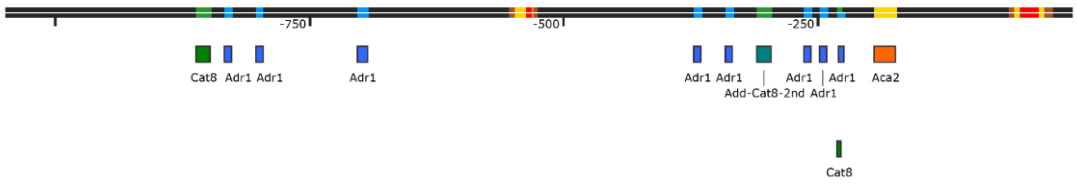


Figure 4.6 Designed promoter architecture of the NEPV $P_{ADH2\text{-}AddCat8\text{-}2(OptCat)}$ having an additional/optimized Cat8 TFBS

Among the zinc cluster proteins, the specificity of binding may be determined by the spacer length while nucleotides in the spacer is less important (Reece and Ptashne, 1993). The designed ENVP is expected also to form a DNA-binding site for the other zinc cluster proteins besides Cat8. The MatInspector analysis of the putative TFBSs of $P_{ADH2\text{-}AddCat8\text{-}2}$ compared to $P_{ADH2\text{-}wt}$ revealed that: ***i)*** TFBS of Leu3 (F$FBAS; matrix: F$LEU3.02) was replaced with TFBS of Cat8 (F$CSRE; matrix: F$CAT8.01); but, ***ii)*** at the overlapping position together with the targeted Cat8 TFBS formation, additional TFBSs for Lys14(F$MGCM; matrix: F$LYS14.01), Rgt1(F$MGCM; matrix: F$RGT1.02), Oaf1(F$YORE; matrix: F$OAF1.01) and Abf1(F$YABF; matrix: F$ABF1.03) were created (Appendix E). Optimization of the sequence for Cat8 could also create TFBSs for the other TFs such as Lys14 and Rgt1, which have Gal4- class DNA-binding motifs like Cat8. Lys14 is a transcriptional regulator activates the genes involved in the lysine biosynthesis (Feller et al., 1994). Oleate-activated transcription factor (Oaf1) act alone or as a heterodimer with Pip2, activates the genes involved in β-oxidation of fatty acids, peroxisome biogenesis and organization, and involved in diauxic shift (Karpichev et al., 1997; Zampar et al., 2013). Yeast Abf1, autonomously replicating sequence binding factor, involves in gene silencing, DNA replication and repair, and transcriptional regulation (Boer, 2000; Reed et al., 1999). Although, Add-Cat2 (Opt-Cat) modification created some other untargeted TFBSs, transcriptional regulation by TFs depend on the conditions and other elements of the promoter architecture. In ethanol, specific Cat8 binding to the modified sequence element and its collaboration with the other TFs are expected. Under the NEPV $P_{ADH2\text{-}AddCat2(OptCat)}$, 3.80- and 4.76- fold higher eGFP expression levels were obtained, respectively, in the fermentations with methanol and ethanol induction compared to wt-ADH2 promoter; which indicates the considerable increase in the promoter strength with the Add-Cat8-2 modification in the promoter design. The results indicate that in *P. pastoris*, although the concentrations of the TFs Cat8-1 and Cat8-2 are sufficient in the ethanol grown cells, the naturally occurring $P_{ADH2\text{-}wt}$ can be a Cat8 TFBS limited promoter, as its strength was improved with two Cat8

TFBS modifications which indicates that Cat8 TFBS motif is a useful tool in P_{ADH2} engineering. In the fermentations with ethanol, the NEPVs $P_{ADH2\text{-}AddCat8\text{-}1(AddCat)}$ and $P_{ADH2\text{-}AddCat8\text{-}2(OptCat)}$ performed 2.36- and 4.76-fold higher eGFP synthesis compared to that of under $P_{ADH2\text{-}wt.}$ The identical Cat8 TFBSs inserted into the NEPVs can perform diverse characteristics due to their designed positions and closeness to adjacent TFBSs and the core promoter, since their TFs perform synergistic effect or prevent binding of the other TFs due to the steric hindrance created. The results show that the strength of the NEPV having AddCat-2(OptCat) modification is higher than the other due to the design of the constructed promoter architectures. The NEPV *P. pastoris* $P_{ADH2\text{-}v}$ AddCat8-2 was designed with two proximal Adr1 binding sites that are located 23 bp upstream and 33 bp downstream of modified position while AddCat-1 modification site has only one Adr1 TFBS that is located 16 bp downstream of the modified position. Since in *S. cerevisiae* Cat8 activates P_{ADH2} in collaborative action with Adr1, closeness of inserted Cat8 binding motif to two putative Adr1 binding motifs are expected to increase the transcriptional activation potential. The increased promoter strength is the result of the synergism created between the designed TFBSs and TFs within $P_{ADH2\text{-}AddCat8\text{-}2(OptCat)}$ context. A decay in the activation effect of TFBS on transcriptional regulation is observed with the increase in distance to the start codon (Sharon et al., 2012); which is with the inserted Cat8 TFBSs in the NEPVs to the start codon are 387 bp and 295 bp in $P_{ADH2\text{-}AddCat8\text{-}1(AddCat)}$ and $P_{ADH2\text{-}AddCat8\text{-}2(OptCat)}$, respectively. Since the 2nd insertion position is 92 bp closer to the translation start site and closer to the core promoter, basal transcriptional machinery assembly makes AddCat8-2(OptCat) design stronger. Add-Cat8-2 design was achieved with 7 nucleotide replacement; and the similarity of its sequence to the original sequence is indeed noteworthy.

4.2.3 NEPVs of P_{ADH2} - Design-3: Modification of TATA-box

One of the parameters in gene expression is the RNA Pol II binding to core promoter sequence. A decisive step in eukaryotic transcriptional regulation is the recruitment of TATA-binding protein (TBP) that facilitates the bending of DNA to allow the RNA Pol II binding and transcription initiation (Struhl, 1995). Genome-wide analysis revealed only ~19 % of yeast promoters contain TATA elements, and many of them (~10% of all yeast promoters) are depend on SAGA co-activator function that are highly regulated and stress responsive (Huisinga and Pugh, 2004). In contrast, ~90% of yeast promoters are more constitutively active and primarily TFIID dependent (Huisinga and Pugh, 2004; Shen, 2003). The canonical binding sequence of *S. cerevisiae* TATA-box is TATA(A/T)A(A/T)(A/G) (Basehoar et al., 2004). Although only ~19% of *S. cerevisiae* genes contain a conventional TATA- box, TATA-less promoters still need TBP to function (Pugh and Tjian, 1991). Transcriptional activation mechanism differs between TATA and TATA-less promoters: In TATA harboring promoters TBP is delivered by the SAGA complex; whereas in TATA-less promoters TBP is delivered by TFIID (Basehoar et al., 2004). Alterations in the sequence of TATA-box act on the interactions between RNA Pol II and DNA; consequently, gene expression and promoter noise (Blake et al., 2006; Raser, 2004; Tirosh et al., 2006). Investigations on TATA-box canonical sequences showed mutations on the TATA-box affects, its strength, and stronger gene expression increases in a predictable way with the strength of the TATA box independent from the upstream arrangement of *cis*-acting regulatory sites (Blake et al., 2006; Mogno et al., 2010). The evidence shows that the TATA-box as well as other TFBSs are modular components of the promoters. TATA-box sequence does not directly affect TBP binding to promoter but rather the stability of TBP and TFIIA complex that is essential for TBP recruitment to transcriptional machinery (Stewart and Stargell, 2001) and the TBP turnover rate, which is defined as the rate of new TBP molecule's binding to the promoter after an old one has been dislocated (van Werven et al., 2009).

A number of promoter specific factors are important in intervening rates of transition between active and inactive promoter states. The TBP is one of the key component of those, since it is highly conserved among all the eukaryotes and can be essential for transcription of all the yeast genes. Binding of TBP to promoter DNA is often a rate-limiting step in the assembly of transcription apparatus and subsequent transcription initiation (Kim and Iyer, 2004). Stably formed transcription scaffold supported by TBP can lead to transcriptional reinitiation, a process associated with strong promoters which efficiently express genes at high levels (Blake et al., 2006). Transcriptional reinitiation process is observed when a subset of transcriptional scaffold containing TFs and cofactors remains bound to the core promoter after the clearance of RNA Pol II which leads to the elimination of *de novo* reassembly of a complete transcriptional machinery and thus resulted in the generation of larger amount of mRNA molecules in shorter periods (Yudkovsky et al., 2000).

P. pastoris $P_{ADH2\text{-}wt}$ promoter nucleotide sequence is:

‘’TCCTTTTTACCACCCAAGTGCGAGTGAAACACCCCATGGCTGCTCTCCG
ATTGCCCCTCTACAGGCATAAGGGTGTGACTTTGTGGGCTTGAATTTTAC
ACCCCCTCCAACTTTTCTCGCATCAATTGATCCTGTTACCAATATTGCAT
GCCCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATCGCA
GGGTGAGACTGTAGAGACCCCACATAGTGACAATGATTATGTAAGAAGA
GGGGGGTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTGAAC
AATGCCCAGCAGTCCTCCCCACTCTTTGACAAATCAGTATCACCGATTAA
CACCCCAAATCTTATTCTCAACGGTCCCTCATCCTTGCACCCCTCTTTGG
ACAAATGGCAGTTAGCATTGGTGCACTGACTGACTGCCCAACCTTAAAC
CCAAATTTCTTAGAAGGGGCCCATCTAGTTAGCGAGGGGTGAAAAATTC
CTCCATCGGAGATGTATTGACCGTAAGTTGCTGCTTAAAAAAAATCAGT
TCAGATAGCGAGACTTTTTTGATTTCGCAACGGGAGTGCCTGTTCCATTC
GATTGCAATTCTCACCCCTTCTGCCCAGTCCTGCCAATTGCCCATGAATC
TGCTAATTTCGTTGATTCCCACCCCCCTTTCCAACTCCACAAATTGTCCA
ATCTCGTTTTCCATTTGGGAGAATCTGCATGTCGACTACATAAAGCGACC
GGTGTCCGAAAAGATCTGTGTAGTTTTCAACATTTTGTGCTCCCCCCGCT
GTTTGAAAACGGGGGTGAGCGCTCTCCGGGGTGCGAATTCGTGCCCAAT
TCCTTTCACCCTGCCTATTGTAGACGTCAACCCGCATCTGGTGC**GAATAT**
AGCGCACCCCCAATGATCACACCAACAATTGGTCCACCCCTCCCCAATC
TCTAATATTCACAATTCACCTCAC**TATAAATA**CCCCTGTCCTGCTCCCAA
ATTCTTTTTTCCTTCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTAC
GAAA’’

*Putative TATA-box elements are given with bold characters, in the NEPV $P_{ADH2\text{-}TATA}$ grey highlighted nucleotides were replaced with the TATA-box motif of *P. pastoris* P_{AOX1} for construction of $P_{ADH2\text{-}TATA}$.

P. pastoris P_{ADH2} is a highly regulated promoter, based on our screening results it is strongly induced by ethanol and to a lesser extent with methanol and repressed by excess glucose, limited glucose and excess glycerol. Depending on its intrinsic regulated nature, a TATA binding domain is expected within its promoter architecture. However, according to the results of *in silico* analysis by MatInspector, any TATA-box motif was not identified in the core promoter region. As represented above in the core promoter element of $P_{ADH2\text{-}wt}$, two TATA-box resembling sequences located between -75 to -82 bp and -155 to -161 bp were observed with manual curation of promoter sequence. It is well established that strength of TATA-box domain affects

the affinity of RNA Pol II to promoter sequence and it acts as a modular scaling factor regardless of the other TFBSs located upstream sequence of the promoter; thus, strong TATA element increase the expression strength of promoters (Mogno et al., 2010). Additionally, nucleotide differences on TATA-box sequences results in differences in the strength of TATA binding domain (Blake et al., 2006). In this context, introducing a strong canonical TATA-box to a rationally designed position on regulated P_{ADH2} was hypothesized as a novel promoter engineering target. For the rational design of NEPV $P_{ADH2\text{-}TATA}$, promoter architectures of strong methanol and ethanol regulated *P. pastoris* P_{AOX1} and *S. cerevisiae* P_{ADH2} promoters are considered.

Table 4.3 Distance between TATA-box domain to translation start site and regulatory important Adr1 (Mxr1) TFBS

Promoter Name	**Distance Between TATA-start codon**	**Distance Between TATA-Adr1 (Mxr1) TFBS**
***P. pastoris* P_{AOX1}**	152 bp	79bp
***S. cerevisiae* P_{ADH2}**	151 bp	102 bp
NEPV $P_{ADH2\text{-}TATA}$	152 bp	70 bp

Table 4.4 *P. pastoris* P_{AOX1} TATA element structure and sequence obtained with MatInspector

	Strand	**Core similarity**	**Matrix Similarity**	**Sequence**
Yeast TATA	+	1	0.943	acagcaaTATAtaaaca
Yeast TATA	-	1	0.852	ttctgttTATAtattgc

Based on the position of TATA-box elements to start codon and to major transcription regulator Adr1 (Mxr1), a TATA-box *cis*-acting sequence integration site was designed as shown in Table 4.3 to create $P_{ADH2\text{-}TATA}$. Distance between TATA-box and start codon on *P. pastoris* P_{AOX1} and *S. cerevisiae* P_{ADH2} promoters are 152 bp and

151 bp, respectively (Table 4.3); and, the distance between TATA-box and Adr1(Mxr1) TFBSs are 79 bp and 102 bp, respectively. By considering the known strong promoters architecture, a canonical *P. pastoris* TATA-box integration site was determined at the site of distal TATA-resembling sequence located between -155 to -161 bp to start codon; and, TATA-box distance to start codon and Adr1(Mxr1) TFBS was designed as 152 bp and 70 bp, respectively. In order to use canonical TATA binding motif of *P. pastoris*, overlapping TATA-binding sequences that are aligned on positive and negative strands (presented in Table 4.4) similar to P_{AOX1} TATA-box architecture were integrated to the designed position.

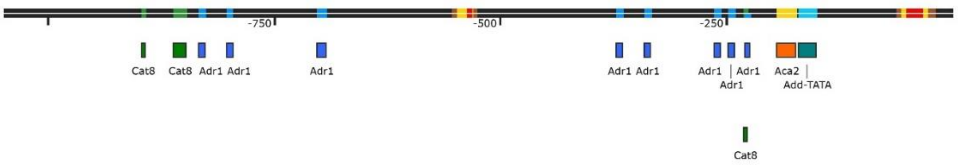


Figure 4.7 Designed promoter architecture of the NEPV $P_{ADH2\text{-}AddTATA}$

The NEPV $P_{ADH2\text{-}AddTATA}$ was analyzed with MatInspector. The design with two yeast TATA-box motifs (O$YTBP; matrix: O$SPT15.01) located at plus and minus strands were also determined by MatInspector; however, two plant TATA box motifs (O$PTBP; matrixes O$PTATA.01, O$PTATA.02) and a vertebrate TATA box element (O$VTBP; matrix: O$VTATA.01) were also identified. Furthermore, according to the MatInspector data, fork head transcription factor Fkh1 binding site (F$FKHD; matrix: F$FKH1.01) was also formed, and iron-sensing transcription factor Aft1 binding site (F$IRTF; matrix: F$AFT1.01) was destructed.

The design strategy with TATA-box integration resulted in a reduced promoter strength of ca. 70% of the $P_{ADH2\text{-}wt}$ in fermentations with ethanol induction. The core promoter region, stretching ca. 200 nucleotide long before the start codon on which the basal transcription machinery assembles, with its sequence has significant impact

on the promoter strength. The results revealed that addition of a TATA element to P_{ADH2} sequence was not appropriate for increasing the strength of the promoter. Indeed interesting, the TATA-box domain was not identified *in silico* analysis, probably because $P_{ADH2\text{-}wt}$ contains a TATA element due to its intrinsic nature, which is probably located at a position within -75 bp to -82 bp and its distal sequence might be different than the canonical yeast core elements. Although regulated promoters usually have TATA-box and preference to utilize SAGA rather than TFIID coactivator, some yeast regulated promoters can still contain a TATA element and depend on TFIID-coactivator system. In *S. cerevisiae*, a physical interaction between transcription activation domains (TADs) of Adr1 and TFIID complex was found, and it was stated that transcriptional activation by Adr1 needs an intact TFIID complex (Komarnitsky et al., 1998). If that was the case, integration of a second TATA binding sequence could change the bending pattern of the promoter for transcriptional activation and lead to confusion in the assembly of transcriptional machinery, which reduces mRNA production rate. Moreover, it is also possible that some regulated promoters can be TATA-less and depend on TFIID complex that can have degenerate or divergent sequences, so TFIID binding motif can not be recognized. If this was the case, integration of TATA-box element to TATA-less promoter may recruit also SAGA coactivator complex for transcriptional activation which can lead to a competition between SAGA and TFIID co-activator complexes, a disordered transcriptional assembly, and weaker expression levels.

4.2.4 NEPVs of P_{ADH2} - Design-4: TFBS Modifications of Aca2

S. cerevisiae TFs Aca1 and Aca2 are ATF/CREB family basic leucine zipper transcription factors, and arose from whole genome duplication and bind to ATF/CREB consensus motif TGACGTCA. *P. pastoris* has only Aca1 (*PAS_chr3_0135*), does not have paralog Aca2 (Cst6). *S. cerevisiae* Aca2 is an important TF for utilization of non-optimal carbon sources, and Aca2 null mutants exhibited growth defect on ethanol and became more vulnerable to ethanol and

oxidative stresses (Liu et al., 2016). In *P. pastoris*, there is no report about the regulatory role of Aca1, yet. *In silico* analysis of *P. pastoris* P_{ADH2} revealed a putative Aca2 TFBS located at -193 to -173 bp. As Aca2 is essential for ethanol utilization in *S. cerevisiae*, effect of an additional Aca2 TFBS on P_{ADH2} was investigated. The NEPV $P_{ADH2\text{-}AddAca2}$ was designed by integration of an additional Aca2 binding motif from -293 to -273 bp by replacement of the original sequence (Figure 4.8). Since the influence of TFBSs and their respective TFs weakens as their distance form core promoter increases, Add-Aca2 integration site was chosen in close proximity to core promoter without disturbing any essential TFBS for P_{ADH2} activity.

Figure 4.8 Designed promoter architecture of the NEPV $P_{ADH2\text{-}AddAca2}$ having an additional Aca2 TFBS.

Putative TFBSs of $P_{ADH2\text{-}AddAca2}$ was analyzed by MatInspector and compared to $P_{ADH2\text{-}wt.}$ GATA type zinc finger protein Gat4 (F$GATC; matrix F$GAT4.01) located at + and – strands at overlapping position was lost, and targeted Aca2 (F$BZIP; matrix: F$CST6.01) was inserted. Three different unintended TFBSs were formed which belong to families of M-box interacting with Mat1-Mc (F$MMAT; matrix: F$MAT1MC.01), yeast mating factors (F$YMAT; matrix: F$MATALPHA2.02), and Telobox-containing general regulatory factor (F$TGRF; matrix: F$TBF1.02) (Appendix E). *In silico* analysis of promoter sequences by MatInspector revealed the putative TFBSs, even minor changes in nucleotide sequence sometimes gives quite few differences in the TFBSs. As MatInspector uses an algorithm to identify DNA

binding motifs; however, the determined putative TFBS motifs cannot always correct, as they also could be interrelated with function of the promoter, as well. From the functional point of view, $P_{ADH2\text{-}AddAca2}$ has an additional Aca2 binding motif located in close proximity to the core promoter region and 79 bp apart from the native Aca2 binding domain.

The addition of an Aca2 binding motif did not change the strength and regulation of the NEPV. Contrary to its paralog identified in *S. cerevisiae*, in *P. pastoris* Aca1 may not be necessary for the ethanol utilization metabolism. Due to methylotrophic nature of *P. pastoris,* its transcription factors should have been evolved to control different regulatory pathways. Since Aca2 is a stress responsive TF and functioning in non-optimal carbon source utilization, in *P. pastoris* it can be active in methanol utilization pathway and peroxisome biogenesis. Indeed, promoter and transcriptional engineering results of P_{AOX1} supported this idea which is presented in the upcoming section 4.6.

4.2.5 NEPVs of P_{ADH2} - Design-5: Nucleosome Modifications of P_{ADH2}

Chromatin is a pivotal constituent of gene regulation since it exerts considerable influence on TF binding, transcription initiation and elongation (Li et al., 2007). The importance of chromatin structure on the promoter strength (Lam et al., 2008) as well as the potential to alter transcription rates by inserting nucleosome disfavoring DNA sequences (Sharon et al., 2012), were reported. Inherent properties of poly (dA:dT) tracts are crucial for nucleosome depletion, promoter accessibility, and transcriptional activity (Raveh-Sadka et al., 2012; Segal and Widom, 2009). The genome-wide organization of nucleosome positioning is determined by combination of DNA sequence, nucleosome remodeling enzymes and transcription factors. These determinants influence each other; therefore, nucleosome positioning patterns are probably differing among genes and cells in a population with eventual effect on gene expression (Struhl and Segal, 2013).

Evaluation of a TEF-promoter library showed that cumulative sum of predicted nucleosome affinity across the entire promoter is inversely proportional to promoter strength in a predictable and robust way, in spite of the vast amount of differences in sequence and TFBS mutations (Curran et al., 2014). Nucleosome architecture influences activities of yeast promoters and identified strong correlation underpins the potential for nucleosome architecture to be used as a promoter engineering tool for *P. pastoris* promoters. Nucleosome architecture optimization of native *S. cerevisiae* promoters resulted in up to 16-fold increase in their expression capacity (Curran et al., 2014).

The investigation of $P_{ADH2\text{-}wt}$ nucleosome occupancy by Nucleosome Positioning Prediction (NuPoP) tool (Wang et al., 2008; Xi et al., 2010) revealed two nucleosome depleted regions located at -552 to -527 and -60 to-19 (Figure 4.1). To increase the strength of P_{ADH2} via minimization of the blocking effect of nucleosome, original nucleosome structure was minimized using computational tools for the redesign of $P_{ADH2\text{-}NucOpt}$. Computational redesign of nucleosome optimized P_{ADH2} was performed using MATLAB algorithms that employ NuPoP for nucleosome affinity prediction (Curran et al., 2014). Throughout the optimization, nucleotides outside of the identified TFBSs and other restricted sites were systematically perturbed via custom MATLAB scripts that employ NuPoP; and a greedy algorithm was utilized to minimize the cumulative affinity score over several cycles of optimization. In the redesign process the core promoter, yeast stress responsive elements, carbon source responsive elements, TFBSs related with non-optimal carbon source utilization, and originally nucleosome depleted regions found on P_{ADH2} were protected from any mutation. To observe specifically the nucleosome modification effect, the method was also restricted from creation or destruction of any known transcription factor binding sites according to YEASTRACT database (Teixeira et al., 2018). For more reliable nucleosome positioning prediction, in addition to promoter sequence, 200bp upstream and 100bp downstream (the first 100 nucleotides of *eGFP*) of P_{ADH2} was provided to the software as an input data. The redesigned $P_{ADH2\text{-}NucOpt}$ nucleotide sequence was

determined at the 100th round of the computational cycle aiming optimization of the NEPV.

>P*ADH2-NucOpt*

TCCTTTTTACCACC**T**AAGTGCGAGTGAAACACCC**T**ATGGCTGCTCTCCGA
TTGCCCCTCTACAGGCATAAGGGTGTGA**T**TTT**T**T**TTTT**TTT**T**AATTTTACA
CCCCCTCCAACTTTT**T**TCGC**G**T**A**AATTGATCCTGTTACCAATATTGCATG
CCCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATCGCAG
GGT**A**A**A**AT**A**TA**T**AGACCCCACA**A**AA**AAA**AA**A**AATGATTATGTAAGAAGA
GGGGGGTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTGAA**A**
AATGCCCAGC**TT**T**TT**TCCC**T**A**T**TCTTTGACAAATCAGTATCAC**TT**ATTAA
CACCCCAAAT**T**TT**T**TTCTCAACGGTCCCTCATCCTTGCACCCCTCTTTGG
ACAAATGGCAGTTAG**T**ATT**A**GTGCACTGACTGACTGCC**T**AACCTTAAAC
CC**T**AATTTCTTAGAAGGGGCCCAT**A**TAGTTAGCGAGGGGTGAAAAATTC
CTCCATCGGAGATGTATT**A**ACCGTAA**T**TT**TT**T**TT**TTAAAAAAAAA**AA**A**A**T
TCAGATAGCGA**A**A**T**TTTTTTGATTTCGC**G**ACG**C**G**C**GT**TTT**T**T**TT**TTT**TT**T**
TTTT**TTTT**TTCTCACCCCTTCTGCCCAGT**T**CTGCCAATTGCCCATGAATCT
ACTAATTTCGTTGATTCCCACCCCCCTTTCCAACTCCA**A**AAATTT**T**T**TT**AA
T**T**T**TT**TTTTT**TTT**TTTGGGAGAATCTG**A**ATGT**AT**A**T**TACATAAAGCGACCG
GTGTCCGAAAA**A**AT**T**T**T**T**T**TTTTT**T**AA**TT**TTTTT**T**T**TT**T**T**CCCCCGCT**T**T
TT**A**AAAACGGGGGT**A**AGCGCTCTCCGGGGTGCGAATTCG**C**GCCC**T**ATTC
CTTTCACCCTGCCTATTGTAGACGTCAACCCGCATCTGGTGCGAATATAG
CGCACCCCCAATGATCACACCAACAATTGGTCCACCCCTCCCCAATCTCT
AATATTCACAATTCACCTCACTATAAATACCCCTGTCCTGCTCCCAAATT
CTTTTTTCCTTCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTACGA
AA

*Green highlighted sequences show restricted sites from any mutation. Mutated nucleotides are underlined and showed with bold characters.

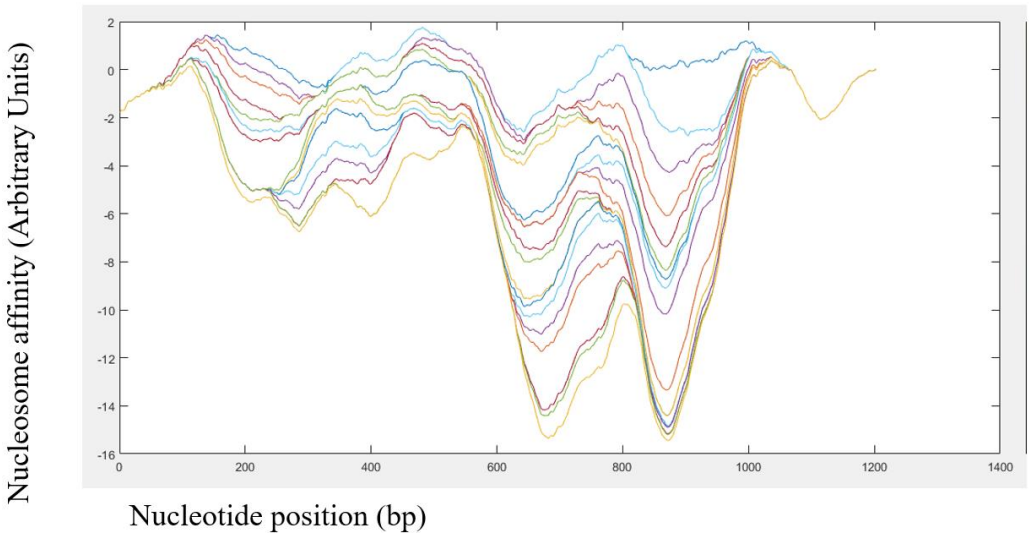


Figure 4.9 Predicted nucleosome affinity profile of the NEPVs throughout nucleosome redesign from $P_{ADH2\text{-}wt}$ to $P_{ADH2\text{-}NucOpt}$. Graphs were drawn in every five-cycle.

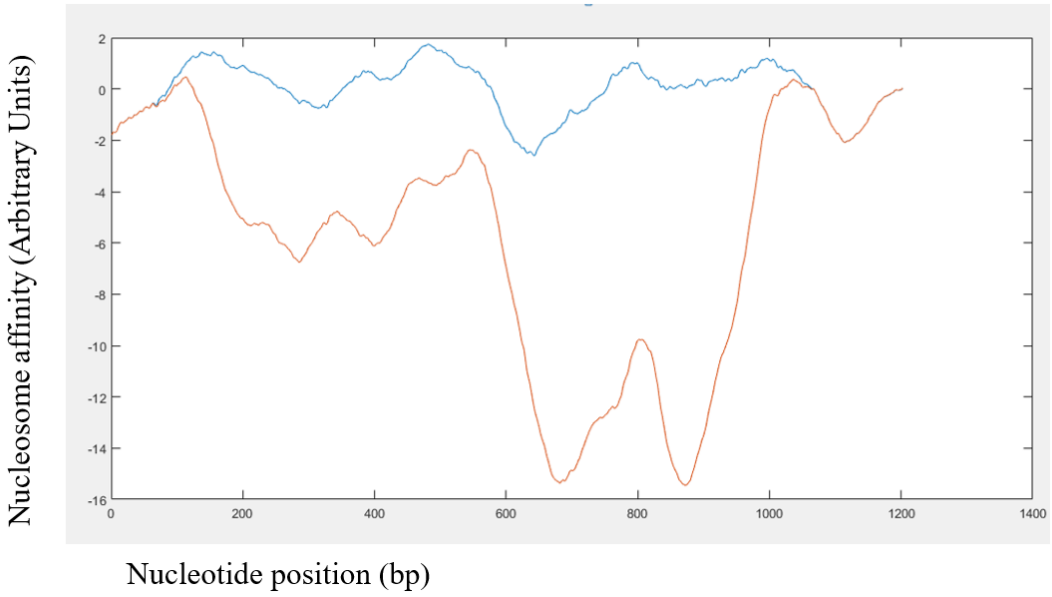


Figure 4.10 Predicted nucleosome affinity profile for the $P_{ADH2\text{-}wt}$ and $P_{ADH2\text{-}NucOpt}$ promoters calculated using the hidden Markov model

Putative TFBSs of redesigned $P_{ADH2\text{-}NucOpt}$ was analyzed by MatInspector and compared with $P_{ADH2\text{-}wt}$ (Appendix E). Although formation or destruction of known TFBSs were prevented by providing YEASTRACT TFBSs list, some deletions and insertions were identified by *in silico* analysis of MatInspector. This situation is probably resulted from the differences in TFBS motifs used in MatInspector and YEASTRACT databases. As described in the construction of the former $P_{ADH2\text{-}v}$ NEPVs, targeted TFBS modifications were not always determined by MatInspector, e.g., addition of an extra Adr1 TFBS was not identified by MatInspector. Therefore, the identified binding site differences of promoter variants need to be functionally evaluated. Many core promoter elements, such as yeast TATA element and vertebrate TATA element, also were identified in upstream sequence of $P_{ADH2\text{-}NucOpt}$; however, the core promoter region was restricted from any mutation and the identified elements on the upstream activation sequence did not exert any effect on the promoter function.

In order to evaluate the redesign for nucleosome optimization, the NEPV $P_{ADH2\text{-}NucOpt}$ was screened in deep well plates with 5 different carbon sources to test for induction and repression properties in comparison to wild-type promoter. In the fermentations with the carbon sources ethanol and methanol, $P_{ADH2\text{-}NucOpt}$ upregulated the expression, respectively, 1.5- and 3.2-fold than that was expressed under $P_{ADH2\text{-}wt}$. However, under limited and excess glucose and excess glycerol conditions, although slightly higher expression values than $P_{ADH2\text{-}wt}$ were determined, $P_{ADH2\text{-}NucOpt}$ is repressed to an extent by these carbon sources. Although MatInspector identified some destructed or newly introduced TFBSs, they did not influence the regulation pattern of $P_{ADH2\text{-}NucOpt}$.

The expression capacities of *S. cerevisiae* native promoters *TDH3* and *GAL1* were improved by Blazeck et al. (2012) by chimeric hybrid promoter engineering; but nucleosome optimization did not lead to any improvement (Curran et al., 2014). As the strength of P_{ADH2}, which is one of the strongest regulated promoter of *P. pastoris*, was enhanced by the computational redesign of the nucleosome, the work presented

in this section conclusively demonstrates that nucleosome is a platform for developing strong NEPVs of *P. pastoris*.

Meanwhile Yang et al. (2018) reported on modification of nucleosome positioning of P_{AOX1} by lengthening or removing of native poly (dA:dT) tracts, created promoter variants whose activities spanning at the range of 0.25- to 3.5-fold of wild-type promoter. Poly (dA:dT) stretches stimulate gene expression by conferring increased accessibility to proximal TFBSs (Iyer and Struhl, 1995; Raveh-Sadka et al., 2012). When poly (dA:dT) stretches were mutated, strong reduction in promoter activities were reported (Sharon et al., 2012). Contrary to the previous reports, Yang et al. (2018) reported that increase in predicted nucleosome affinity as a result of deletion of poly (dA:dT) stretches led to higher promoter activities. Poly (dA:dT) tracts possess shorter helical structure and a narrow minor groove that resist the bending necessary for histone binding (Jansen and Verstrepen, 2011). Intrinsic barrier function of poly (dA:dT) tracts lead highly positioned nucleosome formation at their adjacent locations and accordingly direct the position of neighboring nucleosomes (Mavrich et al., 2008). The study of Yang et al. (2018) aimed to relocate the nucleosome structures within the promoter by extending or deleting poly (dA:dT) tracts other than minimization of nucleosome architecture throughout the entire promoter sequence. Further in Yang et al. (2018), there is not a straight forward logic for addition or deletion of poly (dA:dT) tracts and the effect of these modifications seem location dependent in the context of promoter.

4.2.6 NEPVs of P_{ADH2} - Design-6: Combinatorial Modification of P_{ADH2}

Nucleosome optimization and Add-Cat2(Opt-Cat) design improved the transcriptional capacity of the NEPVs of P_{ADH2}, separately. In order to investigate the combinatorial effect of the two design, the NEPV $P_{ADH2\text{-}NucOpt\text{-}OptCat}$ was designed. Putative TFBSs of $P_{ADH2\text{-}NucOpt\text{-}OptCat}$ were searched by MatInspector and the results compared to that of the $P_{ADH2\text{-}NucOpt}$. The differences in TFBSs are same with the Add-

Cat2(Opt-Cat) design applied to $P_{ADH2\text{-}wt}$. Leu3 (F$FBAS; matrix: F$LEU3.02) binding motif was deleted and the targeted Cat8 (F$CSRE; matrix: F$CAT8.01) binding motif was inserted and three new TFBSs were observed. The first two belong to F$MGCM matrix family: Lys14 (F$MGCM; matrix: F$LYS14.01) and Rgt1 (F$MGCM; matrix: F$RGT1.02), and the third one is oleate activated transcription factor Oaf1(F$YORE; matrixes: F$OAF1.01) (Appendix E).

Under the NEPV $P_{ADH2\text{-}NucOpt\text{-}OptCat}$ 3.68- and 8.8-fold higher eGFP expression levels were achieved compared to the recombinant system under $P_{ADH2\text{-}wt}$ at ethanol and methanol induction conditions, respectively. Furthermore, in the fermentations with the substrates at limited glucose and excess glycerol, the strength and derepression capacity of the NEPV $P_{ADH2\text{-}NucOpt\text{-}OptCat}$ compared to $P_{ADH2\text{-}wt}$ also was increased. With ethanol induction under $P_{ADH2\text{-}AddCat2}$ and $P_{ADH2\text{-}NucOpt}$, respectively, 4.76- and 1.5-fold higher expressions were obtained; however, the combinatorial effect of the two under the NEPV $P_{ADH2\text{-}NucOpt\text{-}OptCat}$ produced an expression value between the above two values, which is higher than that of the expression level obtained under $P_{ADH2\text{-}NucOpt}$. Although the strength of $P_{ADH2\text{-}NucOpt}$ was improved with Add-Cat2 modification, a concerted effect was not observed under the NEPV $P_{ADH2\text{-}NucOpt\text{-}OptCat}$ in the fermentation with ethanol induction. Therefore, this can be the result of: ***i)*** some TFBSs could be destructed during the nucleosome optimization process, ***ii)*** the TFBSs could be occupied by some uncharacterized TFs whose function and DNA binding motifs were not determined yet, and lastly ***iii)*** those TFs could activate some genes in cooperation with Cat8, at ethanol induction conditions.

With methanol induction, however, the combinatorial design of the NEPV $P_{ADH2\text{-}NucOpt\text{-}OptCat}$ created a synergistic effect and enhanced the expression further than that of under the NEPVs designed with its components. The formation of new and yet unknown TFBSs, and their corresponding TFs may be active in the methanol grown cells and activated the genes in coordination with Cat8. Besides, TFBS that can be functional in the methanol grown cells may be highly occupied by nucleosome,

therefore their increased availability can contribute to the increased co-activation with Cat8. These results reveal that combination of promoter engineering tools does not always give a concerted effect (Hartner et al., 2008); but, our results also reveal that TFBS modification and nucleosome optimization can be combined to increase the strength of yeast promoters.

4.2.7 Expression Potential of NEPVs of $P_{ADH2:}$ Influence of Carbon Sources

The expression capacity of the engineered P_{ADH2} library was evaluated in five growth media *i.e*. ethanol, methanol, limited glucose, excess glucose, and excess glycerol. After transformation of *P. pastoris*, least ten clones for each construct were analyzed for statistical comparison and the clones representing the whole population were selected for further analysis. For the measurement of the eGFP synthesized in the cells, using flow cytometry the cell distribution was determined and the cells that show linear regression in terms of FSC-H and FSC-A values were gated for specific selection of the singlets in each population. The specific eGFP synthesis values of the cells based on the fluorescence intensity related to the cell volume (Hohenblum et al., 2003); and, in eGFP fluorescence calculations, geometric mean of the gated population was used.

Final screening experiments were performed with three biological replicas of each strain in deep 24-deep well plate at 25°C, 225 rpm for 20 h in production media and results are given in Figure 4.11-4.12 and Table 4.5-4.6. Precultivation was performed in YP media without glucose, after 20 h of precultivation, the cells were harvested by centrifugation and inoculated into each fermentation medium at specified initial cell concentration. Promoter variants did not affect the cell growth, and the novel strains constructed with the NEPVs reached the same cell concentration under the same induction condition; however, their eGFP production levels showed different patterns, as expected. In order to determine the optimal concentrations of ethanol and methanol

to induce the NEPVs, different alcohol concentrations ranging between 0.5% and 5% (v/v) were used and the results are presented in Figure 4.11 and Table 4.5.

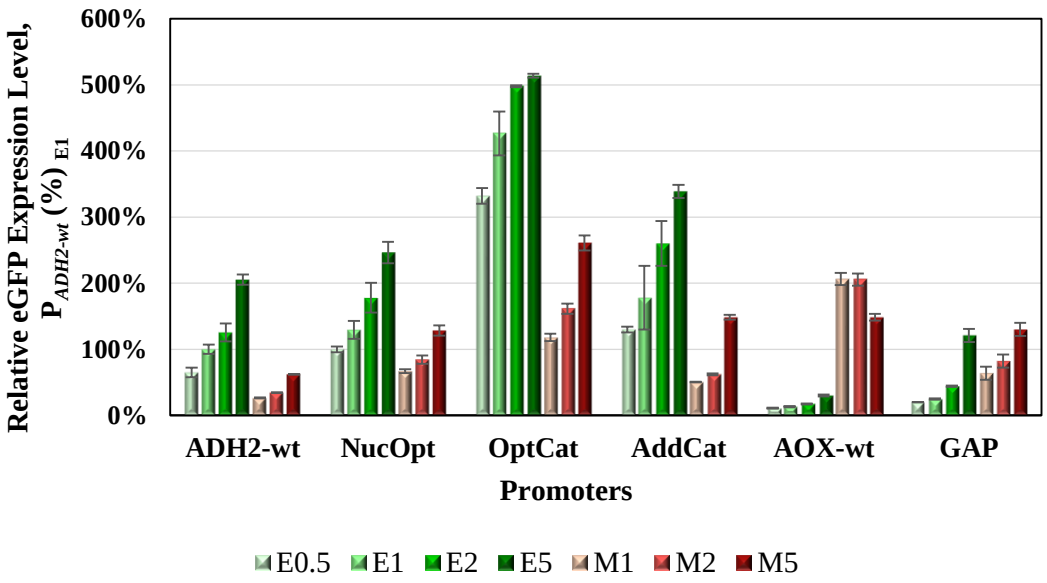


Figure 4.11 Relative eGFP synthesis capacity of constructed novel *P. pastoris* cells under $P_{ADH2\text{-}v}$ control relative to wild-type P_{ADH2} (%)$_{E1}$ at different ethanol and methanol concentrations at the cultivation time of t = 20 h. Calculations were carried out based on arbitrary fluorescence units. Mean values of 3 different clones are presented. Detailed information is presented in Table 4.5. E0.5: 0.5% (v/v) ethanol; E1: 1% (v/v) ethanol; E2: 2% (v/v) ethanol; E5: 5% (v/v) ethanol; M1: 1% (v/v) methanol; M2: 2% (v/v) methanol; M5: 5% (v/v) methanol. Error bars represent the standard deviation (±)

Based on the screening results, eGFP expression level of $P_{ADH2\text{-}wt}$ in response to 5% (v/v) ethanol was equal to eGFP levels produced by P_{AOX1} in response to 1% or 2% (v/v) methanol. Moreover, the NEPV $P_{ADH2\text{-}OptCat}$ outcompeted P_{AOX1} at all tested ethanol concentrations and $P_{ADH2\text{-}AddCat}$ did so under 2% to 5% (v/v) ethanol induction conditions.

As represented in Figure 4.11 and Table 4.5, increase in the ethanol concentration from 0.5% to 5% (v/v) led to an increase in the eGFP synthesis of $P_{ADH2\text{-}v}$. Increased ethanol concentrations created a higher increase in the eGFP synthesis levels in $P_{ADH2\text{-}wt}$, $P_{ADH2\text{-}NucOpt}$ and $P_{ADH2\text{-}AddCat}$ compared to the strong NEPV $P_{ADH2\text{-}OptCat}$. This result shows that expression capacity of the strongest NEPV $P_{ADH2\text{-}OptCat}$ reached to a plateau due to allocations of TFs to other sites in the entire genome which may lead to limited concentrations of Cat8 or other TFs required for the activation of the promoter; thus, the product formation rate decreases. $P_{ADH2\text{-}OptCat}$ also could be saturated with the selective TFs; therefore, further increase in the inducer concentration could not linearly increased the product synthesis. In order to evaluate the productivity of $P_{ADH2\text{-}v}$, a distinguishing distribution pattern was observed with 2% (v/v) ethanol and this concentration was used in the screening experiments.

The eGFP synthesis levels with methanol induction showed different patterns, where by using 1% (v/v) and 2% (v/v) methanol concentrations the same amount of eGFP synthesis under P_{AOX1} were determined; whereas, when methanol concentration was increased to 5% (v/v), ca. 25% reduction was observed in the eGFP synthesis. This phenomenon can be explained with Mxr1 (Adr1) titration effect. The screening results reveal that increasing the concentration of methanol can lead to allocation of Mxr1(Adr1) to different peroxisomal and MUT pathway genes; thus, a reduction in the activation of P_{AOX1} and the growth on methanol occurred due to limited availability of activator TF concentration. Whereas, activity under the P_{GAP} was increased with the increase in ethanol and methanol concentrations, probably due to its requirement for transcriptional activation. Based on the screening results, 1% (v/v) methanol concentration was used in the screening experiments.

Table 4.5 Relative eGFP synthesis capacity of constructed novel *P. pastoris* cells under $P_{ADH2\text{-}v}$ control relative to wild-type P_{ADH2} $(\%)_{E1}$ at different ethanol (C_2H_5OH) and methanol(CH_3OH) concentrations at the cultivation time of t = 20 h. Calculations were carried out based on arbitrary fluorescence units. Mean values of 3 different clones are presented.

Normalized eGFP concentrations in novel *Pichia pastoris* strains constructed with							
Promoter	**0.5%** (v/v) **C_2H_5OH**	**1%** (v/v) **C_2H_5OH**	**2%** (v/v) **C_2H_5OH**	**5%** (v/v) **C_2H_5OH**	**1%** (v/v) **CH_3OH**	**2%** (v/v) **CH_3OH**	**5%** (v/v) **CH_3OH**
ADH2-wt	65±7	100±7	126±13	206±8	27±1	35±1	62±0
NucOpt	100+4	130±13	178±23	247±16	67±3	85±6	129±8
OptCat	332±12	427±33	498±1	514±3	118±6	162±8	261±11
AddCat	130±4	178±48	260±34	339±10	51±1	62±1	149±4
AOX-wt	11±1	13±1	18±1	30±1	207±9	206±9	149±5
GAP	20±0	25±1	44±1	121±10	64±10	82±10	130±10

Among the tested conditions, ethanol or methanol concentrations are not limiting, and higher alcohol concentrations drives the cells into stress conditions which influences the cell metabolism in a negative way. Although the cells still continue to grow with increasing alcohol concentrations, the growth rate is reduced with increasing alcohol concentrations under the tested 24-deep well plate screening conditions (Table 4.6).

Table 4.6 OD_{600} values of *P. pastoris* X33 and the recombinant *P. pastoris* constructed with $P_{ADH2\text{-}V}$ using different initial ethanol and methanol concentrations (v/v) harvested at t = 20 h

The cell concentrations in terms of OD_{600} values							
Strains	**0.5%** (v/v) **C_2H_5OH**	**1%** (v/v) **C_2H_5OH**	**2%** (v/v) **C_2H_5OH**	**5%** (v/v) **C_2H_5OH**	**1%** (v/v) **CH_3OH**	**2%** (v/v) **CH_3OH**	**5%** (v/v) **CH_3OH**
P.pastoris X33	8.2	7.2	4.8	2.8	4.6	3.6	2.6
P.pastoris **$P_{ADH2\text{-}V}$**	7.5±0.3	6±0.7	4.2±0.1	2.2±0.1	3.8±0.2	2.8±0.3	1.8±0.2

In order to evaluate the regulation and expression strength of NEPVs, *P. pastoris* cells containing $P_{ADH2\text{-}v}$ with reporter eGFP protein gene were screened with fermentations using five different carbon sources, *i.e.* E2: 2% (v/v) ethanol, M1: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol. eGFP expression levels were calculated relative to $P_{ADH2\text{-}wt}$ productivity under E2 condition. The results are given in Figure 4.12 and Table 4.7.

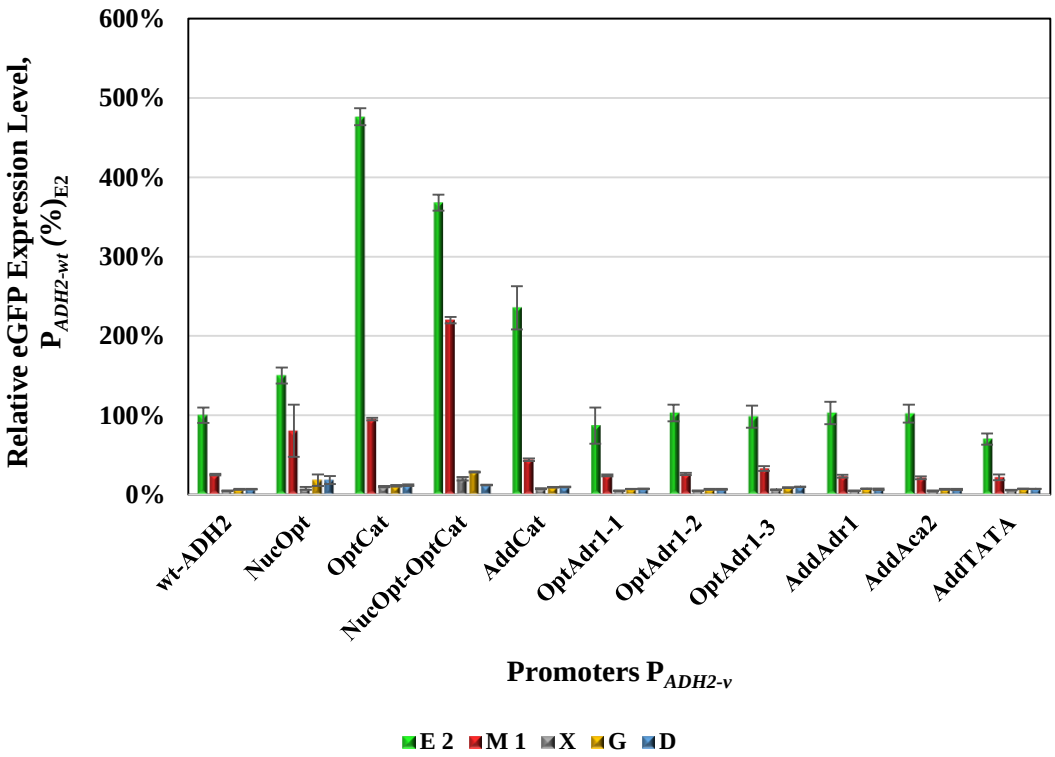


Figure 4.12 Normalized eGFP synthesis capacity of constructed novel *P. pastoris* strains with the NEPVs relative to $P_{ADH2\text{-}wt}$ (%)$_{E2}$ in different carbon sources at the cultivation time of t = 20 h. Calculations were carried out based on arbitrary fluorescence units. Mean values of 3 different clones are presented. Detailed information including statistical analysis is presented in Table 4.7. E2: 2% (v/v) ethanol, M1: 1% (v/v) methanol, X: limited glucose, G: excess glycerol, D: excess glucose. Error bars represent the standard deviation (±)

The screening results indicate that, all the NEPVs of P_{ADH2} sustain the basic P_{ADH2} regulation pattern to an extent, e.g., still repressed by excess glucose, limited glucose, and excess glycerol; and activated with ethanol. Derepression effect was more obvious in combinatorial promoter variant $P_{ADH2\text{-}NucOpt\text{-}OptCat}$ in response to limited glucose and excess glycerol conditions. The most influential effect on the strength of P_{ADH2} was observed with Cat8 modifications. In *P. pastoris*, there is not any study for characterization of Cat8/Sip4 genes, yet; however, in *S. cerevisiae*, Cat8 is an activator of P_{ADH2} together with Adr1. These results reveal that *P. pastoris* P_{ADH2} is a Cat8-limited promoter as its expression strength is increasing with introduction of new Cat8 TFBSs. Nucleosome optimization method also contributed to increase the promoter activity ca. 50% with ethanol induction, and even more in the methanol grown cells. The combinatorial effect of OptCat design together with nucleosome optimization did not show concerted effect with ethanol induction but a synergistic effect was created with methanol. Based on eGFP synthesis results, by the promoter engineering a P_{ADH2} library was designed and constructed which allows an 476% increase in r-protein synthesis compared to the that of under $P_{ADH2\text{-}wt}$.

Table 4.7 Normalized eGFP synthesis capacity of constructed novel *P. pastoris* strains with the NEPVs relative to $P_{ADH2\text{-}wt}$ $(\%)_{E2}$ in different carbon sources at the cultivation time of t = 20 h. Mean values of 3 different clones are presented. Significance of differences to the control strain was calculated with Student's t-test. (* p-values <0.05; ** p-values <0.01).

Normalized eGFP concentrations in r-*P. pastoris* constructed with					
P_{ADH2} Variant	**2% (v/v) C_2H_5OH**	**1% (v/v) CH_3OH**	**Limited Glucose**	**Excess Glycerol**	**Excess Glucose**
ADH2-wt	100±10	25±1	5±1	6±1	7±0
NucOpt	150±10*	80±33*	8±2	18±7	18±5
OptCat	476±11**	95±2**	10±1**	11±1**	12±1**
NucOpt-OptCat	368±10**	220±4**	20±2**	28±1**	12±0**
AddCat	236±27**	44±2**	7±1	9±1**	10±0**
OptAdr1-1	87±23	24±1	5±0	7±0	7±0
OptAdr1-2	103±10	26±1	5±0	7±0	7±0
OptAdr1-3	98±14	33±3*	6±0	9±1**	10±0**
AddAdr1	103±14	23±2	5±0	7±0	7±1
AddAca2	102±11	21±2	5±1	7±1	6±1
AddTATA	70±7**	22±4	6±0	7±0	7±0

4.2.8 eGFP Transcript Levels of $P_{ADH2\text{-}V}$

To confirm the enhancement in the strength of P_{ADH2} promoter variants, the transcript levels of *eGFP* in *P. pastoris* variants were analyzed by means of qPCR experiments. Comparative Ct method (ΔΔCt) was used for the relative quantification. mRNA expressions of *eGFP* were normalized relative to *Actin* mRNA expression; thereafter, divided by normalized *eGFP* expression performed with $P_{ADH2\text{-}wt}$, and mRNA transcription ratios (mTRs) were determined and presented in Figure 4.13 and Table 4.8.

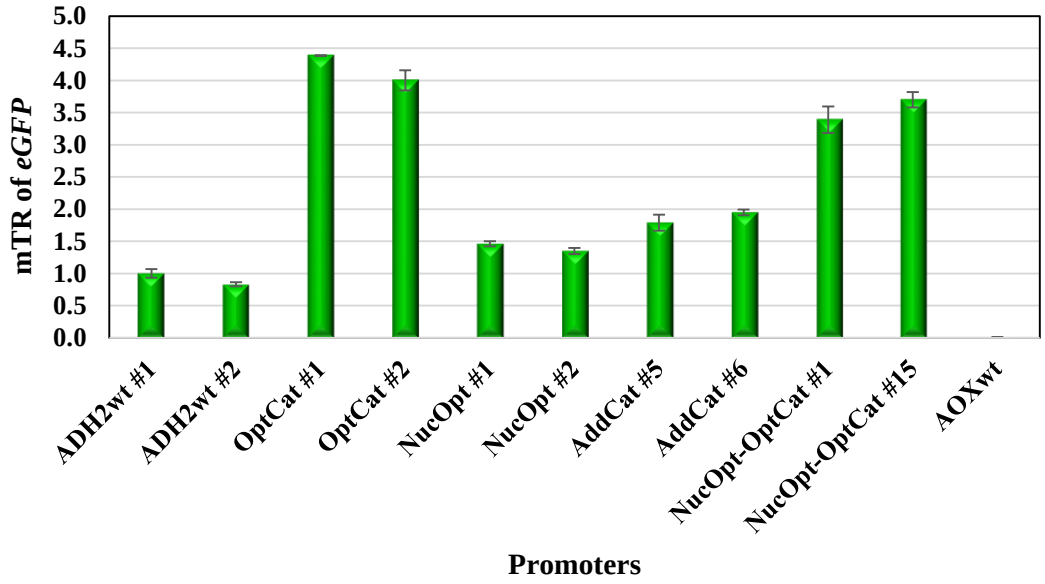


Figure 4. 13 mRNA transcription ratios (mTRs) of *eGFP* in 2% (v) ethanol grown *P. pastoris* cells under $P_{ADH2\text{-}v}$ control at t = 20 h. Transcript levels were normalized to *ACT1* mRNA expression and presented relative to expressions with $P_{ADH2\text{-}wt}$. The error bars represent the standard deviation of three technical replicates.

Table 4.8 mRNA transcription ratios (mTRs) of *eGFP* in 2% (v) ethanol grown *P. pastoris* cells with $P_{ADH2\text{-}v}$ control at t = 20 h. Transcript levels were normalized to *ACT1* mRNA expression and presented relative to $P_{ADH2\text{-}wt}$. StDev: standard deviation of three technical replicates.

Promoter	mTR of *eGFP*	StDev
ADH2wt #1	1.0	0.1
ADH2wt #2	0.8	0.0
OptCat #1	4.4	0.0
OptCat #2	4.0	0.2
NucOpt #1	1.5	0.0
NucOpt #2	1.4	0.0
AddCat #5	1.8	0.1
AddCat #6	2.0	0.0
NucOpt-OptCat #1	3.4	0.2
NucOpt-OptCat #15	3.7	0.1
AOXwt	0.0	0.0

Subsequent measurements of the mTRs confirmed stronger transcriptional capacity of the NEPVs $P_{ADH2\text{-}OptCat}$, $P_{ADH2\text{-}NucOpt}$, $P_{ADH2\text{-}AddCat}$, and $P_{ADH2\text{-}NucOpt\text{-}OptCat}$ being ca. 4.2-, 1.5-, 1.9-, and 3.6-fold higher transcriptional activity of $P_{ADH2\text{-}wt}$, respectively.

4.2.9 Extracellular Human Serum Albumin Production by $P_{ADH2\text{-}V}$

The strength of the two prominent NEPVs $P_{ADH2\text{-}NucOpt}$ and $P_{ADH2\text{-}OptCat}$ were tested in the production of extracellular human serum albumin (HSA). With its native signal sequence, *HSA* gene was cloned under the control of the NEPVs $P_{ADH2\text{-}NucOpt}$, and $P_{ADH2\text{-}OptCat}$. To be independent from the effects of genomic integration sites, the designed recombinant plasmids were integrated into AOX1 transcription termination locus. Least ten clones per each construct were screened to generate reliable results. Final screening experiments were performed with four clones for each strain representing the whole population. The extracellular HSA production was performed in 24-deep well plates at 25 °C, 225 rpm for 49 hours. The precultivation step was performed in YP media at 25 °C, 225 rpm for 20 h; and then the cells were harvested by centrifugation. The fermentations were initiated at OD_{600} = 1 in ASMv6 minimal medium. At t = 0 1% (v/v) ethanol was added into each bioreactor; thereafter, at t = 5, t = 17, t = 29, and t = 41 h, 1% (v/v) ethanol was also added to induce r-protein production, and the cells were harvested at t = 49 h.

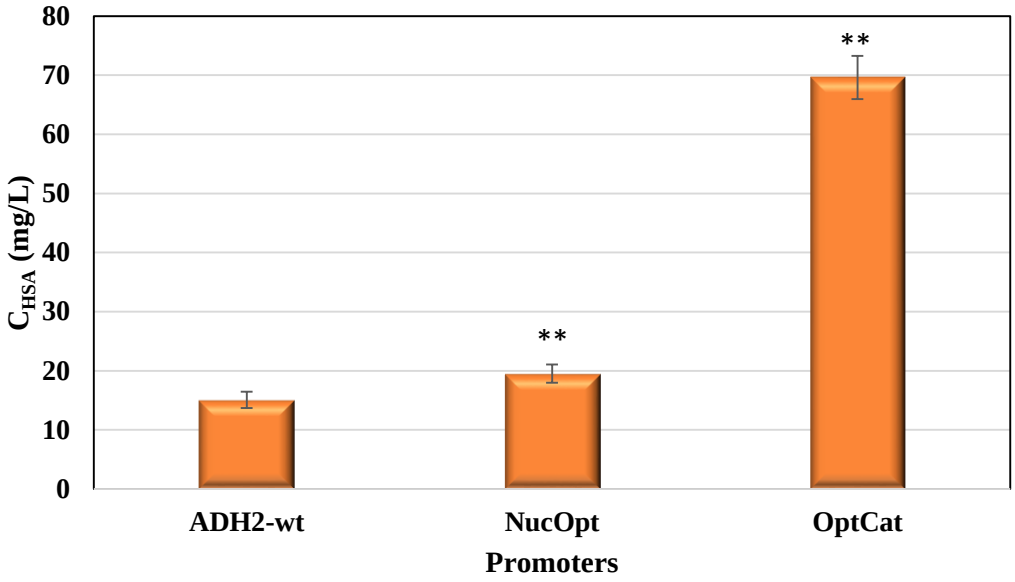


Figure 4.14 Extracellular human serum albumin (HSA) concentrations of *P. pastoris* cells with different $P_{ADH2\text{-}v}$ in ethanol pulse feed fermentation at t = 49 h. Error bars represent the standard deviation of four different clones (±). Significance of the differences to the $P_{ADH2\text{-}wt}$ was calculated with Student's t-test (** p-values <0.01).

Table 4.9 Extracellular human serum albumin (HSA) concentrations of *P. pastoris* cells under control of different P_{ADH2} variants in ethanol pulse feed fermentation at t = 49 h. StDev: Standard deviation among four different clones. Significance of differences to the $P_{ADH2\text{-}wt}$ was calculated with Student's t-test (** p-values <0.01).

P_{ADH2} Variant	**C_{HSA} (mg/L)**	**StDev**
ADH2-wt	15.1	1.4
NucOpt	19.5**	1.5
OptCat	69.6**	3.7

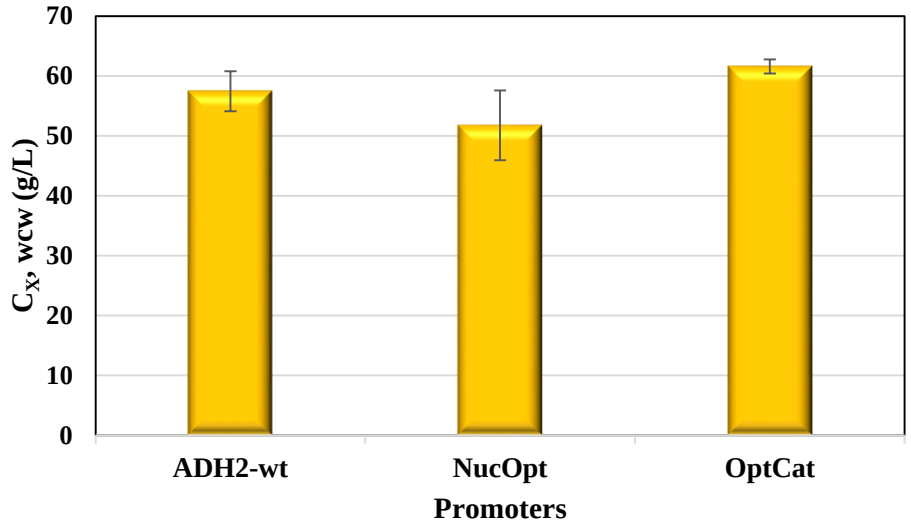


Figure 4.15 The cell concentrations of *P. pastoris* strains producing extracellular HSA under control of the NEPVs at t = 49 h of ethanol pulse feed fermentation. Error bars represent the standard deviation among four different clones (±) C_X, wet cell weight (wcw, g/L)

Table 4.10 The cell concentrations of *P. pastoris* strains producing HSA under the two prominent NEPVs at t = 49 h of ethanol pulse feed fermentation. StDev: Standard deviation among four different clones.

Promoter	Cx, wcw (g/L)	StDev
ADH2-wt	57.5	3.3
NucOpt	51.8	5.8
OptCat	61.6	1.2

The average extracellular HSA yield per gram wet-cell-weight produced by r-*P. pastoris* strains constructed with $P_{ADH2\text{-}wt}$, $P_{ADH2\text{-}NucOpt}$, and $P_{ADH2\text{-}OptCat}$ were calculated as 0.263 mg/g, 0.377 mg/g, and 1.130 mg/g, respectively. Thus, *P. pastoris* strains constructed with the NEPVs $P_{ADH2\text{-}NucOpt}$ and $P_{ADH2\text{-}OptCat}$ enhanced the hSA production 1.43- and 4.30-fold than that of the $P_{ADH2\text{-}wt}$.

Expression potentials of the strong NEPVs of the naturally occurring P_{ADH2} were confirmed by the intracellular eGFP synthesis and eGFP transcript levels, and then by extracellular HSA production. Based on the results, the NEPVs $P_{ADH2\text{-}OptCat}$, $P_{ADH2\text{-}AddCat}$, $P_{ADH2\text{-}NucOpt}$ and $P_{ADH2\text{-}NucOpt\text{-}OptCat}$ are on a par with the strength of commonly used strongest inducible promoters of *P. pastoris*. These NEPVs are promising candidates for industrial r-protein production.

4.3 Design and Performances of NEPVs of P_{AOX1}

The aim of *P. pastoris* alcohol oxidase 1 promoter engineering is developing methanol-free expression systems. Regulatory models on the *S. cerevisiae* ADH2 and *P. pastoris* ADH2 were used as a basis in engineering of P_{AOX1}. Considering the knowledge about ethanol regulated promoters, P_{AOX1} architecture was engineered for its inducible regulation by ethanol which is a repressor of the naturally occurring P_{AOX1}.

P. pastoris Mxr1 (methanol expression regulator 1) was identified as the global regulator of genes involved in the methanol utilization pathway (Lin-Cereghino et al., 2006). *P. pastoris mxr1Δ* mutants cannot grow on methanol and form functional peroxisomes; furthermore, production of methanol utilization pathway enzymes are considerably decreased (Lin-Cereghino et al., 2006). Six different Mxr1 binding sites were identified on P_{AOX1} using electrophoretic mobility shift assay (EMSA) and DNAse I footprinting methods; and, the regions were denoted by P_{AOXI} Mxr1 response elements (Mxres) (Figure.4.16) (Kranthi et al., 2009); where, 5′CYCC3′ (5′GGRG3′ in the complementary strand) motif was found to be conserved in all the six P_{AOX1} Mxres (Kranthi et al., 2009). According to *in silico* analysis by MatInspector, only one Adr1 binding site was found to be located at bp -577 to -569 from translation start site, which is overlapping with Mxre4.

***P. pastoris* wild-type P_{AOX1} ($P_{AOX1\text{-}wt}$) nucleotide sequence.**

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCC
GACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAGGAGG
GGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTT
CTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGC
TTGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCA
TGACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTG
TTTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCA
GATGAGGGCTTTCTGAGTGTGGGGGTCAAATAGTTTCATGTTCCCCAAAT
GGCCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAA
AGCGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAA
CGGCCAGTTGGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTC
TTGTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTA
GCGCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGC
AAATGGGGGAAACACCCGCTTTTTGGATGATTATGCATTGTCTCCACATT
GTATGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCAT
GATCAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATAAACA
GAAGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATTAGCTT
ACTTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGA
CTTTTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAAC
G

*Mxr1 response elements (Mxre) are highlighted in blue; and, core regions of Mxre are underlined and represented with bold characters.

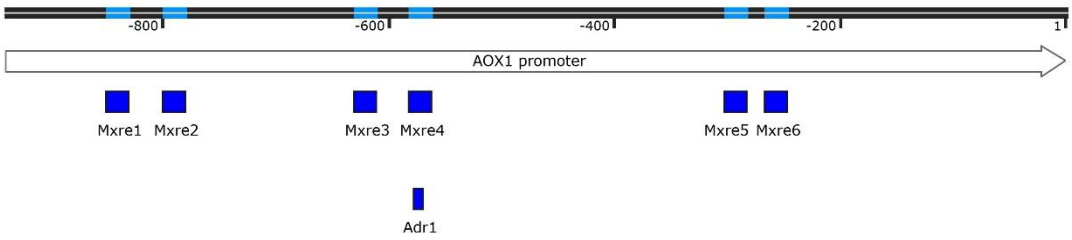


Figure 4.16 Mxr1(Adr1) binding motifs (annotated as Mxre) identified on wild-type P_{AOX1}

In *P. pastoris*, Mxr1 is the only identified homolog of *S. cerevisiae* Adr1 (Lin-Cereghino et al., 2006). Therefore, we hypothesized that *P. pastoris* Adr1(Mxr1) is a transcriptional activator of both ethanol regulated P_{ADH2} and methanol regulated P_{AOX1}

genes. Therefore, in the design of the NEPVs of P_{AOX1}, denoted by $P_{AOX1\text{-}v}$, *S. cerevisiae* Adr1, Cat8 DNA binding motifs, and *P. pastoris* Aca2 DNA binding motifs were used.

4.3.1 NEPVs of P_{AOX1}-Design-1: TFBS Modification of Adr1 (Mxr1)

Two Adr1 monomers bind symmetrically to a palindromic sequence on *S. cerevisiae* P_{ADH2} to activate transcription (Thukral et al., 1991). When Adr1 binding palindromic repeats were analyzed in different promoters, longer spacer length is observed in *P. pastoris* P_{AOX1}. Adr1 binding core motifs are separated with a spacer length of 12 bp in *S. cerevisiae* P_{ADH2}; however, the spacer length is 24 bp, 25 bp, and 9 bp, respectively, in *P. pastoris* P_{ADH2} Adr1-1, Adr1-2, and Adr1-3 palindromic repeats (Figure 4.2). *P. pastoris* P_{AOX1} core sequences are separated with a spacer length of: ***i)*** 44 bp in Mxre1 and Mxre2, ***ii)*** 47 bp in Mxre3 and Mxre4, and ***iii)*** 32 bp in Mxr5 and Mxr6. The dissociation constant (K_d) values for the Mxr1(Adr1) binding onto six Mxre motifs in P_{AOX1} are high and in the range of 50-100nM (Kranthi et al., 2009), whereas the reported K_d values for ADH2UAS1-Adr1 interactions in *S. cerevisiae* are lower and within the range of 2-9 nM (Thukral et al., 1991), which indicates that the chemisorption of Mxr1(Adr1) onto Mxre motifs occurs with a lower affinity than the chemisorption of Adr1 onto Adr1 binding motif of *S. cerevisiae* (UAS1). Therefore, we can hypothesize that decreasing the length of the spacer region can contribute to a higher binding affinity of Adr1(Mxr1) onto the corresponding binding sites, which can contribute to the activation of P_{AOX1}. In this context, three Mxre sites were replaced by the optimized Adr1 palindromic binding sequences of *S. cerevisiae* (Cheng et al., 1994). The redesign studies resulted in duplicated Adr1 binding core DNA regions in Mxre2, Mxre 4 and Mxre6 positions, with shorter spacer lengths ranging between 10-15 nucleotides. The NEPVs $P_{AOX1\text{-}Adr1}$, $P_{AOX1\text{-}Adr2}$ and $P_{AOX1\text{-}Adr3}$ were designed according to the architecture depicted in Figure 4.17; where, original nucleotides located in determined positions were engineered, as follows:

Design of $P_{AOX1\text{-}Adr1}$ for Optimization of Mxre2:

Mxre2 was replaced by the optimized *S. cerevisiae* Adr1 binding palindromic repeats with a spacer length of 10 bp, as follows:

TCCTCTT**CTCC**TCAACACCC sequence was replaced with:

A**CCCC**AATATTATTT**GGGG**T

Design of $P_{AOX1\text{-}Adr2}$ for Optimization of Mxre4:

Mxre4 was replaced by two Adr1 binding sequences identified on *P. pastoris* P_{ADH2} by MatInspector with a spacer length of 15 bp, as follows:

TTTCTGAGTGT**GGGG**TCAAATAGTTT sequence was replaced with:

GA**CCCC**ACATTTTTTTTTTGA**CCCC**A

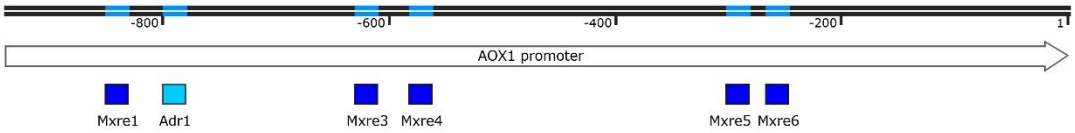
Design of $P_{AOX1\text{-}Adr3}$ for Optimization of Mxre6:

Mxre6 was replaced by the *S. cerevisiae* Adr1 binding palindromic repeat, with a spacer length of 10 bp, as follows:

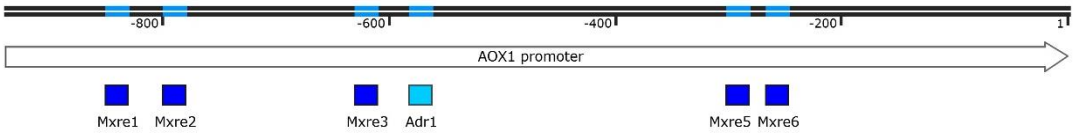
TGCATTGT**CTCC**ACATTGTA sequence was replaced with:

A**CCCC**AATACATTTT**GGGG**T

a) $P_{AOX1\text{-}Adr1}$



b) $P_{AOX1\text{-}Adr2}$



c) $P_{AOX1\text{-}Adr3}$

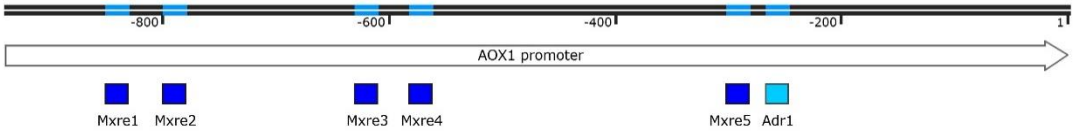


Figure 4. 17 Design of the promoter architectures of the NEPVs: **a)** $P_{AOX1\text{-}Adr1}$, **b)** $P_{AOX1\text{-}Adr2}$ and **c)** $P_{AOX1\text{-}Adr3}$

Putative TFBSs of the designed promoter variants $P_{AOX1\text{-}Adr1}$, $P_{AOX1\text{-}Adr2}$ and $P_{AOX1\text{-}Adr3}$ were analyzed by MatInspector. *In silico* analysis revealed that the redesign for $P_{AOX1\text{-}Adr1}$ did not lead to a destruction in any known TFBSs; but, created five novel TFBSs on $P_{AOX1\text{-}Adr1}$ compared to $P_{AOX1\text{-}wt}$. The novel TFBSs are: ***i)*** two targeted Adr1 binding motifs (F$YADR; matrix: F$ADR1.01); ***ii)*** two Mig1 binding motifs (F$YMIG; matrix: F$MIG1.01); and, ***iii)*** an iron-regulated transcriptional activator, Aft2 binding motif (F$IRTF; matrix: F$AFT2.01). Mig1 and Mig2 proteins are yeast metabolic regulator zinc finger proteins that mediate glucose repression. *P. pastoris* Mig1 and Mig2 are repressors of P_{AOX1}, especially responsible for glycerol repression (Shi et al., 2018). MatInspector analysis revealed only one Mig2 binding motif on the naturally occurring P_{AOX1} overlapping with the Mxre4. According to MatInspector analysis in $P_{AOX\text{-}Adr2}$ design, replacement of the Mxre4 destroyed a Mig2 binding motif (F$YMIG; matrix: F$MIG2.01), while two Mig1 (F$YMIG; matrix: F$MIG1.01) and

one Adr1 (F$YADR; matrix: F$ADR1.01) binding motifs were created. Two Adr1 binding motifs were inserted as designed; however, since Mig1 and Mig2 binding motif core regions (CCCC) share identical core sequence with the Adr1 binding site (GGGG), their identification at the overlapping position may be either due to an *in silico* lapse by the analysis tool or simultaneous formation of the Mig1/2 binding motifs. In $P_{AOX1\text{-}Adr3}$ design, a Phd1, transcription factor involved in the regulation of filamentous growth, (F$PHD1; matrix: F$PHD1.01) was deleted while two Adr1 (F$YADR; Matrixes: F$ADR1.01), two Tbf1 (TTAGGG repeat binding factor 1) (F$TGRF; matrixes: F$TBF1.01), and an inverted CGG motif (F$ICGG; matrixes: F$ICGG_N10.01) were formed according to MatInspector. Contrary to *in silico* analysis results of $P_{ADH2\text{-}v}$ on which no Adr1 TFBS was identified, in the $P_{AOX1\text{-}v}$ all the inserted Adr1 palindromic repeats were recognized even though original Mxre motifs were not identified as Adr1 binding sequences.

Expression capacity and regulation of the NEPVs $P_{AOX1\text{-}Adr1}$, $P_{AOX1\text{-}Adr2}$ and $P_{AOX1\text{-}Adr3}$ were evaluated with fermentations using five different carbon sources. The results show that all the designed variants were repressed by glucose, glycerol, and ethanol; furthermore, their activation rate in response to methanol was similar to $P_{AOX1\text{-}wt}$. The results conclusively demonstrate that the promoter engineering approaches targeted to Adr1(Mxr1) binding site did not lead to significant changes in the transcriptional activity of P_{AOX1}, similar to P_{ADH2} engineering results. Thus, it is possible to assert that either: ***i)*** Adr1 concentration can be limited in the cell; or, ***ii)*** Adr1 cannot be the single limiting factor, but the other essential TFs also can have a limiting effect which avoids a concerted transcriptional activation of P_{AOX1}. Consequently, addition and optimization of an Adr1(Mxr1) binding site on P_{AOX1}, neither led to an increase in the promoter strength nor changed the regulation of P_{AOX1}.

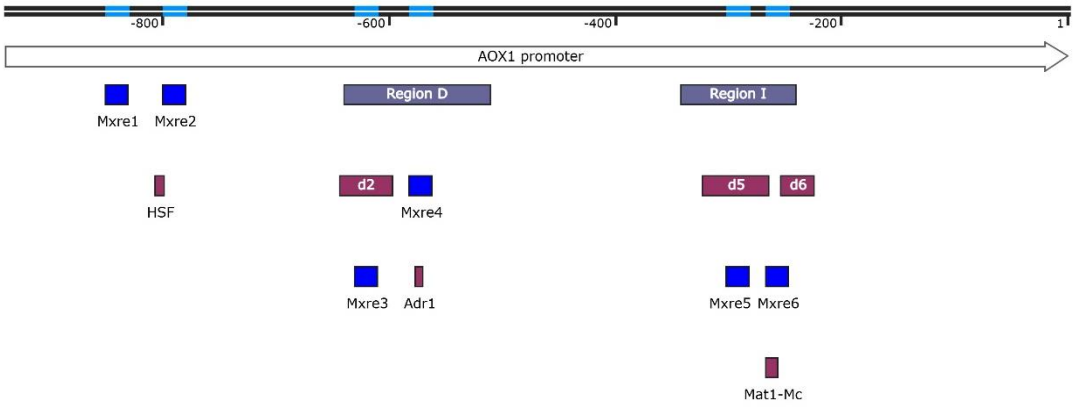


Figure 4.18 Identified overlapping DNA regions with Mxre motifs important for transcriptional activation of P_{AOX1}

Functional analysis of *P. pastoris* P_{AOX1} revealed the important DNA sequences for transcriptional activation (Figure 4.18). Region D and Region I were identified as important regulatory parts of P_{AOX1} since deletion of Region D and Region I resulted in, respectively, 84% and 60% decrease in eGFP expression levels than that of the wild-type promoter (Xuan et al., 2009). Hartner et al. (2008) reported that deletion of Adr1 which overlaps with Mxre4, and deletion of dMat1-Mc and d6 which overlap with Mxre6, reduced the promoter activity to 30%, 45%, and 42% of the $P_{AOX1\text{-}wt}$. Therefore, our results reveal that replacing native Mxre regions with alternative Adr1 binding sequences does not alter transcriptional activation of P_{AOX1}. Contrary to the results of Kranthi et al. (2009) which reported that Mxr1 cannot bind to UAS1 of *S. cerevisiae* P_{ADH2}, the design and construction with the replacement of the original Mxre elements with *S. cerevisiae* Adr1 binding motifs did not lead to any reduction in the P_{AOX1} activity, indicating that Mxr1(Adr1) binds on the newly introduced binding motifs to sustain its activation. The sequence: tgcaTTGTctc was identified as a Mat1-Mc motif by MatInspector-Release 6.1 (Hartner et al., 2008); however, the sequence was not identified as a Mat1-Mc motif in our analysis by the upgraded MatInspector-Release 8.4.1. Moreover, Hartner et al. (2008) reported 55% decrease

in the P_{AOX1} activity with the deletion of the sequence in P_{AOX1}; contrarily, with the design of the NEPV $P_{AOX1\text{-}Adr3}$ the destruction of the sequence not disrupt the P_{AOX1} activity is indeed noteworthy, as our results are confronting to those reported by Hartner et al. (2008).

4.3.2 NEPVs of P_{AOX1}-Design-2: Replacement of Mxr1 Binding Motifs with Cat8 Binding Motifs

In silico analysis of P_{AOX1} putative TFBSs by MatInspector did not identified any Cat8 or Sip4 binding sites. There is only one putative carbon source-responsive element for the binding of Rds2, located at -618 to -604 bp. The regulator of drug sensitivity 2, Rds2, is a zinc cluster transcriptional activator involved in the regulation of the gluconeogenesis and glyoxylate shunt genes, conferring resistance to antifungal drug ketoconazole (Soontorngun et al., 2007; Turcotte et al., 2010).

Transcriptional regulatory tools have received attention in synthetic biology applications, since synthetic biology aims to assemble biological components in new ways to obtain novel systems. A main goal of promoter engineering is development of NEPVs using modular and robust genetic elements to fine-tune gene expression and regulation. Ethanol regulated activation of *S. cerevisiae* P_{ADH2} requires collaborative function of Adr1 and Cat8 TFs (Walther and Schüller, 2001). Based on the current knowledge about Adr1 and Cat8, the hypothesis in this work is that introducing Cat8 binding motifs to rationally designed positions in P_{AOX1} can lead to transcriptional activation of P_{AOX1} in response to ethanol which is a repressor molecule for wild-type promoter. The Cat8 binding motif is considered as a modular and context-independent *cis*-acting element to be used for the design of new genetic tools with novel regulations. In this context, Cat8 binding motifs were integrated to three different Mxre positions namely Mxre1, Mxre3, and Mxre5 in close proximity to other native Mxre motifs by replacing the original sequence. The NEPVs $P_{AOX1\text{-}Cat1}$, $P_{AOX1\text{-}Cat2}$, and $P_{AOX1\text{-}Cat3}$ were designed with the architectures depicted in Figure 4.19;

where original nucleotides located in determined positions were mutated as represented, as follows:

Design of $P_{AOX1\text{-}Cat1}$

Mxre1 was replaced by the optimized Cat8 binding site for *S. cerevisiae* (Roth et al., 2004).

ACAGGA**GGGG**ATACACTAGC sequence was replaced with:
ACAATTCCGTTCGTCCGATT

Design of $P_{AOX1\text{-}Cat2}$

Mxre3 was replaced by Cat8 TFBS that was identified on *P. pastoris* P_{ADH2} by MatInspector.

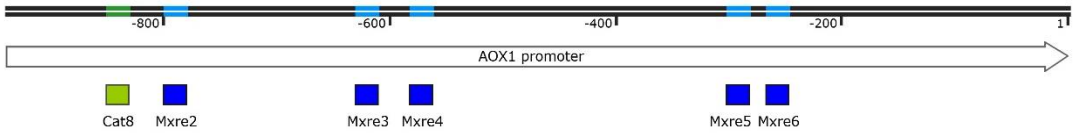
CAACAAG**CTCC**GCATTACAC sequence was replaced with:
CCCTCTCGTCCGGGCTTTTT

Design of $P_{AOX1\text{-}Cat3}$

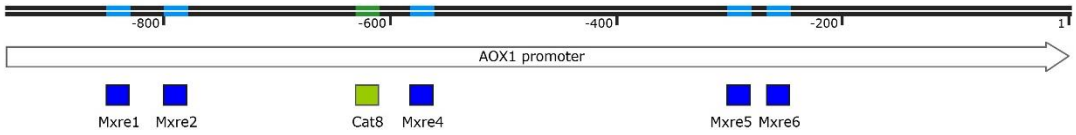
Mxre6 was replaced by the optimized Cat8 binding site for *S. cerevisiae* (Roth et al., 2004)

CGCAAAT**GGGG**AAACACCCG sequence was replaced with:
CATATTCCGTTCGTCCGAAT

a) $P_{AOX1\text{-}Cat1}$



b) $P_{AOX1\text{-}Cat2}$



c) $P_{AOX1\text{-}Cat3}$

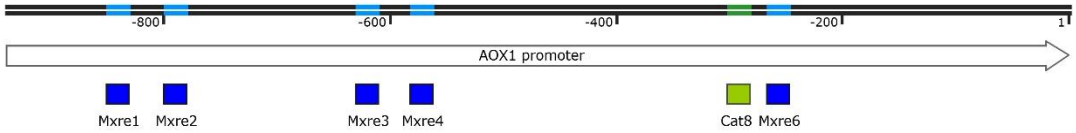


Figure 4.19 Design of the promoter architectures of the NEPVs: **a)** $P_{AOX1\text{-}Cat1}$, **b)** $P_{AOX1\text{-}Cat2}$ and **c)** $P_{AOX1\text{-}Cat3}$

Putative TFBSs of the NEPVs $P_{AOX1\text{-}Cat1}$, $P_{AOX1\text{-}Cat2}$, and $P_{AOX1\text{-}Cat3}$ were analyzed by MatInspector. Deletion of Mxre1 in the design of $P_{AOX1\text{-}Cat1}$ caused a deletion of Msn2 binding motif, which is a transcriptional activator for the genes in the multi-stress response (F$YSTR; matrix: F$MSN2.01); while, a Cat8 (F$CSRE; matrixes: F$CAT8.01), and two Gal4-class motifs (F$MGCM; matrixes: F$LYS14.01; F$RGT1.02) were created. Furthermore, a GATA-type zinc finger protein Gat1 (F$GATA; matrix: F$GAT1.01), and homeodomain transcriptional repressor Matalpha2 (F$YMAT; Matrix: F$MATALPHA2.02) were identified on $P_{AOX1\text{-}Cat1.}$

Deletion of Mxre3 in the $P_{AOX1\text{-}Cat2}$ design resulted in the loss of some potential TFBBs; which are, a Rfx1 (F$RFXP; matrix: F$RFX1.02), two Rap1 (F$YRAP; matrixes: F$RAP1.02, F$RAP1.03), a CSRE denoted by RDS2 (F$CSRE; matrixes: F$RDS2.01), a Aft2 (F$IRTF; matrix: F$AFT2.01), a Swi4 (F$YMCB; matrix:

F$SWI4.01), a Mcm1 (F$YMCM; matrix: F$MCM1.02), a Ste12 (F$PRES; matrix: F$STE12.01), and a Hmra2 (F$YMAT; matrix: F$HMRA2.01) binding motif were deleted based on MatInspector results.

In $P_{AOX1\text{-}Cat2}$ design some putative TFBBs were formed in addition to the targeted Cat8 (F$CSRE; matrix: F$CAT8.01). According to the *in silico* analysis these are: a Gal4 (F$YGAL; matrix: F$GAL4.01), a Asg1 (F$ASG1; matrix: F$ASG1.01), a Pr1 (F$ARPU; matrix: F$PPR1.01), a Rgt1 (F$MGCM; matrix: F$RGT1.02), and two Hal9 (F$HAL9; matrix: F$HAL9.01) binding motifs.

In the design of $P_{AOX1\text{-}Cat3}$, deletion of Mxre5 resulted in the loss of Forkhead transcription factor Fhl1 (F$WHBS; matrix: F$FHL1.01) binding site. Compared to $P_{AOX1\text{-}wt}$, a Cat8 (F$CSRE; matrix: F$CAT8.01), and two Rgt1 (F$MGCM; matrixes: F$RGT1.02) binding motifs were identified (Appendix F).

Introduction of Cat8 binding motifs to natural Cat8 TFBS-less P_{AOX1} led to significant changes in the regulation and expression strength of the promoter. When a Cat8 binding motif was introduced onto the Mxre5 position, ethanol repressive P_{AOX1} transformed to an ethanol-inducible variant with a single modification. The eGFP expression with the NEPV $P_{AOX1\text{-}Cat3}$ with ethanol induction was ca. 74% of the eGFP synthesized with methanol induction with $P_{AOX1\text{-}wt.}$ Contrariwise, the expression capacity with $P_{AOX1\text{-}Cat3}$ increased to 169% of $P_{AOX1\text{-}wt}$ in methanol. Hartner et al. (2008) reported that deletion of d5 (-322 to -264 bp) that includes Mxre5 and some extra nucleotides (Figure 4.16) resulted in a 60% decrease in the activity with P_{AOX1}, is a confronting result. The architecture of $P_{AOX1\text{-}Cat3}$ which is based on Cat8 binding sequence addition and deletion of Mxre5 that is a binding site for the main activator of P_{AOX1} in methanol, contributed to the strength of $P_{AOX1\text{-}Cat3}$ in methanol, is indeed noteworthy. In the methanol-grown cells, replacement of Mxre3 with Cat8 binding motif also increased the expression strength under $P_{AOX1\text{-}Cat2}$ up to 180% that of the obtained with $P_{AOX1\text{-}wt}$; but, $P_{AOX1\text{-}Cat2}$ sustained the ethanol suppression characteristics of $P_{AOX1\text{-}wt}$. Replacement of Mxre1 with Cat8 binding site did not lead to any

significant change in P_{AOX1} regulation and expression. Although, the NEPVs $P_{AOX1\text{-}Cat2}$ and $P_{AOX1\text{-}Cat3}$ exhibit different regulations and strength in response to ethanol and methanol, both are repressed by glucose and glycerol, as stronger inducible variants of the naturally occurred $P_{AOX1\text{-}wt}$ by performing the same tight regulation pattern of P_{AOX1} in response to glucose and glycerol.

Among the designed three NEPVs with Cat8 modifications, the performance of the NEPV $P_{AOX1\text{-}Cat3}$ was the best, and then $P_{AOX1\text{-}Cat2}$. Effects of Cat8 binding sites on the strength and regulation of $P_{AOX1\text{-}v}$ decrease as they are getting farther away from the core promoter and translation start site which is in agreement with the findings of Sharon et al., (2012). Sharon et al., (2012) observed this effect with a shorter (~100-bp) length of the promoter region. Our results reveal the decay in the effectiveness of TFs occurs after a critical length from the start codon as demonstrated by the NEPVs $P_{AOX1\text{-}Cat3}$, $P_{AOX1\text{-}Cat2}$, and $P_{AOX1\text{-}Cat1}$, which are located, respectively, with a spacer length of 282 bp, 610 bp, and 830 bp.

In the design of $P_{AOX1\text{-}Cat2}$ Cat8 binding motif of *P. pastoris* P_{ADH2} was integrated. This motif is found at the minus strand of P_{ADH2} and integrated into the minus strand of P_{AOX1}. After the destruction of the Mxre3 the increased expression with $P_{AOX1\text{-}Cat2}$ in methanol shows that even inversely oriented Cat8 motif of $P_{AOX1\text{-}Cat2}$ exhibits similar function with the straight Cat8 motif of $P_{AOX1\text{-}Cat3}$. Influences of the changed orientations of TFBSs on transcriptional activation were compared by Sharon et al. (2012) and a significant change has been found for 6 yeast transcription factors out of 75 tested in *S. cerevisiae*. Sharon et al. (2012) reported that orientation of Adr1 binding site has a significant effect on transcriptional regulation; whereas, orientation of Cat8, as well as the Sip4 binding site, has no significant effect on transcriptional regulations. Our results demonstrate that Cat8 can substitute the decrease in P_{AOX1} activity resulted from the loss of Adr1(Mxr1) binding site. Cat8 binding motif is not identified on naturally occurring P_{AOX1}; however, the results reveal that Cat8 can collaboratively

function with the native transcriptional architecture for efficient activation of the promoter in methanol.

4.3.3 NEPVs of P_{AOX1}-Design-3: Integration of Aca2 TFBS

Based on *in silico* analysis results by MatInspector, neither Aca1 nor Aca2 binding motifs were identified on P_{AOX1}. *S. cerevisiae* Aca2 was considered important for utilization of non-optimal carbon sources, and Aca2 null mutants were found to be more vulnerable to ethanol and oxidative stresses (Liu et al., 2016). However, the designed $P_{ADH2\text{-}AddAca}$ did not show a significant change in the promoter activity in ethanol.

P. pastoris has only Aca1 (*PAS_chr3_0135*) and not have paralog Aca2(Cst6). Thus, the identified putative Aca1/Aca2 biding sites should be allocating as a binding site for only Aca1, as there is no report on the function of *P. pastoris* Aca1, yet. Therefore, the aim of designing the NEPV $P_{AOX1\text{-}Aca}$ is discovering the function of *P. pastoris* Aca in the AOX1 promoter architecture. Thus, Aca2 binding sequence that was identified in *P. pastoris* P_{ADH2} was integrated at -439 to -419 bp positions of P_{AOX1} by replacement of the original sequence. The NEPV was designed according to the architecture depicted in Figure 4.20, where the original nucleotides located in the determined positions were mutated as presented:

Design of $P_{AOX1\text{-}Aca}$

TGGTCAAAAAGAAACTTCCAA was replaced by:
GCCTATTGTAGACGTCAACCC

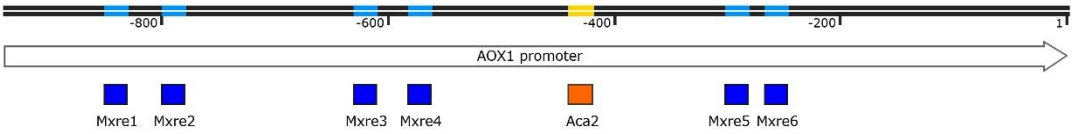


Figure 4.20 Design of the promoter architecture of the NEPV $P_{AOX1\text{-}Aca}$

Putative TFBSs of the NEPV was analyzed using MatInspector. In the design of $P_{AOX1\text{-}Aca}$ a glucose-dependent zinc finger transcriptional activator, Azf1 (F$AZF1; matrix: F$AZF1.01) and Mcm1 (F$YMCM; matrix: F$MCM1.02) binding motifs were destructed. Instead, a targeted Aca2 (F$BZIP; matrix: F$CST6.01), a Mot3 (Modifier of transcription 3) (F$MOT3; matrix: F$MOT3.01), and two mating related TFs Mat1Mc (F$MMAT; matrix: F$MAT1MC.01) and Matalpha 2 (F$YMAT; Matrixes: F$MATALPHA2.02) binding motifs were created.

Integration of an Aca2 binding site into originally Aca2-less AOX1 promoter significantly increased the eGFP expression in the methanol-grown cells. $P_{AOX1\text{-}Aca}$ expression level reached 138% activity of P_{AOX1} in methanol, and the repression by glycerol, glucose, and ethanol were sustained. Hartner et al., (2008) reported that duplication of Mat1-Mc motif led to a 3-fold increase in the activity of P_{AOX1} under derepression conditions. Based on MatInspector results, a Mat1-Mc site was generated; but screening results demonstrated that $P_{AOX1\text{-}Aca}$ was still repressed by glycerol and there was no difference in the activity of the NEPV compared to $P_{AOX1\text{-}wt}$ at limited glucose condition.

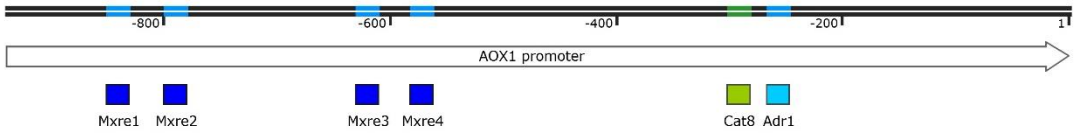
S. cerevisiae Aca2 is a stress-responsive element and used in the utilization of non-optimal carbon sources (Liu et al., 2016). Aca2 null-mutants cannot grow on respiratory carbon sources like ethanol or glycerol (Garcia-Gimeno and Struhl, 2000). Although Aca2 binding motif was identified in naturally occurring *P. pastoris* P_{ADH2}, integration of an additional Aca2 motif did not lead to any change in the activity with $P_{ADH2\text{-}ACA}$ in response to ethanol; however, $P_{AOX1\text{-}Aca}$ demonstrated an increased

transcriptional capacity in response to methanol. P_{ADH2} and P_{AOX1} engineering results advocate that due to the methylotrophic nature of *P. pastoris*, Aca1 may gain new functions; and, instead of ethanol it may become more active on the regulation of methanol utilization pathway genes. However, Aca2 binding motifs found by MatInspector disputed this argument, since there was an Aca2 binding site on P_{ADH2} but not on P_{AOX1}, as these bioinformatics tools cannot always determine the real situation, yet. Therefore, MatInspector could not identify some of the Adr1(Mxr1) binding sites on $P_{ADH2\text{-}v}$. As there is not any TFBS database for *P. pastoris* yet, *in silico* analysis of *P. pastoris* promoters by MatInspector depending on its release number still have a developing capacity which limits high fidelity identification of the TFBSs.

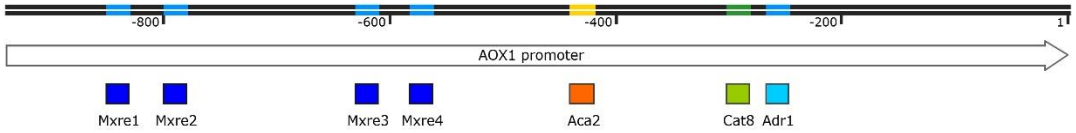
4.3.4 NEPVs of P_{AOX1}-Design-4: Combinatorial Engineering of P_{AOX1}

In the next phase of P_{AOX1} engineering, combinatorial effects of the performed design and constructions were analyzed. As the most challenging results were observed with the mutations that were in close proximity to the core promoter, the findings related to the design were applied together in the design of the NEPVs $P_{AOX1\text{-}Cat3Adr3}$ and $P_{AOX1\text{-}AcaCat3Adr3}$. Consequently, the NEPV $P_{AOX1\text{-}mod}$ was constructed in order to make an evaluation in a broader context. The NEPVs $P_{AOX1\text{-}Cat3Adr3}$, $P_{AOX1\text{-}AcaCat3Adr3}$, and $P_{AOX1\text{-}mod}$ were designed according to the architectures depicted in Figure 4.21.

a) $\mathbf{P}_{AOX1\text{-}Cat3Adr3}$



b) $\mathbf{P}_{AOX1\text{-}AcaCat3Adr3}$



c) $\mathbf{P}_{AOX1\text{-}mod}$

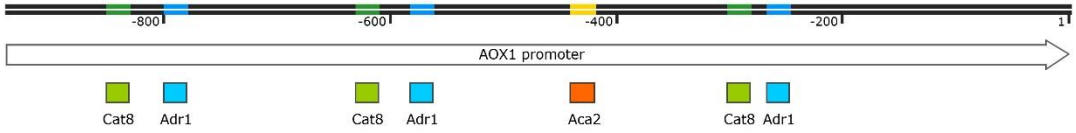


Figure 4. 21 Design of the promoter architectures of the NEPVs: **a)** $P_{AOX1\text{-}Cat3Adr3}$, **b)** $P_{AOX1\text{-}AcaCat3Adr3}$, **c)** $P_{AOX1\text{-}mod}$

In the design of $P_{AOX1\text{-}Cat3Adr3}$ Mxre5 and Mxre6 were replaced by Cat8 and Adr1 binding motifs, respectively. In the single modifications, $P_{AOX1\text{-}Adr3}$ did not show any significant difference compared to $P_{AOX1\text{-}wt}$; while $P_{AOX1\text{-}Cat3}$ was considerably active with ethanol and performed ca. 74% of the expression obtained with that of methanol-induced $P_{AOX1\text{-}wt}$. Furthermore, under $P_{AOX1\text{-}Cat3}$ the expression reached 169% that of the obtained with $P_{AOX1\text{-}wt}$ in the methanol-grown cells. The ethanol and methanol-grown cells constructed with $P_{AOX1\text{-}Cat3Adr3}$ demonstrated 85% and 133% of the expression strength of methanol-induced $P_{AOX1\text{-}wt}$, respectively, is indeed important. Although single AOX1-Adr3 modification did not create any difference, when combined with the AOX1-Cat3 modification, the novel design contributed to the expression in ethanol. Therefore, we can conclude that replacement of Mxre5 with the optimized *S. cerevisiae* Adr1 binding palindromic repeats provide a docking

platform for *P. pastoris* Mxr1(Adr1) TF which function together with the corresponding activator of the newly introduced Cat8 binding site.

When Aca2 binding motif was integrated to P_{AOX1}, the construct enabled ca. 33% increase in activation in methanol, whereas there is not an increase in induction in ethanol. Likewise, Aca2 engineering did not improve the strength of P_{ADH2} in ethanol. Further, when Aca2 binding motif was integrated to $P_{AOX1\text{-}Cat3Adr3}$ variant, the activity of the NEPV in ethanol was reduced from 85% to 60%, while the activity in methanol induction was increased from 133% to 165%. The contribution of Aca2 on transcriptional activation in methanol was observed with the $P_{AOX1\text{-}Aca}$ design with the single-modification, and also with the $P_{AOX1\text{-}AcaCat3Adr3}$ design created with combinatorial modification. Some mutations probably need additional design attempts to demonstrate their intrinsic potential, as was observed with $P_{AOX1\text{-}Cat3Adr3.}$

In order to determine and demonstrate the concerted effect of the seven distinct modifications, $P_{AOX1\text{-}mod}$ was constructed and its expression strength and regulation potential were tested. Among all the P_{AOX1} variants, the highest expression capacity was observed with $P_{AOX1\text{-}mod}$ which reached 130% and 197% activity of the wild-type promoter in the ethanol and methanol-grown cells, respectively. Thus, we conclusively demonstrated that Aca1, Adr1, Cat8-1, Cat8-2, together with native transcriptional architecture of P_{AOX1}, operate in harmony and the concerted effect of which enhanced the transcriptional strength and expression potential under the NEPV $P_{AOX1\text{-}mod}$. P_{AOX1} engineering results demonstrated that P_{AOX1} tight regulation is mainly provided by its *cis*-acting elements. Ethanol repression characteristic of P_{AOX1} was mainly controlled by the absence of Cat8 binding motifs. Here it is shown that replacement of Mxre5 by Cat8 binding motif altered the ethanol repression characteristic of P_{AOX1} to be ethanol-inducible, is indeed important. The significant change in the regulation of P_{AOX1} was achieved with only a single TFBS modification; where, strong ethanol regulated $P_{AOX1\text{-}mod}$ was created using the seven design tools

simultaneously. Expression strength of $P_{AOX1\text{-}mod}$ outcompeted $P_{AOX1\text{-}wt}$ in methanol and ethanol-induced $P_{AOX1\text{-}mod}$ expression strength was on par with $P_{AOX1\text{-}wt}$.

4.3.5 Expression Potential of NEPVs of P_{AOX1}: Influence of Carbon Sources

After the transformation of *P. pastoris* X33 cells, least ten clones for each construct were analyzed for a reliable comparison; and, the clones representing whole population were selected for further analysis. The specific eGFP synthesis values of the constructed novel strains based on the fluorescence intensity related to the cell volume (Hohenblum et al., 2003); and, in eGFP fluorescence calculations, geometric mean of the cell population was used. eGFP synthesis level of $P_{AOX1\text{-}wt}$ under 1% (v/v) methanol induction is considered as 100%; and, eGFP synthesis levels of the constructed strains were correlated to this value.

In order to evaluate the transcriptional potential of $P_{AOX1\text{-}v}$, final screening experiments were performed with three biological replicas of each strain in 24 deep-well plates at 25°C, 225 rpm for 20 h in each production medium. Precultivation was performed in YP Media without glucose. After 20 h precultivation, the cells were harvested by centrifugation and inoculated into each production medium at specified initial cell concentrations, as explained in section 3.7.2. *P. pastoris* cells that carry intracellular eGFP expression cassettes with the NEPVs of P_{AOX1} demonstrated similar cell growth under the same induction condition. Nevertheless, their eGFP production levels showed different patterns. In order to evaluate the regulation and expression strength of the NEPVs of P_{AOX1}, *P. pastoris* cells containing P_{AOX1} variants with *eGFP* gene were screened with fermentations using five different carbon sources, *i.e.* E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol. The eGFP expression levels were calculated relative to $P_{AOX1\text{-}wt}$ productivity in methanol. The results are presented in Figure 4.22 and Table 4.11.

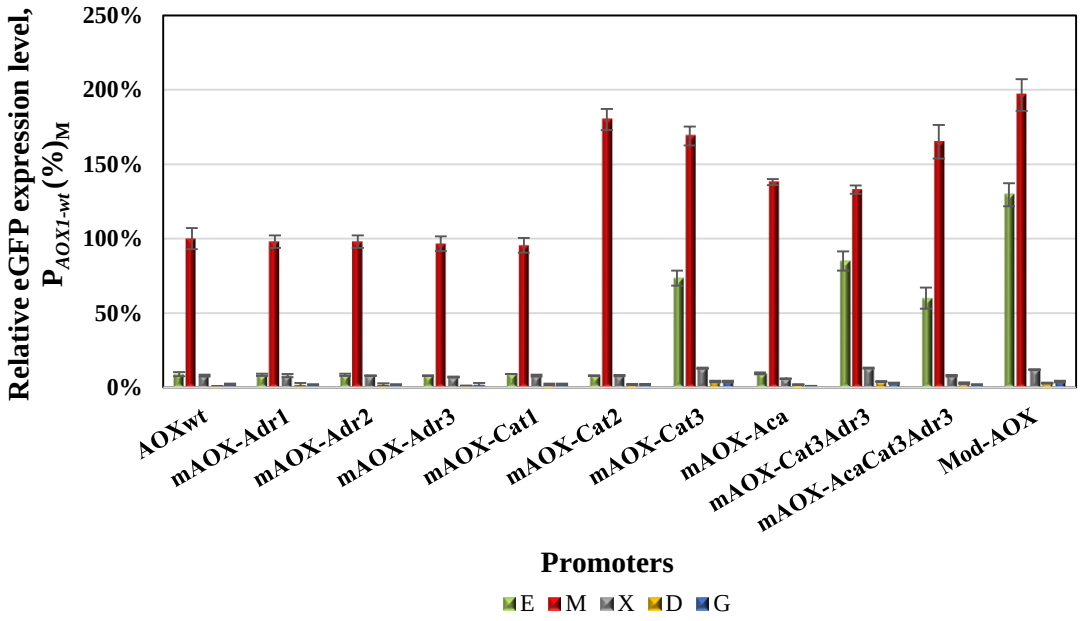


Figure 4.22 Normalized eGFP synthesis capacity of novel *P. pastoris* strains with the NEPVs relative to $P_{AOX1\text{-}wt}$ (%)$_M$ in different carbon sources at the cultivation time of t = 20 h. Calculations were carried out based on arbitrary fluorescence units. Mean values of 3 different clones are presented. Detailed information including statistical analysis is presented in Table 4.11. Carbon sources: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, G: excess glycerol, D: excess glucose. Error bars represent the standard deviation (±)

Screening results demonstrate that all the NEPVs of P_{AOX1} sustain their regulated gene expression pattern. Although some of them exhibit increased strength in response to methanol and ethanol, all are still repressed by the excess glucose and excess glycerol. At limited glucose condition, $P_{AOX1\text{-}Cat3}$, $P_{AOX1\text{-}Cat3Adr3}$, and $P_{AOX1\text{-}mod}$ exhibited slightly higher derepression activities than $P_{AOX1\text{-}wt}$.

Table 4.11 The normalized eGFP synthesis capacity of novel *P. pastoris* strains with the NEPVs relative to $P_{AOX1\text{-}wt}$ $(\%)_M$ in different carbon sources at the cultivation time of t = 20 h. Mean values of 3 different clones are presented. Significance of differences to the control strain was calculated with Student's t-test. (* p-values <0.05; ** p-values <0.01). Carbon sources: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol.

Normalized eGFP concentrations					
P_{AOX1} Variant	**E**	**M**	**X**	**D**	**G**
AOX1-wt	9±1	100±7	8±1	1±0	2±1
mAOX1-Adr1	9±1	98±4	8±1	2±1	2±0
mAOX1-Adr2	9±1	98±4	8±0	2±1	2±0
mAOX1-Adr3	8±0	97±5	7±0	1±0	2±1
mAOX1-Cat1	9±0	96±5	8±1	2±1	2±1
mAOX1-Cat2	8±0	180±7**	8±0	2±0	2±0
mAOX1-Cat3	74±5**	169±6**	13±1**	4±1	4±0
mAOX1-Aca	10±1	138±2**	6±0	2±0	1±0
mAOX1-Cat3Adr3	85±6**	133±3**	13±1**	4±1	3±0
mAOX1-AcaCat3Adr3	60±7**	165±11**	8±1	3±1	2±0
Mod-AOX1	130±8**	197±11**	12±0**	3±0	4±1

The contribution of Cat8 to methanol utilization pathway was not identified yet. Further in the literature, *in silico* analysis of $P_{AOX1\text{-}wt}$ did not reveal any Cat8 binding motif. Our results indicate that Cat8 can collaboratively function with inherent promoter architecture of P_{AOX1} for increasing the activation in methanol; and, more importantly achieve ethanol-induced activation. *P. pastoris* Aca1 also exhibited an activator function in the MUT pathway although its homolog *S. cerevisiae* Aca2 known as an activator in ethanol, glycerol, and raffinose utilization pathways (Garcia-Gimeno and Struhl, 2000). In accordance with the literature, the Aca2 binding motif was not identified on $P_{AOX1\text{-}wt}$ but was identified on P_{ADH2}. Overall, our findings demonstrated that Aca2 engineering was mainly effective in P_{AOX1}. Due to the methylotrophic nature of *P. pastoris*, we conclude that *P. pastoris* Aca1 gained new functions while losing some others. Therefore, since methanol is the prominent

carbon source of *P. pastoris,* Aca1 may function on the activation of the methanol utilization genes. The reason of not identifying any Aca1 and Aca2 binding motif in P_{AOX1} by *in silico* analysis can be the lapse of the software due to differences between the DNA binding motif of *P. pastoris* Aca1 and *S. cerevisiae* TFBS database. The influence of the TFBS is linked to their proximal and distal elements which can be stimulatory or inhibitory on the promoter. Therefore, Aca1 and Aca2 probably influence the activity of P_{ADH2} if its binding motif was inserted in some other nucleotide positions within the promoter sequence.

The results revealed that, promoter engineering approaches on P_{AOX1} created a P_{AOX1} library that span in the range of 74% to 197% activity of $P_{AOX1\text{-}wt}$; furthermore, P_{AOX1} repression in ethanol was altered with a single TFBS modification. Although promoters have very complicated activation and repression mechanisms, *P. pastoris* P_{AOX1} repression in ethanol is mainly controlled in the absence of the Cat8 binding site. Therefore, if a Cat8 motif is introduced to a rationally designed position on P_{AOX1}, the NEPV can be induced by ethanol as much as methanol. Aca1 can also be used to strengthen P_{AOX1} in methanol.

4.3.6 eGFP Transcript Levels of the NEPVs of P_{AOX1}

To confirm the enhancement in the strength and altered regulation of the the AOX1 promoter variants, the transcription levels of *eGFP* in *P. pastoris* variants were analyzed by means of qPCR experiments. Comparative Ct method (ΔΔCt) was used for the relative quantification. mRNA expressions of *eGFP* were normalized relative to *Actin* mRNA expression; thereafter, divided by normalized *eGFP* expression performed with $P_{AOX1\text{-}wt}$, and mRNA transcription ratios (mTRs) were determined and presented in Figure 4.23 and Table 4.12.

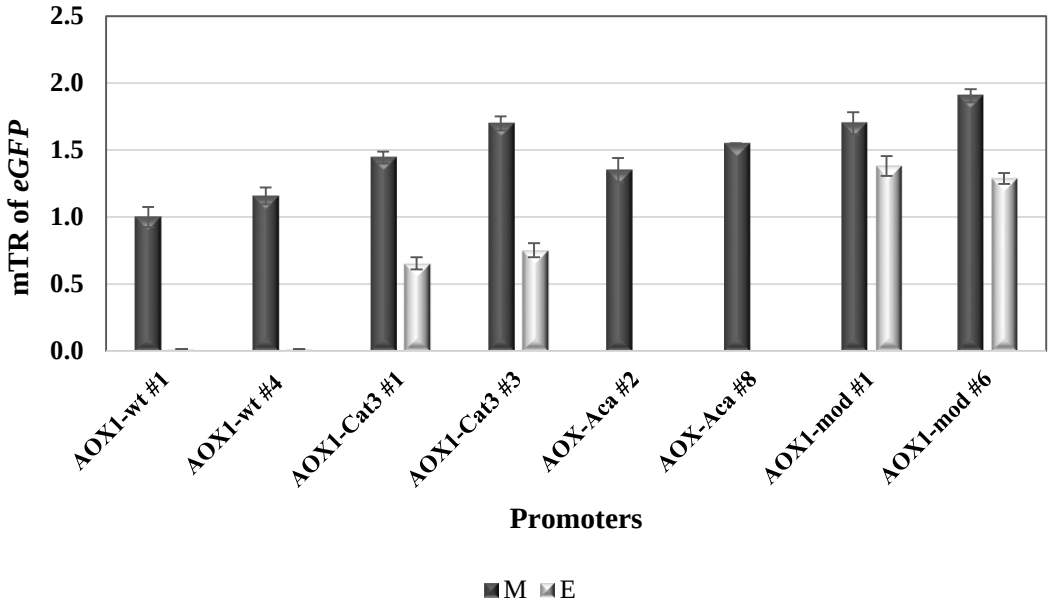


Figure 4. 23 mRNA transcription ratios (mTRs) of *eGFP* in 2% (v) ethanol and 1% (v/v) methanol grown *P. pastoris* cells under $P_{AOX1\text{-}v}$ control at t = 20 h. Transcript levels were normalized to *ACT1* mRNA expression and presented relative to methanol induced $P_{AOX1\text{-}wt}$. The error bars represent the standard deviation of three technical replicates. M: 1% (v/v) methanol; E: 2% (v/v) ethanol.

Subsequent measurement of the transcription levels confirmed that the NEPVs $P_{AOX1\text{-}Cat3}$, $P_{AOX1\text{-}Aca}$, and $P_{AOX1\text{-}mod}$ have stronger transcription capacity with the mTRs of 1.6, 1.5, and 1.8 of $P_{AOX1\text{-}wt}$, respectively. Further, in 2% (v/v) ethanol, no eGFP transcription was detected by $P_{AOX1\text{-}wt}$. However, at the same ethanol concentration, $P_{AOX1\text{-}Cat3}$ and $P_{AOX1\text{-}mod}$ enhanced the mTR, respectively, to 0.7- and 1.3 compared to that of the 1% (v/v) methanol-induced P_{AOX1}.

Table 4. 12 mRNA transcription ratios (mTRs) of *eGFP* in methanol and ethanol grown *P. pastoris* cells carrying $P_{AOX1\text{-}v}$ at t = 20 h. Transcript levels were normalized to *ACT1* mRNA expression and presented relative to methanol induced $P_{AOX1\text{-}wt}$. (M):1% (v/v) methanol, (E): 2% (v/v) ethanol; StDev: standard deviation of three technical replicates

Promoter	mTR of *eGFP*	StDev
AOX1-wt #1 (M)	1.0	0.1
AOX1-wt #4 (M)	1.2	0.1
AOX1-Cat3 #1 (M)	1.4	0.0
AOX1-Cat3 #3 (M)	1.7	0.1
AOX-Aca #2 (M)	1.4	0.1
AOX-Aca #8 (M)	1.6	0.0
AOX1-mod #1 (M)	1.7	0.1
AOX1-mod #6 (M)	1.9	0.1
AOX-wt #1 (E)	0.0	0.0
AOX1-wt #4 (E)	0.0	0.0
AOX1-Cat3 #1 (E)	0.7	0.0
AOX1-Cat3 #3 (E)	0.8	0.1
AOX-mod #1 (E)	1.4	0.1
AOX-mod #6 (E)	1.3	0.0

4.3.7 Extracellular Human Serum Albumin Production by the NEPVS of P_{AOX1}

In order to test the potential of the prominent NEPV of P_{AOX1} in extracellular r-protein production, human serum albumin gene (*HSA*), with its native signal sequence was cloned under the control of $P_{AOX1\text{-}mod}$ and for the comparison under $P_{AOX1\text{-}wt}$. Recombinant plasmids were integrated into AOX1 transcription termination locus, where the promoter variants were independent of the effects of genomic integration sites; and least ten clones per each construct were screened to obtain reliable results. Extracellular HSA production was performed in 24 deep-well plates at 25 °C, 225 rpm for 49 h. The precultivation step was performed in YP media at 25 °C and 225 rpm for 20 h, then the cells were harvested by centrifugation and the fermentations were started with inoculation of the cells at an OD_{600} value of 1 in ASMv6 minimal

medium. At t = 0 h, 0.5% (v/v) methanol or 1% (v/v) ethanol was added into shake bioreactors; and, 1% (v/v) methanol or ethanol was introduced at t = 5, t = 17, t = 29, and t = 41 h of the fermentations to induce the r-protein production. The cells were harvested at t = 49 h. Extracellular HSA production concentrations were analyzed with ELISA kit as explained in Section 3.8.3, and the results are presented in Figure 4.24 and Table 4.13.

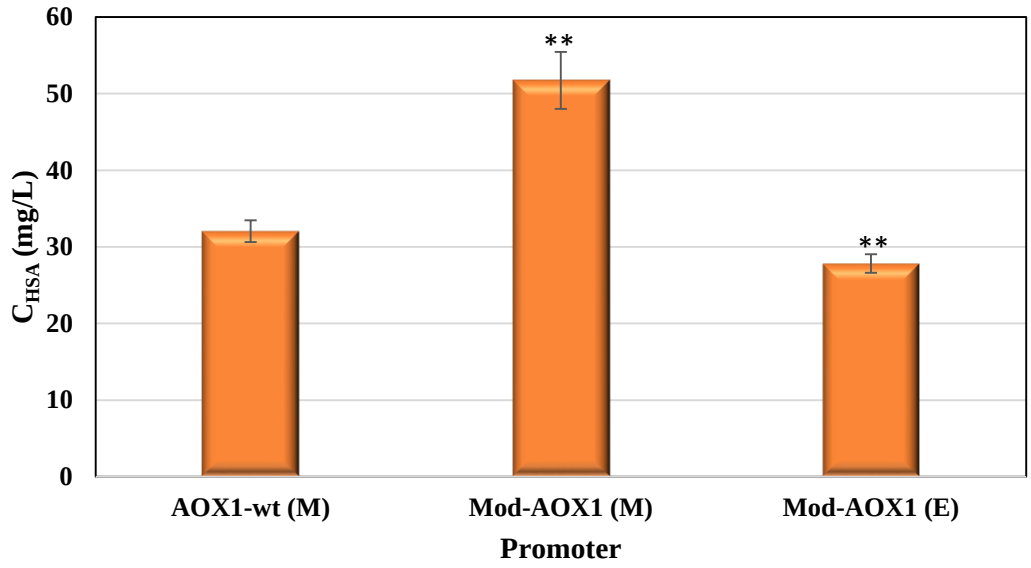


Figure 4.24 Extracellular human serum albumin concentrations (C_{HSA}) of the *P. pastoris* cells under control of $P_{AOX1\text{-}wt}$ and $P_{AOX1\text{-}mod}$ at t = 49 h. (M): methanol pulse feed fermentation, (E): ethanol pulse feed fermentation. Error bars represent the standard deviation among ten different clones (±). Significance of the differences to the $P_{ADH2\text{-}wt}$ was calculated with Student's t-test (** p-values <0.01).

Table 4. 13 Extracellular human serum albumin production titers of *P. pastoris* cells under $P_{AOX1\text{-}wt}$ and $P_{AOX1\text{-}mod}$ at t = 49 h of pulse feed fermentation. StDev: Standard deviation among ten different clones. Significance of differences to the $P_{ADH2\text{-}wt}$ was calculated with Student's t-test (** p-values <0.01).

P_{AOX1} Variant	**Carbon source**	**C_{HSA} (mg/L)**	**StDev**
AOX1-wt	Methanol	32.1	1.4
Mod-AOX1	Methanol	51.7**	3.7
Mod-AOX1	Ethanol	27.8**	1.2

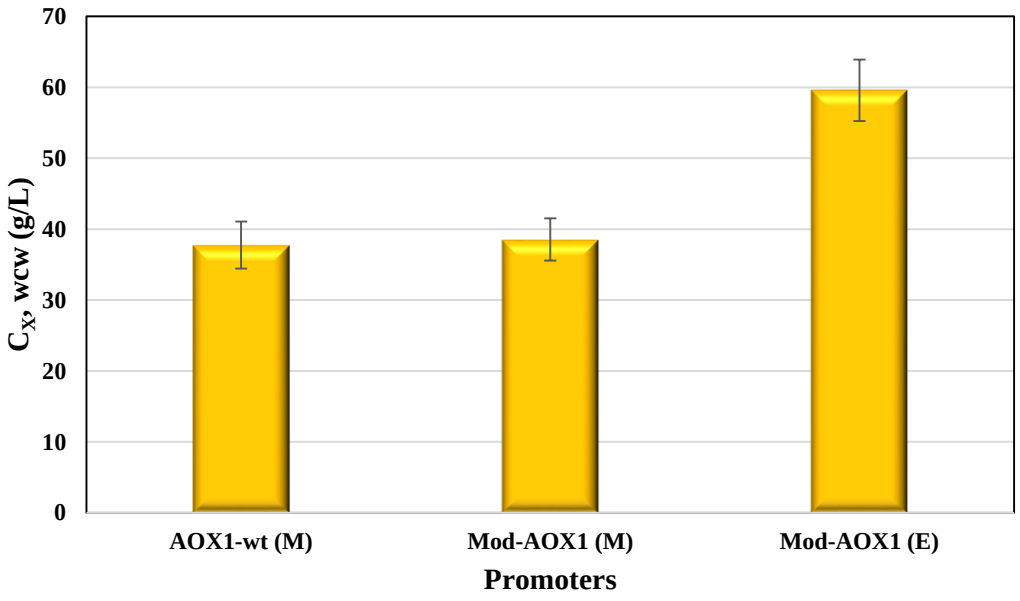


Figure 4.25 The cell concentrations of methanol- and ethanol-grown *P. pastoris* strains producing HSA under P_{AOX1} variants at t = 49 h. (M): methanol pulse feed fermentation, (E): ethanol pulse feed fermentation. Error bars represent the standard deviation among ten different clones (±). C_X, wet cell weight (wcw, g/L)

Table 4.14 The cell concentrations of methanol- and ethanol-grown *P. pastoris* strains producing HSA under different P_{AOX1} variants at t = 49 h of pulse feed fermentation. StDev: standard deviation among ten different clones.

P_{AOX1} Variant	Carbon source	C_X, wcw (g/L)	StDev
AOX1-wt	Methanol	37.7	3.3
Mod-AOX1	Methanol	38.5	2.9
Mod-AOX1	Ethanol	59.5	4.3

The average HSA yield per gram wet-cell-weight was calculated as $Y_{P/X}$ = 0.848 mg/g for the *P. pastoris* strains producing extracellular HSA under the control of $P_{AOX1\text{-}wt}$; whereas, for the *P. pastoris* strains producing HSA under the control of $P_{AOX1\text{-}mod}$ the yield coefficients $Y_{P/X}$ were calculated as 1.341 mg/g and 0.476 mg/g in methanol- and ethanol-based medium, respectively. The *P. pastoris* strains with $P_{AOX1\text{-}mod}$ reached 1.59-fold higher product yield than that of the $P_{AOX1\text{-}wt}$ in methanol and, 0.55-fold of the $Y_{P/X}$ value of $P_{AOX1\text{-}wt}$ in ethanol. Under the designed fermentation conditions the cell yield was higher in ethanol containing minimal medium. The productivity of the cells carrying $P_{AOX1\text{-}mod}$ in ethanol (27.8 µg/ml) is almost equal to the productivity of the methanol-induced $P_{AOX1\text{-}wt}$ (32.1 µg/ml).

Differences between extracellular HSA production and intracellular eGFP synthesis mainly can be because of different screening and induction conditions between eGFP and HSA production processes. However, differences in amino acid composition between two model proteins and changing secretion conditions with different carbon sources should also lead to differences in expression and secretion potential of r-*P. pastoris* strains. The preferred ethanol concentration in the extracellular HSA production was kept quite low compared to eGFP screening to avoid stress-related protease activity, is indeed important. However, under the controlled fermentation conditions, higher ethanol concentrations can be used for more efficient r-protein production. Here it is demonstrated that stronger expression potential and different

ethanol based regulation pattern of the NEPV $P_{AOX1\text{-}mod}$ was confirmed with the intracellular eGFP synthesis, eGFP transcript levels, and extracellular HSA production. Based on the results in this work, it is conclusively demonstrated that $P_{AOX1\text{-}mod}$ is a promising candidate for r-protein production.

4.4 Transcriptional Engineering

4.4.1 Transcriptional Engineering of P_{ADH2} Variants

The functions of the *P. pastoris* transcription factors Adr1(Mxr1), Aca1, Cat8-1, and Cat8-2 on the regulation of P_{ADH2} and the cell metabolism are not clearly known yet. In order to clarify the roles of TFs whose binding sequences were used in P_{ADH2} engineering in section 4.2, TF overexpression and knock-out *P. pastoris* strains were designed and constructed. Since strong and constitutive overexpression of Mxr1 damage the *P. pastoris* viability (Lin-Cereghino et al., 2006; Vogl et al., 2018), the repressible promoter of THI11 (P_{THI11}) was selected with moderate strength in the design of TF overexpression cassettes. The protein THI11 (PP7435_Chr4–0952) is involved in the synthesis of thiamine precursor hydroxymethylpyrimidine. P_{THI11} reached 63% activity of P_{GAP} in the absence of thiamine but is repressed by the addition of thiamine (Delic et al., 2013). Transcription factor overexpression cassettes were designed with P_{THI11} and the transcription terminator of the protein component of the 40S ribosome (PP7435_Chr1–0118) (*RPS3tt*). For the construction of the knock-out strains, GoldenPiCS was used to assemble CRISPR/Cas9 plasmids (Prielhofer et al., 2017). Humanized Cas9 was cloned with *P. pastoris* P_{PFK300} and terminator Sc*CYCtt*. The designed single guide RNA was cloned with P_{GAP} and terminator *RPS25Att*. Cas9 and single guide RNA expression cassettes were cloned into a single episomal plasmid.

P. pastoris strains harboring eGFP synthesis cassettes under the control of $P_{ADH2\text{-}v}$ were used as parent strains for the development of overexpression (OE) and knock-out (KO) strains. Among the TFs used, deletion of the *ACA1* could not be achieved

after several mutation deletion trials with two different guide RNAs, which enables to put forward that Aca1 can be an essential gene for *P. pastoris* viability.

P. pastoris strains carrying $P_{ADH2\text{-}wt}$ and the NEPVs $P_{ADH2\text{-}AddAdr}$, $P_{ADH2\text{-}OptAdr1\text{-}3}$, and $P_{ADH2\text{-}NucOpt}$ were transformed with Adr1 OE-cassette. eGFP synthesis levels of the transcriptionally engineered cells (TECs) were investigated in ethanol and methanol containing fermentation media. None of the *P. pastoris* Adr1-OE strains showed increased eGFP synthesis compared to the parent strain in response to ethanol; however, all became highly active in methanol with a similar improvement in their productivity.

P. pastoris strains carrying $P_{ADH2\text{-}wt}$, and the NEPVs $P_{ADH2\text{-}AddAca}$, and $P_{ADH2\text{-}NucOpt}$ were transformed with Aca1 OE-cassette. Aca1-OE strains demonstrated similar activity with that of the Adr1-OE strains. ACA1-OE cassette contributed to the activation of P_{ADH2} variants with similar increases in eGFP concentrations in methanol, but not with a similar locus in ethanol.

P. pastoris strains carrying $P_{ADH2\text{-}wt}$ and the NEPVs $P_{ADH2\text{-}OptCat}$, and $P_{ADH2\text{-}NucOpt}$ were transformed with Cat8-1-OE and Cat8-2-OE cassettes. None of the overexpression strains constructed with the P_{ADH2} variants demonstrated a change in eGFP expression in ethanol containing medium; while demonstrated usually stronger expressions in methanol. Since the TECs carrying P_{ADH2} variants and $P_{ADH2\text{-}wt}$ demonstrated similar progress in expression levels, further experiments for screening were carried out with the strains constructed with $P_{ADH2\text{-}wt}$. In order to evaluate the P_{ADH2} activity in TF-OE and TF-KO strains, the screening experiments were carried out in 24-deep well plates with fermentations using five different carbon sources. *P. pastoris* OE and KO strains were precultivated in YP media at 25 °C, 280 rpm for 20 h. The cells were harvested by centrifugation and inoculated into ASMv6 minimal media containing one of the tested carbon sources. The fermentations were carried out at 25 °C, 280 rpm for 20 h. The cells were harvested and diluted in PBS buffer to an OD_{600} value of 0.4, and fluorescence intensities of the cells were measured with Flow Cytometry. Normalized

eGFP concentrations of OE and KO strains are presented in Figure 4.26 and Table 4.15.

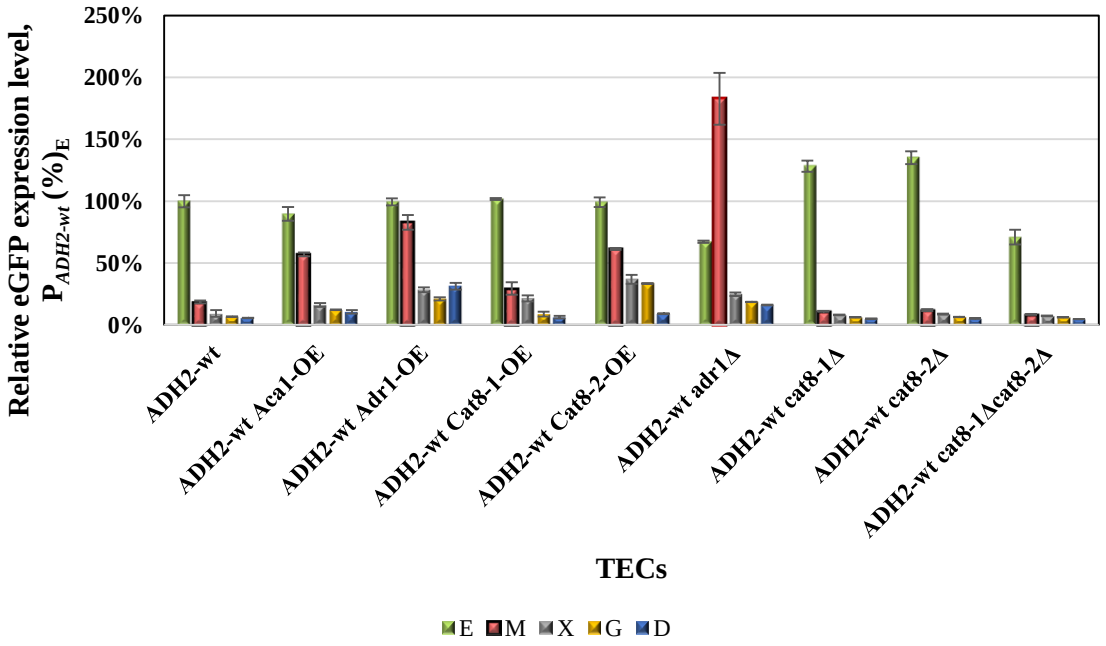


Figure 4.26 Normalized eGFP synthesis capacity of transcriptionally engineered *Pichia pastoris* strains with $P_{ADH2\text{-}wt}$ with different carbon sources at t = 20 h. Mean values of 3 different clones are presented. Error bars represent the standard deviation (±). Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol. Detailed information including statistical analysis is presented in Table 4.15.

Table 4. 15 Normalized eGFP synthesis capacity of TF-OE and TF-KO *P. pastoris* strains with $P_{ADH2\text{-}wt}$ relative to ethanol grown wild-type *P. pastoris* strain carrying $P_{ADH2\text{-}wt}$ at the cultivation time of t = 20 h. Mean values of 3 different clones are presented. Significance of differences to the control strain was calculated with Student's t-test. (* p-values <0.05; ** p-values <0.01). Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol

Normalized eGFP concentrations in the TECs					
***P. pastoris* Strain**	**E**	**M**	**X**	**G**	**D**
ADH2-wt	100±5	19±1	9±3	7±0	6±0
ADH2-wt Aca1-OE	90±6	57±2**	16±1*	12±0**	11±1*
ADH2-wt Adr1-OE	99±3	83±6**	28±2**	21±1**	31±3**
ADH2-wt Cat8-1-OE	102±1	29±5*	21±2**	9±2	6±1
ADH2-wt Cat8-2-OE	99±4	61±1**	37±4**	33±0**	9±0**
ADH2-wt *adr1*Δ	67±1**	183±21**	25±1**	19±0**	16±0**
ADH2-wt *cat8-1*Δ	128±5**	11±0**	8±0	6±0	5±0
ADH2-wt *cat8-2*Δ	135±5**	12±1**	9±0	6±0	5±0
ADH2-wt *cat8-1*Δ*cat8-2*Δ	71±6**	8±0**	7±0	6±0	5±0

Ethanol induced intracellular eGFP relative concentrations in wild-type *P. pastoris* strain carrying naturally occurring $P_{ADH2\text{-}wt}$ was considered as 100 units, and eGFP concentrations were correlated accordingly. For simplicity purposes in this part of the thesis, ethanol induced wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$ will be mentioned as "the wild-type strain". In ethanol-grown *P. pastoris* cells, overexpression of TFs did not lead to any change in the promoter performance; but in methanol, the eGFP expressions in the strains constructed with P_{ADH2} transformed into Aca1-OE, Adr1-OE, Cat8-1-OE, and Cat8-2-OE reached, respectively, 57%, 83%, 29%, and 61% expression level of the wild type strain. The highest derepression effect was observed with Cat8-2-OE strain at limited glucose and excess glycerol; where, with $P_{ADH2\text{-}wt}$ the expression reached 37% and 33% of the ethanol induced wild-type promoter activity, respectively.

In excess glucose, only Adr1-OE strain increased the expression while others continue to sustain repressed characteristics of P_{ADH2}. Deletion of *ADR1* gene reduced the expression under P_{ADH2} to 67 % that of the wild-type strain. But in methanol, *adr1Δ* mutant with P_{ADH2} showed a significant increase in activity reaching 183% of the wild-type strain despite the cells failed to grow. These results demonstrate that Adr1 is not the main regulator of *P. pastoris* P_{ADH2}. Even in the absence of Adr1, P_{ADH2} shows remarkable expression performance in the ethanol-grown cells. Further, in *adr1*Δ mutants, the expression performance of P_{ADH2} increased significantly with the carbon sources methanol, limited glucose, excess glucose, and excess glycerol. The results reveal that Adr1 is an activator of P_{ADH2} in the ethanol-grown cells. But with the substrates methanol, glucose, and glycerol, Adr1 functions as a transcriptional activator and also a repressor of P_{ADH2}. Since eGFP synthesis is high in Adr1-OE and as well as in Adr1-KO strains, the changes in the concentration of Adr1 in the TECs can influence the other TFs that are involved in the activation or repression of P_{ADH2}.

In the ethanol-grown *cat8-1*Δ and *cat8-2*Δ mutants, P_{ADH2} expressions increased, respectively, to 128% and 135% of the wild-type strain, while both deletions create repression on P_{ADH2} in the methanol-grown cells. The evidence for *P. pastoris* is contradictory to that of reported for *S. cerevisiae*; where Cat8 is a positive regulator of P_{ADH2}, and deletion of Cat8 reduced the activity of *S. cerevisiae* P_{ADH2} to 12.5% of the wild-type *S. cerevisiae* (Walther and Schüller, 2001). In order to gain a deeper understanding, *P pastoris* Cat8-1 and Cat8-2 transcript levels were investigated; and, a regulation between *P. pastoris* Cat8-1 and Cat8-2 was identified, which is indeed interesting. *CAT8-1* transcript levels increased 1.5-fold in the ethanol-grown *cat8-2*Δ strains, while *CAT8-2* transcript levels increased 1.5-fold in the ethanol grown *cat8-1*Δ strains. No significant regulation between Cat8-1 and Cat8-2 was identified in the methanol-grown cells. Therefore, we found out that when one of the Cat8 TFs, either Cat8-1 or Cat8-2, exists in *P. pastoris,* P_{ADH2} activation enhances. P_{ADH2} activation in the carbon source ethanol with Cat8-1, and in methanol with Cat8-2, enhanced the expression performance in *P. pastoris* strains. To clarify the regulatory functions of

Cat8-1 and Cat8-2 on P_{ADH2}, a double knock-out strain was designed and constructed. The *cat8-1*Δ*cat8-2*Δ mutant P_{ADH2} expression was decreased to 71% of the wild-type strain; furthermore, *P. pastoris* lost its ethanol utilization ability. Thus, it is conclusively demonstrated that Cat8-1 and Cat8-2 TFs are the transcriptional activators of P_{ADH2} in *P. pastoris*.

However in *S. cerevisiae*, CSRE is occupied by the TFs Cat8 or Sip4 in activation of the gluconeogenesis, glyoxylate shunt, and ethanol utilization pathway genes (Hiesinger et al., 2001). Both Cat8 and Sip4 can bind to the CSRE independently but unequally. Under glucose derepression condition, *S. cerevisiae* Cat8 and Sip4 contribute to the activation of CSRE-dependent genes 85% and 15%, respectively (Hiesinger et al., 2001). *S. cerevisiae* P_{ADH2} UAS2-dependent gene activation is specifically fulfilled by Cat8, but not with the related zinc cluster transcriptional activator Sip4 (Walther and Schüller, 2001). In *P. pastoris,* two homolog proteins, namely, Cat8-1 (*PAS_chr2-1_0757*) and Cat8-2 (*PAS_chr4_0540*) are determined, and both show high homology to *S. cerevisiae* Cat8 and *S. cerevisiae* Sip4. The results demonstrate that both *P. pastoris* Cat8-1 and Cat8-2 can equally contribute to activation of *P. pastoris* P_{ADH2} which exhibits different regulation than in *S. cerevisiae*.

Table 4.16 The cell concentrations of *P. pastoris* TF-OE and TF-KO strains at t = 20 h, in terms of OD_{600} values. Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol

The cell concentrations in terms of OD_{600} values					
TECs	**E**	**M**	**X**	**G**	**D**
ADH2-wt	3.1	3.3	11.2	3.4	2.7
ADH2-wt Aca1-OE	3.1	1.8	11.4	1.8	1.8
ADH2-wt Adr1-OE	3.1	2.0	12.2	2.6	2.4
ADH2-wt Cat8-1-OE	2.8	3.5	11.6	3.5	2.3
ADH2-wt Cat8-2-OE	2.8	3.2	11.6	3.3	2.1
ADH2-wt *adr1*Δ	1.5	1.0	1.8	0.5	0.5
ADH2-wt *cat8-1*Δ	1.5	3.3	9.6	3.7	2.5
ADH2-wt *cat8-2*Δ	2.1	3.5	5.7	3.1	2.4
ADH2-wt *cat8-1*Δ*cat8-2*Δ	1.0	2.9	5.7	3.8	3.0

The growth characteristics of TECs reveal that, the growth of Aca1-OE strain was low in the fermentations in methanol, excess glucose, and excess glycerol; but not in ethanol and limited glucose fermentations (Table 4.16). The cell concentration of the Adr1-OE strain was also low in methanol, although Adr1 (Mxr1) is the main regulator of the MUT pathway in *P. pastoris* (Lin-Cereghino et al., 2006). On the other hand, for the overexpression of Adr1 under the promoter P_{GAP}, meanwhile Vogl et al (2018) reported the decrease in the transformation rate. Adr1-OE and Adr1-KO results indicate that there is a fine-tuned Adr1 concentration level in the cell. Even small changes significantly affect the cell metabolism, especially in methanol. Lin-Cereghino et al., (2006) showed that *adr1*Δ mutants cannot grow in methanol, while there is no difference in the cell growth with other carbon sources. Similar to the findings of Lin-Cereghino et al., (2006), *P. pastoris adr1*Δ mutants lost their ability to grow in methanol containing minimal media. However, contrary to the results of Lin-Cereghino et al., (2006), *P. pastoris adr1*Δ mutants demonstrated growth impairment in all the other carbon sources, as well. Whole genome transcriptome profiling of *S. cerevisiae* using DNA microarrays revealed 108 Adr1-dependent genes

for expression (Young et al., 2003). The major categories of the functionally annotated Adr1-dependent genes are related to non-fermentative carbon metabolism (21%), peroxisome biogenesis and β-oxidation (11%), amino acid transport and metabolism (11%), meiosis and sporulation (8%), and transcriptional regulation, and signal transduction (6%); whereas, 30% of Adr1-dependent genes have unknown functions yet (Young et al., 2003). The growth impairment observed by *P. pastoris adr1*Δ in minimal media containing methanol, ethanol, limited glucose, excess glucose, or excess glycerol, can be because of genome-wide deeply integrated regulatory effects in *P. pastoris* Adr1(Mxr1). Cat8-1 and Cat8-2 OE strains demonstrated similar cell yield to their parent strains. *P. pastoris cat8-1*Δ strain reached lower cell concentrations in ethanol containing media while *CAT8-2* deletion has a less pronounced effect on the cell growth. Moreover, *cat8-1*Δ*cat8-2*Δ mutant strain cannot grow in minimal media with ethanol as sole carbon source.

TECs carrying P_{ADH2} generated some novel ethanol-free r-protein expression systems with $P_{ADH2\text{-}wt}$ including derepression in limited glucose and excess glycerol. The highest derepression potential was observed in Cat8-2-OE strain, in which 37% of the wild-type activity was achieved at limited-glucose fermentation. The highest ethanol induction was achieved in *cat8-2*Δ strain, where P_{ADH2} performed 135% activity of the wild-type strain. Adr1-OE increased the methanol induction potential of P_{ADH2} to 83% that of the wild-type. However in methanol, deletion of the Adr1 resulted in higher expression with P_{ADH2} that reached 183% expression level of the wild-type; where, the cell concentration of *adr1*Δ did not increase in methanol but the eGFP expression continued. Therefore we can assume that *P. pastoris adr1*Δ strain can maintain the cellular activities at a basal level at limited methanol utilization condition.

4.4.2 Transcriptional Engineering of P_{AOX1} Variants

The TECs carrying $P_{AOX1\text{-}wt}$ and the NEPV $P_{AOX1\text{-}mod}$ were designed by overexpression and deletion of *P. pastoris* Adr1(Mxr1), Aca1, Cat8-1, and Cat8-2 transcription factors. Same overexpression and knock-out cassettes were used for the design and construction of the TECs carrying P_{AOX1}, as in section 4.4.1 for the P_{ADH2}. *P. pastoris* strains harboring eGFP synthesis cassettes under the control of $P_{AOX1\text{-}wt}$ and $P_{AOX1\text{-}mod}$ were used as parent strains for the development of overexpression (OE) and knock-out (KO) strains. Among the TFs, deletion of ACA1 could not be achieved after several deletion mutation trials with two different guide RNAs. In order to evaluate the P_{AOX1} variants activity in TF-OE and TF-KO strains, screening experiments were performed in 24-deep well plates with fermentations with five different carbon sources. Precultivation of *P. pastoris* OE and KO strains were performed in YP media at 25 °C, 280 rpm, for 20 h. The cells were harvested by centrifugation and inoculated into ASMv6 minimal media containing one of the tested carbon sources. Production phase was performed at 25 °C, 280 rpm for 20 h. The cells were harvested and diluted in PBS buffer to an OD_{600} value of 0.4, and eGFP fluorescence intensities of the cells were measured by Flow Cytometry. Normalized eGFP concentrations of OE and KO strains are presented in Figure 4.27 and Table 4.17.

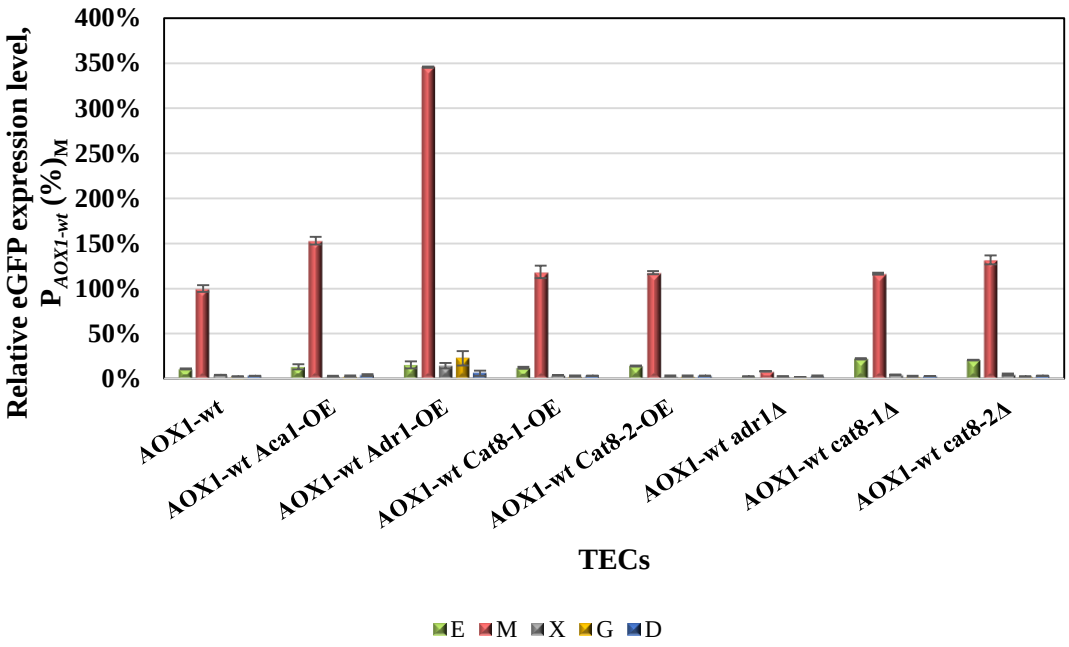


Figure 4.27 Normalized eGFP synthesis capacity of transcriptionally engineered *Pichia pastoris* strains with $P_{AOX1-wt}$ with different carbon sources at t = 20 h. Mean values of 3 different clones are presented. Error bars represent the standard deviation (±). Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol. Detailed information including statistical analysis is presented in Table 4.17.

Methanol induced eGFP synthesis in wild-type *P. pastoris* with $P_{AOX1-wt}$ was considered as 100 units and expression measurments were correlated accordingly. For simplicity, methanol induced wild-type *P. pastoris* with $P_{AOX1-wt}$ will be mentioned as "the wild-type strain".

Table 4.17 Normalized eGFP synthesis capacity of TF-OE and TF-KO *P. pastoris* strains with $P_{AOX1\text{-}wt}$ relative to methanol grown wild-type *P. pastoris* strain carrying $P_{AOX1\text{-}wt}$ at the cultivation time of t = 20 h. Mean values of 3 different clones are presented. Significance of differences to the control strain was calculated with Student's t-test. (* p-values <0.05; ** p-values <0.01).Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol

Normalized eGFP concentrations in *P. pastoris*					
TECs	**E**	**M**	**X**	**G**	**D**
AOX1-wt	11±1	100±4	4±0	2±0	3±0
AOX1-wt Aca1-OE	13±3	153±4**	2±1	3±1	4±1
AOX1-wt Adr1-OE	15±4	345±1**	14±3*	23±8*	6±3
AOX1-wt Cat8-1-OE	12±1	118±7*	4±0	3±1	3±0
AOX1-wt Cat8-2-OE	14±1**	118±2**	3±1	3±1	3±0
AOX1-wt *adr1*Δ	3±0**	8±0**	2±1	2±0	3±1
AOX1-wt *cat8-1*Δ	22±1**	117±1**	4±1	2±1	3±0
AOX1-wt *cat8-2*Δ	21±0**	132±5**	5±1	2±0	3±0

Overexpression of all engineered TFs increased the $P_{AOX1\text{-}wt}$ expression in methanol while the highest progress was observed with Adr1(Mxr1) overexpression. Under derepressed conditions, only Adr1-OE increased the strength of P_{AOX1}. Besides, in excess glycerol only the TECs transformed with Adr1-OE reached to 23% expression strength of the wild-type strain. Deletion of *ADR1* gene caused total loss of $P_{AOX1\text{-}wt}$ activity in all the fermentations carried out with different carbon sources. As described in section 4.4.1, deletion of *ADR1* decreased $P_{ADH2\text{-}wt}$ expression strength to 67% of the ethanol induced wild-type *P. pastoris* with $P_{ADH2\text{-}wt}$; while, in the other carbon sources $P_{ADH2\text{-}wt}$ expression levels increased in the *adr1*Δ cells. However in *S. cerevisiae adr1*Δ mutant, P_{ADH2} activity reduced to 19.5% of the wild-type strain (Walther and Schüller, 2001). Thus, Adr1 was annotated as alcohol dehydrogenase 2 synthesis regulator. In *P. pastoris* Adr1is the main activator of P_{AOX1} in methanol; therefore, its annotation as "methanol expression regulator 1" (Mxr1) clearly defines its role in *P. pastoris* accurately.

The TEC transformed with overexpression of the TF Aca1 increased P_{AOX1} activity to 153% of the wild-type strain, although no Aca1 and Aca2 binding motif was identified on $P_{AOX1\text{-}wt}$ by *in silico* analysis of putative TFBSs by MatInspector. This shows that Aca1 may have an indirect effect on activation of P_{AOX1}; where, can interact with other DNA-bound TFs to increase transcriptional activation or, *P. pastoris* Aca1 may have a different DNA binding motif. The function of Aca2 was identified as a TF for non-optimal carbon source utilization in *S. cerevisiae*, where, *aca2*Δ mutants showed delayed or no growth on respiratory carbon sources such as ethanol, glycerol, and raffinose (Garcia-Gimeno and Struhl, 2000). However, the influence of the Aca1 deletion was very detrimental and no viable *P. pastoris* cell was obtained after the transformation, which allows to conclude that Aca1 can be an essential gene for *P. pastoris*. However, the promoter- and transcriptional- engineering results conclusively demonstrated that the yet uncharacterized *P. pastoris* Aca1 TF acts as an activator on the regulation of P_{AOX1} during methanol utilization. The data obtained with the quantification of mRNA transcription levels is supporting the result. An 8-fold increase was measured in the transcription of *ACA1* in the methanol grown *adr1*Δ mutants, while no change was observed with the ethanol-grown cells. The results indicate that Aca1 contributes to the activation of Adr1-dependent genes in the methanol grown *P. pastoris*. Since the widely used carbon source of *P. pastoris* is methanol, it is plausible for Aca1 to be active on the MUT pathway genes.

In the methanol-grown cells, overexpression of Cat8-1 and Cat8-2 increased the P_{AOX1} activity to 118% of the wild-type strain, while there is no change in ethanol. Deletions of *CAT8-1* and *CAT8-2* have prominent effects on *P. pastoris* $P_{AOX\text{-}wt}$. In methanol and ethanol-grown *cat8-1*Δ mutant strains, P_{AOX1} activity increased, respectively, to 117% and 22% of the wild-type strains. In methanol and ethanol-grown *cat8-2*Δ mutant strains, P_{AOX1} activity increased, respectively, to 132% and 21% of the wild-type strain. This shows that when there is only one Cat8-protein, can be either Cat8-1 or Cat8-2, activation of P_{AOX1} becomes stronger. Similar results were also observed with P_{ADH2}.

The mTRs of *CAT8-1* and *CAT8-2* in *adr1*Δ mutant grown in methanol were high and calculated as 3.3 and 24.0, respectively. Contrarily, no difference was observed in the ethanol-grown *adr1*Δ mutants. These results underpin the function of Cat8-1 and Cat8-2 in methanol metabolism and P_{AOX1} regulation although no putative Cat8 binding motif was identified on P_{AOX1} by *in silico* analysis.

Table 4.18 The cell concentrations of *P. pastoris* TF-OE and TF-KO strains at t = 20 h, in terms of OD_{600} values. Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol

The cell concentrations in terms of OD_{600} values					
TECs	**E**	**M**	**X**	**G**	**D**
AOX1-wt	2.8	3.2	10.8	2.1	1.8
AOX1-wt Aca1-OE	2.5	2.0	10.6	1.7	1.4
AOX1-wt Adr1-OE	2.6	1.8	10.6	1.8	1.6
AOX1-wt Cat8-1-OE	2.7	2.4	10.4	2.1	1.7
AOX1-wt Cat8-2-OE	2.5	2.6	11.2	1.4	1.5
AOX1-wt *adr1*Δ	1.8	1.0	1.8	0.4	0.3
AOX1-wt *cat8-1*Δ	1.5	3.3	10.6	2.1	1.8
AOX1-wt *cat8-2*Δ	2.3	2.5	11.0	2.1	1.5

The TECs of *P. pastoris* under $P_{AOX1\text{-}wt}$ transformed with TF-OE and TF-KO demonstrated similar cell growth with the TECs under $P_{ADH2\text{-}wt}$ transformed with TF-OE and TF-KO; and the OD_{600} values of TECs are given in Table 4.18.

Effects of Adr1(Mxr1), Aca1, Cat8-1, and Cat8-2 on $P_{AOX1\text{-}mod}$ were investigated further by the TECs transformed with TF-OE and TF-KO casettes. Screening experiments were performed using five different carbon sources, and the eGFP synthesis levels of the TECs with $P_{AOX1\text{-}mod}$ in different TF-OE and TF-KO strains are summarized in Figure 4.28 and Table 4.19.

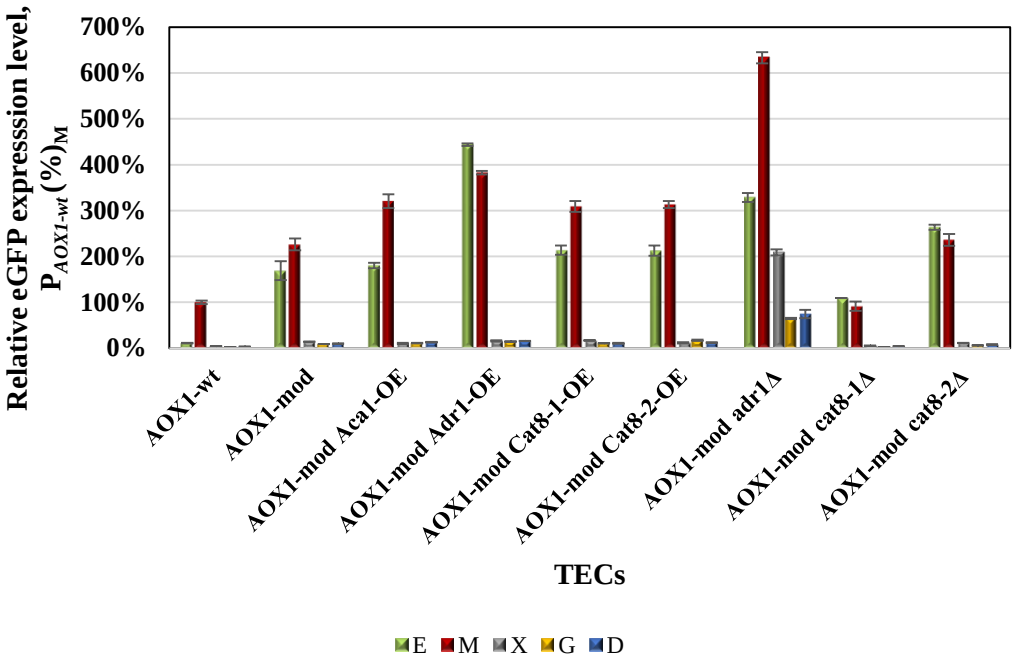


Figure 4. 28 Normalized eGFP synthesis capacity of transcriptionally engineered *Pichia pastoris* strains with $P_{AOX1\text{-}mod}$ with different carbon sources at t = 20 h. Mean values of 3 different clones are presented. Error bars represent the standard deviation (±). Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol. Detailed information including statistical analysis is presented in Table 4.19.

In ethanol and methanol containing minimal media, expression levels of $P_{AOX1\text{-}mod}$ reached 169% and 226% of the $P_{AOX\text{-}wt}$ without any TF overexpression or deletion transformations. Among the TECs, overexpression of all the TFs increased the activity of $P_{AOX1\text{-}mod}$ in methanol and ethanol; which is different than that of the TECs constructed with $P_{AOX1\text{-}wt}$ with an increased activity only in methanol.

Table 4.19 Normalized eGFP synthesis capacity of TF-OE and TF-KO *P. pastoris* strains with $P_{AOX1\text{-}mod}$ relative to methanol grown wild-type *P. pastoris* strain carrying $P_{AOX1\text{-}wt}$ at the cultivation time of t = 20 h. Mean values of 3 different clones are presented. Significance of differences to the control strain was calculated with Student's t-test. (* p-values <0.05; ** p-values <0.01).Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol

Normalized eGFP concentrations in *P. pastoris*					
TECs	**E**	**M**	**X**	**G**	**D**
AOX1-wt	11±1	100±4	4±0	2±0	3±0
AOX1-mod	169±21	226±13	14±0	9±0	10±0
AOX1-mod Aca1-OE	180±6	321±15**	10±1*	11±1	13±1*
AOX1-mod Adr1-OE	444±3**	383±4**	16±1**	15±0	16±0**
AOX1-mod Cat8-1-OE	214±10*	309±12**	17±1*	11±0	11±1
AOX1-mod Cat8-2-OE	213±11*	313±8**	12±1	17±1**	12±1
AOX1-mod *adr1*Δ	329±10**	633±12**	209±7**	65±1**	75±9**
AOX1-mod *cat8-1*Δ	109±0*	91±10**	5±0**	3±0**	4±0**
AOX1-mod *cat8-2*Δ	263±6**	236±13	11±0**	6±0**	8±1

The most significant effect was observed with Adr1-OE and Adr1-KO modifications, which is a high expression level obtained with $P_{AOX1\text{-}mod}$ in Adr1-OE strain in ethanol that increased the expression to 444% of the methanol-induced $P_{AOX1\text{-}wt}$ in the wild-type strain. Adr1 overexpression increased $P_{AOX1\text{-}mod}$ activity also in methanol from 226% to 383% of the wild-type. The highest expression level with $P_{AOX1\text{-}mod}$ was obtained in the methanol-induced *adr1*Δ strain that reached 633% of the activity of the wild-type strain; however, the cells cannot grow in methanol containing minimal media. In the *adr1*Δ strain, the expression strength of $P_{AOX1\text{-}mod}$ also increased to 329% of the wild-type in the ethanol-grown cell. The activity of the *adr1*Δ TECs with P_{ADH2} and P_{AOX1} reduced in ethanol, is indeed noteworthy. Only the *adr1*Δ strain exhibited an increased production level with $P_{AOX1\text{-}mod}$ under derepression conditions. 209% and 65% of the wild-type strain expressions levels was reached with the limited glucose and excess glycerol, respectively. However, in the *adr1*Δ strain, $P_{AOX1\text{-}mod}$ lost its

glucose repressebility. These results demonstrate that regulation of $P_{AOX1\text{-}mod}$ seems like P_{ADH2} rather than P_{AOX1}. Introduction of novel *cis*-acting elements to P_{AOX1} created not only stronger NEPVs but also influenced the regulation patterns.

In methanol, overexpression of Aca1 increased the activity of $P_{AOX1\text{-}mod}$ from 226% to 312% of the wild-type. With Aca1-OE strain, not a significant difference was observed with the other carbon sources. Overexpression of Cat8-1 and Cat8-2 contributed to $P_{AOX1\text{-}mod}$ activation in a similar way. In ethanol-grown Cat8-1-OE and Cat8-2-OE strains, $P_{AOX1\text{-}mod}$ reached to 214% and 213% of the expression of the wild-type. Whereas, with the same strains in methanol, $P_{AOX1\text{-}mod}$ reached 309% and 313% of the expression of the wild-type.

Deletion of the Cat8-1 resulted in a reduction in $P_{AOX1\text{-}mod}$ activity with all the carbon sources. In response to ethanol, Cat8-1 deletion reduced expression with $P_{AOX1\text{-}mod}$ from 169% to 109%, whereas in response to methanol, the expression with $P_{AOX1\text{-}mod}$ reduced from 226% to 91%. However, the deletion of Cat8-2 did not affect the $P_{AOX1\text{-}mod}$ activity in methanol. Even in ethanol $P_{AOX1\text{-}mod}$ showed increased performance in the *cat8-2*Δ mutant. These results demonstrated that in ethanol and methanol, Cat8-1 is sufficient and superior to Cat8-2 for the activation of $P_{AOX1\text{-}mod}$.

Transcriptional engineering of $P_{AOX1\text{-}wt}$ and $P_{AOX1\text{-}mod}$ generated novel r-protein expression systems allow more efficient r-protein production in ethanol and methanol as well as r-protein expression by derepression systems. The highest methanol induction capacity with $P_{AOX1\text{-}mod}$ was achieved in Adr1-OE TEC, eGFP expression level reached 345% of methanol induced wild-type *P. pastoris* with $P_{AOX1\text{-}wt}$.

The TFs Aca1, Cat8-1 and Cat8-2, which are not characterized yet, enhanced the expression with $P_{AOX1\text{-}wt}$, respectively, by 1.53-, 1.18-, and 1.18- fold. Transcriptional engineering with deletions revealed that the TF Cat8-1 could efficiently activates $P_{AOX1\text{-}wt}$ in methanol. Overexpression of Adr1 created derepression effect to produce r-proteins at limited glucose and excess glycerol. However, the productivity of $P_{AOX1\text{-}}$

$_{wt}$ in Adr1 OE strain by derepression was still limited, only reaching to 23% of the methanol-induced activity of wild-type strain.

The transcriptionally engineered $P_{AOX1\text{-}mod}$ characteristics is different from the characteristics of the $P_{AOX1\text{-}wt}$. This shows that introduction of novel *cis*-acting elements into the P_{AOX1} generated the NEPVs which have different transcriptional strengths and regulations. The highest expression level in methanol was obtained in *adr1*Δ strain, in which $P_{AOX1\text{-}mod}$ activity reached 633% of the expression with the $P_{AOX1\text{-}wt}$ in the wild-type strain. *P. pastoris adr1*Δ strain could not grow in methanol; but our results reveal that methanol was consumed through the less-efficient pathways and the cell metabolism was operative at a sustainable level. In ethanol, the highest expression level was obtained with Adr1-OE system, and $P_{AOX1\text{-}mod}$ expression was enhanced by 4.44-fold of the $P_{AOX1\text{-}wt}$ in methanol. Also, the TEC constructed with $P_{AOX1\text{-}mod}$ by *adr1*Δ demonstrated highly derepressed eGFP synthesis in limited glucose with a 2.09-fold enhancement than that of the cells in methanol constructed with $P_{AOX1\text{-}wt}$. Overexpression of Cat8-1 and Cat8-2 enhanced the strength of $P_{AOX1\text{-}mod}$ in the ethanol and methanol-grown cells. $P_{AOX1\text{-}mod}$ enhanced the eGFP expression 2.14-fold and 3.13-fold in ethanol and methanol, respectively, of the expression capacity of $P_{AOX1\text{-}wt}$ in methanol. Although Cat8-1 and Cat8-2-OE affected the TEC in a similar manner, the results indicates that Cat8-1-OE has a more pronounced effect on $P_{AOX1\text{-}mod}$.

4.4.3 Transcription Levels of the Transcription Factors in Overexpression and Knock-out Strains

In order to evaluate the mRNA expression levels of *ADR1*(*MXR1*), *ACA1*, *CAT8-1* and *CAT8-2* quantitative PCR was used. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol- or methanol- grown wild-type *P. pastoris* cells.

4.4.3.1 Transcription Levels of the TFs in Overexpression Strains

In order to gain more insight on the transcriptional engineering of the cells, mRNA transcription ratios (mTRs) of the TF genes in the overexpression strains were measured in the TECs grown in ethanol and methanol. The mTRs of *ACA1* are given in Figure 4.29 and Table 4.20.

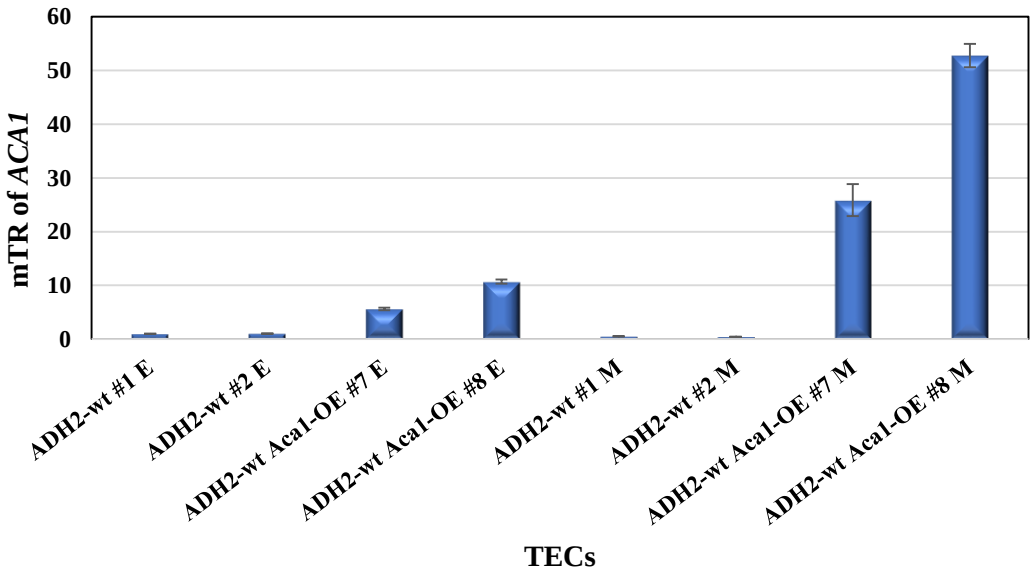


Figure 4. 29 mRNA transcription ratios (mTRs) of *ACA1* in *P. pastoris* TF-OE strains constructed with $P_{ADH2\text{-}wt}$ at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. The error bars represent the standard deviation of three technical replicates (±). Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v)

Table 4.20 mRNA transcription ratios (mTRs) of *ACA1* in *P. pastoris* TF-OE strains constructed with $P_{ADH2\text{-}wt}$ at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v). StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	**Carbon Source**	**mTR of *ACA1***	**StDev**
ADH2-wt #1	E	1.0	0.0
ADH2-wt #2	E	1.0	0.1
ADH2-wt Aca1-OE #7	E	5.6	0.2
ADH2-wt Aca1-OE #8	E	10.7	0.4
ADH2-wt #1	M	0.5	0.0
ADH2-wt #2	M	0.4	0.0
ADH2-wt Aca1-OE #7	M	25.9	3.0
ADH2-wt Aca1-OE #8	M	52.8	2.2

In the ethanol-grown wild-type *P. pastoris,* the mTR of *ACA1* was ca. 2-fold higher than in the methanol-grown cells. In the ethanol-grown overexpression strains, mTRs of *ACA1* increased 5- to 10-fold, whereas in methanol increased 25- to 52-fold, with changes from clone to clone.

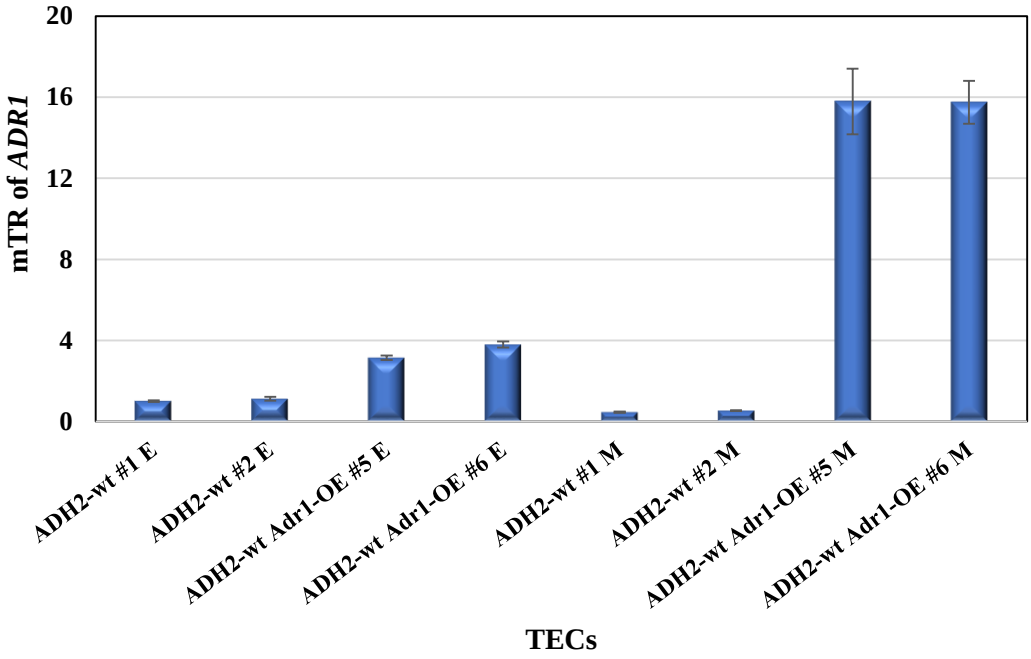


Figure 4.30 mRNA transcription ratios (mTRs) of *ADR1* in *P. pastoris* TF-OE strains constructed with $P_{ADH2\text{-}wt}$ at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. The error bars represent the standard deviation of three technical replicates (±). Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v)

Table 4.21 mRNA transcription ratios (mTRs) of *ADR1* in *P. pastoris* TF-OE strains constructed with $P_{ADH2\text{-}wt}$ at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v). StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	**Carbon Source**	**mTR of *ADR1***	**StDev**
ADH2-wt #1	E	1.0	0.0
ADH2-wt #2	E	1.1	0.1
ADH2-wt Adr1-OE #5	E	3.1	0.1
ADH2-wt Adr1-OE #6	E	3.8	0.2
ADH2-wt #1	M	0.5	0.0
ADH2-wt #2	M	0.5	0.0
ADH2-wt Adr1-OE #5	M	15.8	1.6
ADH2-wt Adr1-OE #6	M	15.8	1.1

The mRNA transcription ratios (mTRs) of *ADR1* are given in Figure 4.30 and Table 4.21. In the ethanol-grown wild-type *P. pastoris*, mTR of *ADR1* was 2-fold higher than in the methanol-grown cells, similar to mTRs of *ACA1*. In the ethanol-grown overexpression strains mTRs of *ADR1* increased ca. 3.5-fold, whereas in methanol mTRs of *ADR1* increased ca. 16-fold.

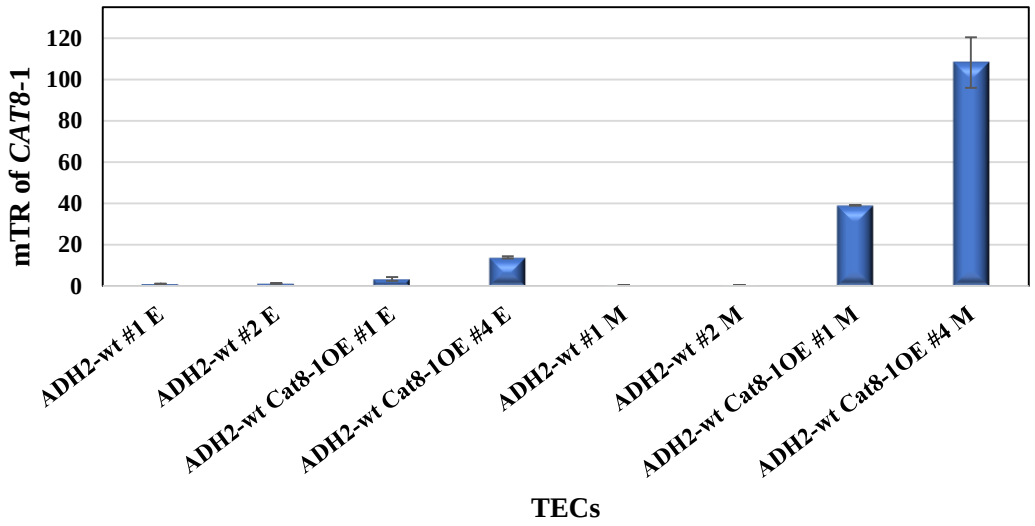


Figure 4.31 mRNA transcription ratios (mTRs) of *CAT8-1* in *P. pastoris* TF-OE strains constructed with $P_{ADH2\text{-}wt}$ at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. The error bars represent the standard deviation of three technical replicates (±). Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v)

Table 4.22 mRNA transcription ratios (mTRs) of *CAT8-1* in *P. pastoris* TF-OE strains constructed with $P_{ADH2\text{-}wt}$ at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v). StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	**Carbon Source**	**mTR of *CAT8-1***	**StDev**
ADH2-wt #1	E	1.0	0,1
ADH2-wt #2	E	1.2	0,1
ADH2-wt Cat8-1OE #1	E	3.4	1.0
ADH2-wt Cat8-1OE #4	E	13.8	0.5
ADH2-wt #1	M	0.4	0.0
ADH2-wt #2	M	0.4	0.0
ADH2-wt Cat8-1OE #1	M	38.9	0.0
ADH2-wt Cat8-1OE #4	M	108.1	12.2

The mTRs of *CAT8-1* are given in Figure 4.31 and Table 4.22. In the ethanol-grown wild-type *P. pastoris*, the mTR of *CAT8-1* was ca. 2-fold higher than in the methanol-grown cells. In the ethanol-grown overexpression strains mTR of *CAT8-1* increased between 3.4- to 13.8-fold clone-to-clone, whereas in methanol *CAT8-1* mTRs increased ca. 39- to 108-fold.

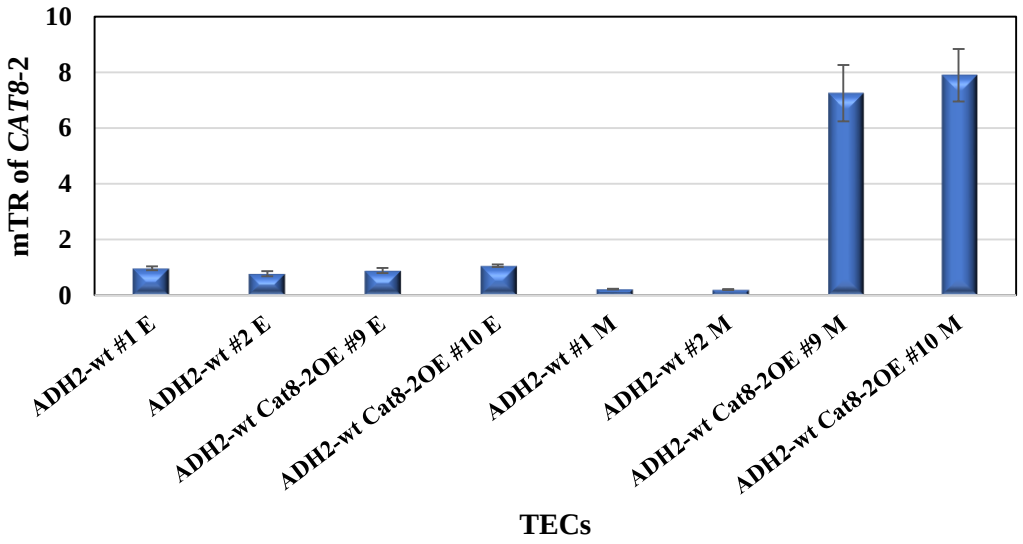


Figure 4.32 mRNA transcription ratios (mTRs) of *CAT8-2* in *P. pastoris* TF-OE strains constructed with $P_{ADH2\text{-}wt}$ at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. The error bars represent the standard deviation of three technical replicates (±). Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v)

Table 4. 23 mRNA transcription ratios (mTRs) of *CAT8-2* in *P. pastoris* TF-OE strains constructed with $P_{ADH2\text{-}wt}$ at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v). StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	Carbon Source	mTR of *CAT8-2*	StDev
ADH2-wt #1	E	0.97	0.07
ADH2-wt #2	E	0.77	0.09
ADH2-wt Cat8-2OE #9	E	0.88	0.09
ADH2-wt Cat8-2OE #10	E	1.06	0.04
ADH2-wt #1	M	0.22	0.01
ADH2-wt #2	M	0.20	0.02
ADH2-wt Cat8-2OE #9	M	7.26	1.01
ADH2-wt Cat8-2OE #10	M	7.90	0.94

The mTRs of *CAT8-2* are given in Figure 4.32 and Table 4.23. In the ethanol-grown wild-type *P. pastoris*, mTRs of *CAT8-2* was ca. 4-fold higher than in the methanol-grown cells. In the ethanol-grown overexpression strains no significant increase in the mTRs of *CAT8-2* was observed; however, in methanol mTRs of *CAT8-2* increased ca. 8-fold.

All targeted transcription factors were overexpressed less in the ethanol grown TECs compared with that of in methanol. The mTRs of *CAT8-2* gene in the OE strains of the TECs in ethanol could not be increased. This result states that the transcription of *CAT8-2* gene is naturally high in the cells; consequently, in the TEC constructed with P_{THI11} an increase in transcription was not observed. The mTRs of *ACA1*, *ADR1* and *CAT8-1* increased 8-, 3.5-, and 8-fold, respectively, in the ethanol-grown OE strains. The mTRs of *ACA1*, *ADR1*, *CAT8-1*, and *CAT8-2* -OE strains in methanol increased 39-, 16-, 73-, and 7-fold, respectively.

Due to the differences in transcription levels in the methanol and ethanol-grown TF-OE strains constructed with P_{THI11}, the expression capacity of the TECs were tested with fermentations using different carbon sources. The mTRs of *THI11* are given in Figure 4.33 and Table 4.24; and compared with the wild-type *P. pastoris* in ethanol.

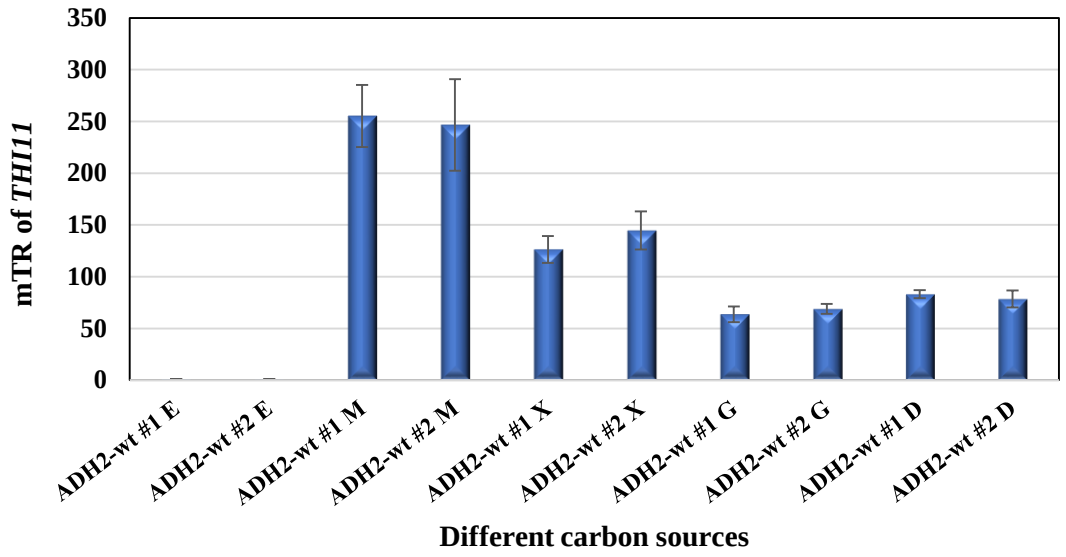


Figure 4.33 mRNA transcription ratios (mTRs) of *THI11* in *P. pastoris* strains grown in different carbon sources at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented is relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. Error bars represent the standard deviation among three technical replicates (±). Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol

Table 4.24 mRNA transcription ratios (mTRs) of *THI11* in *P. pastoris* strains grown in different carbon sources at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol. StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	Carbon Source	mTR of *THI11*	StDev
ADH2-wt #1	E	1.0	0.1
ADH2-wt #2	E	0.8	0.0
ADH2-wt #1	M	255.1	30.0
ADH2-wt #2	M	246.5	44.2
ADH2-wt #1	X	126.3	13.0
ADH2-wt #2	X	144.6	18.4
ADH2-wt #1	G	63.8	7.6
ADH2-wt #2	G	68.9	4.7
ADH2-wt #1	D	83.0	3.9
ADH2-wt #2	D	78.4	8.1

Thiamine was reported to be the only regulator of P_{THI11}; where, P_{THI11} is totally repressed in the presence of thiamine and activated in its absence (Delic et al., 2013; Landes et al., 2016). When *P. pastoris* cells were cultivated in five different carbon sources, their *THI11* transcript levels showed substantial differences. According to qPCR results, P_{THI11} showed a carbon-source dependent regulation as well. P_{THI11} was very active in the methanol-grown cells. Its activity gradually decreased in limited glucose, excess glycerol and excess glucose. The lowest activity was observed in the ethanol-grown cells.

Thiamine pyrophosphate (TPP), the active derivative of thiamine, is the co-factor of decarboxylases, transketolases, and phosphoketolases. In *P. pastoris* induction of the MUT pathway results in an intracellular thiamine deficiency which is caused by a strong induction of TPP-containing enzymes DAS1/2 (Rußmayer et al., 2015). *P. pastoris* pyruvate decarboxylase is also a TPP-containing enzyme that catalyzes

decarboxylation of pyruvate to acetaldehyde for ethanol production. According to our results, the mTRs of *THI11* indicate that in *P. pastoris* ethanol utilization pathway enzymes are not dependent on TPP; and, thus THI11 promoter does not have a high activity in the ethanol grown cells.

Landes et al. (2016) reported on recombinant HSA production with P_{THI11} resulted in different transcription profiles of *HSA* and *THI11* in response to the presence and uptake of extracellular thiamine. The potential occurrence of a riboswitch was proposed as a reason for the regulation differences between the native *THI11* and r-protein genes (Landes et al., 2016), which is concordant with our findings. In this study, P_{THI11} exhibited different transcription profiles for *THI11* and TF genes in different carbon sources.

4.4.3.2 Transcription Levels of the TFs in Different TF Knock-out Strains

Gene deletions of the specified transcription factors influenced the regulations in the TECs. Some of the TF deletions improved the strength of $P_{ADH2\text{-}v}$ and $P_{AOX1\text{-}v}$. In order to gain a deeper insight, transcription levels of the TF genes were studied in TF knock-out strains. In order to calculate the mTRs, the transcription levels of the TF genes were normalized with the transcription level of the *Actin* gene; and then divided by the normalized transcription values of of the wild-type *P. pastoris* cells in ethanol. The mTRs of *ACA1* in the ethanol and methanol-grown TF knock-out *P. pastoris* strains are given in Figures 4.34 and 4.35, and Table 4.25 and 4.26.

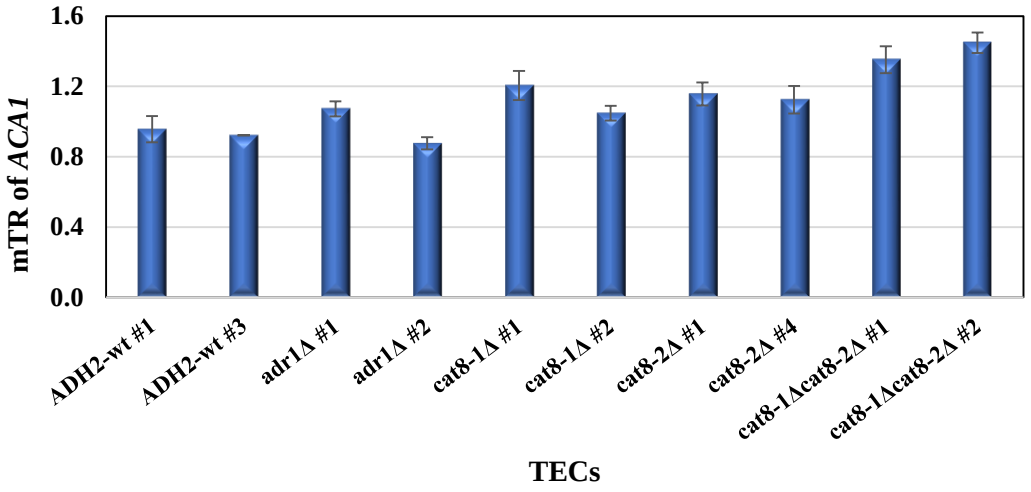


Figure 4.34 mRNA transcription ratios (mTRs) of *ACA1* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 2% (v/v) ethanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. The error bars represent the standard deviation of three technical replicates (±).

Table 4. 25 mRNA transcription ratios (mTRs) of *ACA1* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 2% (v/v) ethanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	**mTR of *ACA1***	**StDev**
ADH2-wt #1	0.96	0.07
ADH2-wt #3	0.92	0.00
***adr1*Δ #1**	1.07	0.04
***adr1*Δ #2**	0.88	0.04
***cat8-1*Δ #1**	1.20	0.08
***cat8-1*Δ #2**	1.05	0.04
***cat8-2*Δ #1**	1.16	0.07
***cat8-2*Δ #4**	1.12	0.08
***cat8-1*Δ*cat8-2*Δ #1**	1.35	0.08
***cat8-1*Δ*cat8-2*Δ #2**	1.45	0.06

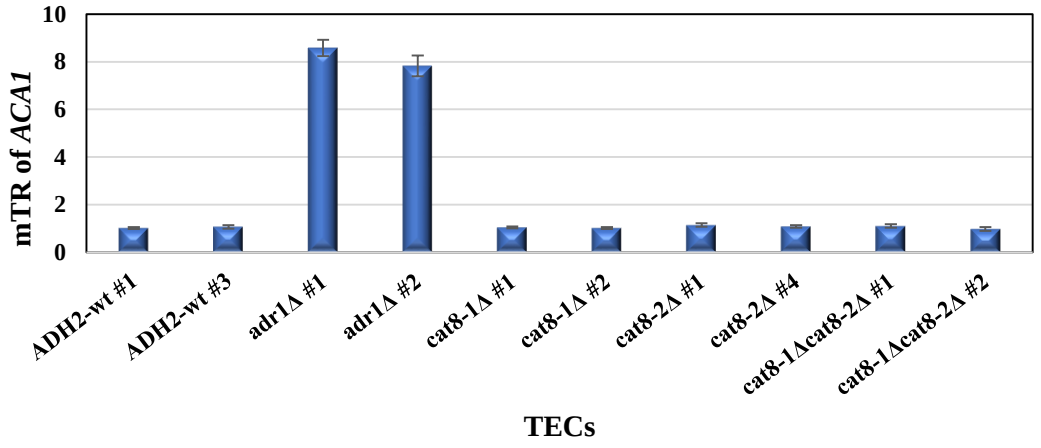


Figure 4.35 mRNA transcription ratios (mTRs) of *ACA1* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 1% (v/v) methanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented is relative to methanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. The error bars represent the standard deviation of three technical replicates (±).

Table 4.26 mRNA transcription ratios (mTRs) of *ACA1* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 1% (v/v) methanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to methanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	**mTR of *ACA1***	**StDev**
ADH2-wt #1	1.02	0.04
ADH2-wt #3	1.07	0.07
***adr1*Δ #1**	8.58	0.35
***adr1*Δ #2**	7.83	0.44
***cat8-1*Δ #1**	1.05	0.04
***cat8-1*Δ #2**	1.02	0.04
***cat8-2*Δ #1**	1.14	0.07
***cat8-2*Δ #4**	1.08	0.05
***cat8-1*Δ*cat8-2*Δ #1**	1.10	0.07
***cat8-1*Δ*cat8-2*Δ #2**	0.98	0.08

The mTRs of *ACA1* in *P. pastoris adr1*Δ, *cat8-1*Δ, *cat8-2*Δ and *cat8-1*Δ*cat8-2*Δ strains were determined. When the cells were grown in ethanol, *ACA1* transcription in the TF-KO strains were almost same. However, when the cells were grown in methanol a significant increase in mTR was observed in *P. pastoris adr1*Δ strain. After deletion of *ADR1*, exposure in methanol increased *ACA1* transcription ca. 8-fold.

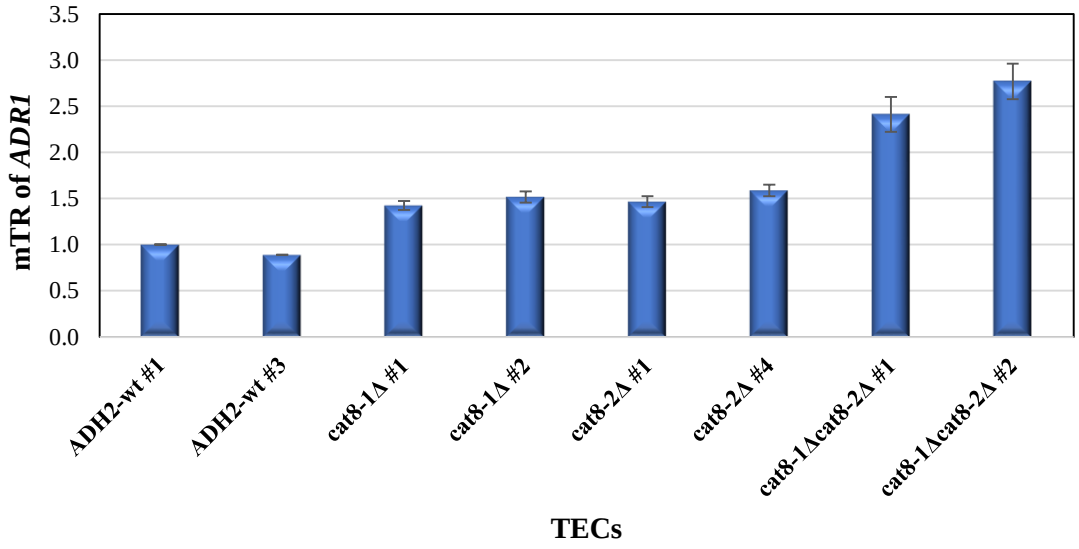


Figure 4. 36 mRNA transcription ratios (mTRs) of *ADR1* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 2% (v/v) ethanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. The error bars represent the standard deviation of three technical replicates (±).

Table 4. 27 mRNA transcription ratios (mTRs) of *ADR1* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 2% (v/v) ethanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	mTR of *ADR1*	StDev
ADH2-wt #1	1.00	0.00
ADH2-wt #3	0.89	0.00
***cat8-1Δ* #1**	1.42	0.05
***cat8-1Δ* #2**	1.52	0.06
***cat8-2Δ* #1**	1.46	0.06
***cat8-2Δ* #4**	1.59	0.06
***cat8-1Δcat8-2Δ* #1**	2.41	0.19
***cat8-1Δcat8-2Δ* #2**	2.77	0.19

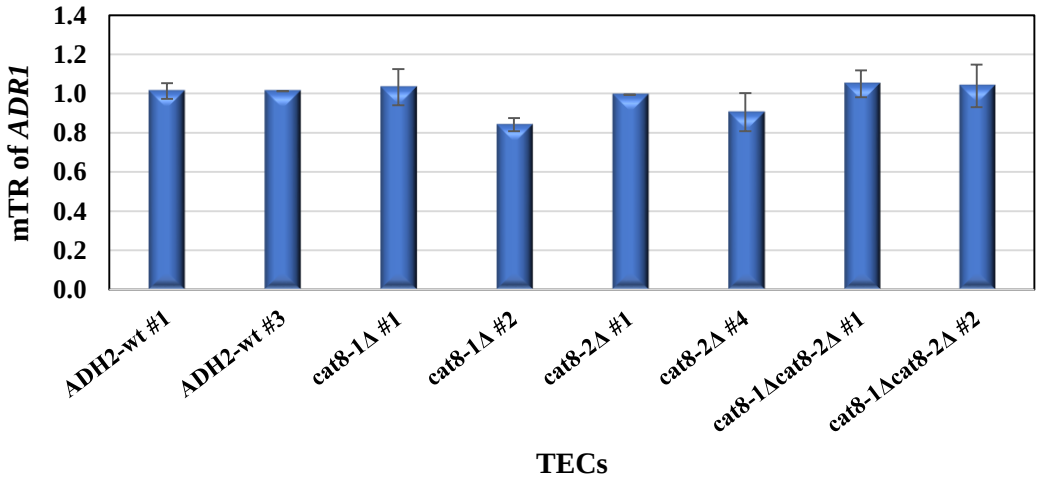


Figure 4.37 mRNA transcription ratios (mTRs) of *ADR1* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 1% (v/v) methanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to methanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. The error bars represent the standard deviation of three technical replicates (±).

Table 4. 28 mRNA transcription ratios (mTRs) of *ADR1* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 1% (v/v) methanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to methanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	**mTR of *ADR1***	**StDev**
ADH2-wt #1	1.0	0.0
ADH2-wt #3	1.0	0.0
***cat8-1*Δ #1**	1.0	0.1
***cat8-1*Δ #2**	0.8	0.0
***cat8-2*Δ #1**	1.0	0.0
***cat8-2*Δ #4**	0.9	0.1
***cat8-1*Δ*cat8-2*Δ #1**	1.1	0.1
***cat8-1*Δ*cat8-2*Δ #2**	1.0	0.1

The mTRs of *ADR1* in the ethanol and methanol-grown TF knock-out *P. pastoris* strains are given in Figure 4.36 and 4.37, and Table 4.27 and 4.28. mTRs of *ADR1* were determined in *P. pastoris cat8-1*Δ, *cat8-2*Δ, and *cat8-1*Δ*cat8-2*Δ strains. In the ethanol-grown *P. pastoris cat8-1*Δ, *cat8-2*Δ and *cat8-1*Δ*cat8-2*Δ strains, mTR of *ADR1* increased 1.54-, 1.60-, and 2.72 fold, respectively. When the cells were grown in methanol, *ADR1* transcriptions in the TF-KO strains were close to each other.

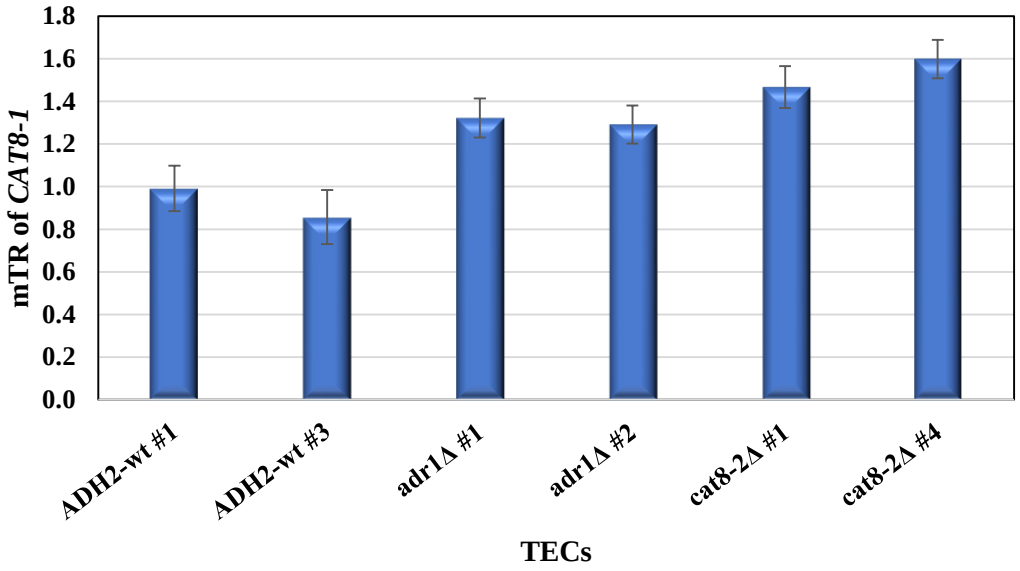


Figure 4.38 mRNA transcription ratios (mTRs) of *CAT8-1* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 2% (v/v) ethanol containing minimal media at t = 20 h. Transcript levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. The error bars represent the standard deviation of three technical replicates (±).

Table 4.29 mRNA transcription ratios (mTRs) of *CAT8-1* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 2% (v/v) ethanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	**mTR of *CAT8-1***	**StDev**
ADH2-wt #1	0.99	0.11
ADH2-wt #3	0.86	0.13
***adr1*Δ #1**	1.32	0.09
***adr1*Δ #2**	1.29	0.09
***cat8-2*Δ #1**	1.47	0.10
***cat8-2*Δ #4**	1.60	0.09

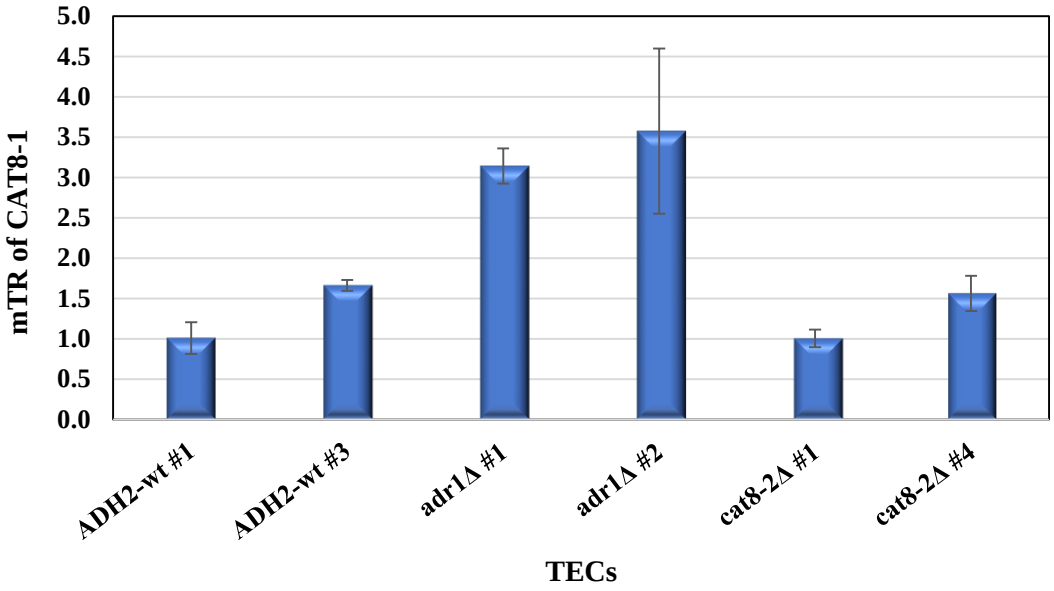


Figure 4. 39 mRNA transcription ratios (mTRs) of *CAT8-1* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 1% (v/v) methanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to methanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. The error bars represent the standard deviation of three technical replicates (±).

Table 4.30 mRNA transcription ratios (mTRs) of *CAT8-1* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 1% (v/v) methanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to methanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	**mTR of *CAT8-1***	**StDev**
ADH2-wt #1	1.0	0.2
ADH2-wt #3	1.7	0.1
***adr1*Δ #1**	3.1	0.2
***adr1*Δ #2**	3.6	1.0
***cat8-2*Δ #1**	1.0	0.1
***cat8-2*Δ #4**	1.6	0.2

The mTRs of *CAT8-1* in the ethanol and methanol-grown TF knock-out *P. pastoris* strains are given in Figures 4.38 and 4.39, and Table 4.29 and 4.30. *CAT8-1* transcriptions were determined in *P. pastoris adr1*Δ and *cat8-2*Δ strains. In the ethanol grown *P. pastoris adr1*Δ and *cat8-2*Δ strains, *CAT8-1* transcription increased ca. 1.4-, and 1.65-fold, respectively. When the cells were grown in methanol, *CAT8-1* transcription increased ca. 3-fold in *adr1*Δ strain. But there was no significant difference in *cat8-2*Δ strain.

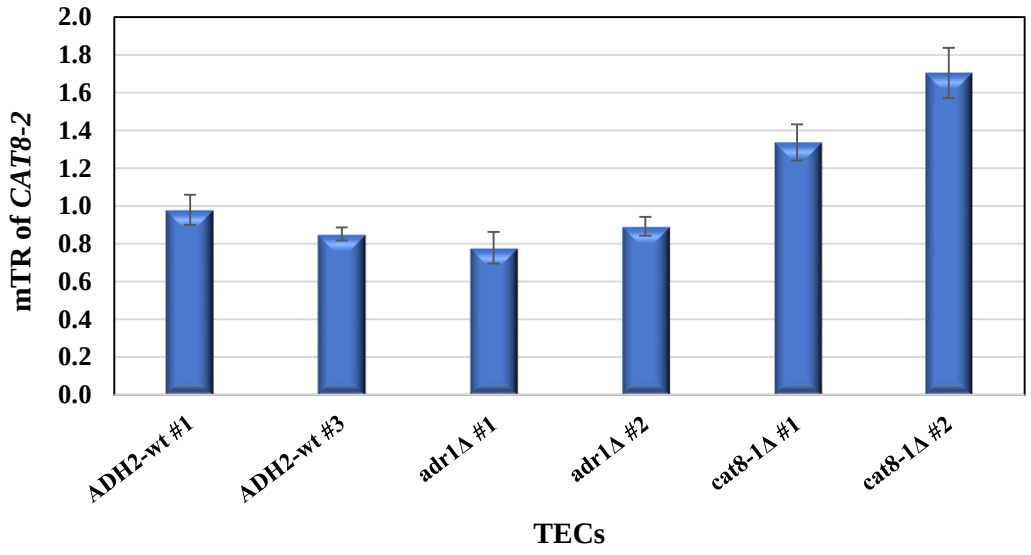


Figure 4.40 mRNA transcription ratios (mTRs) of *CAT8-2* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 2% (v/v) ethanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. The error bars represent the standard deviation of three technical replicates (±).

Table 4.31 mRNA transcription ratios (mTRs) of *CAT8-2* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 2% (v/v) ethanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	mTR of *CAT8-2*	StDev
ADH2-wt #1	1.0	0.1
ADH2-wt #3	0.9	0.0
***adr1*Δ #1**	0.8	0.1
***adr1*Δ #2**	0.9	0.1
***cat8-1*Δ #1**	1.3	0.1
***cat8-1*Δ #2**	1.7	0.1

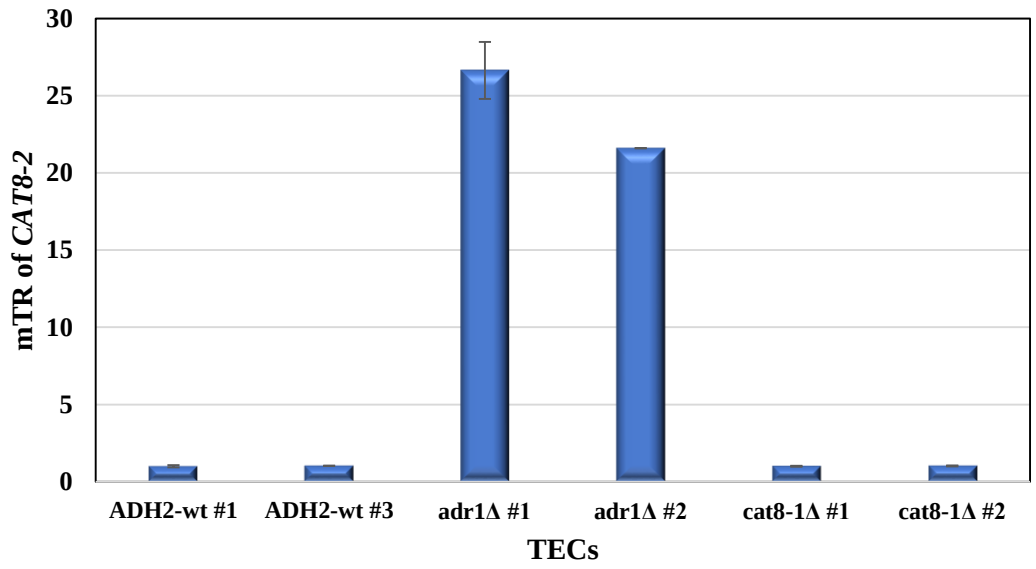


Figure 4.41 mRNA transcription ratios (mTRs) of *CAT8-2* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 1% (v/v) methanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to methanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. The error bars represent the standard deviation of three technical replicates (±).

Table 4. 32 mRNA transcription ratios (mTRs) of *CAT8-2* in *P. pastoris* TF-KO strains constructed with $P_{ADH2\text{-}wt}$ grown in 1% (v/v) methanol containing minimal media at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to methanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	**mTR of *CAT8-2***	**StDev**
ADH2-wt #1	1.0	0.1
ADH2-wt #3	1.0	0.0
***adr1*Δ #1**	26.6	1.9
***adr1*Δ #2**	21.6	0.0
***cat8-1*Δ #1**	1.0	0.0
***cat8-1*Δ #2**	1.0	0.0

The mTRs of *CAT8-2* in the ethanol and methanol-grown TF knock-out *P. pastoris* strains are given in Figure 4.40 and 4.41, and Table 4.31 and 4.32. The mTRs of *CAT8-2* in *P. pastoris adr1*Δ and *cat8-1*Δ strains were calculated. In the ethanol grown *P. pastoris cat8-1*Δ strains, mTR of *CAT8-2* gene increased ca. 1.66-fold. When the cells were grown in methanol, *CAT8-2* transcription increased ca. 24-fold in *adr1*Δ strain while there was not any significant difference in *cat8-1*Δ

Transcription results of Aca1, Adr1, Cat8-1, and Cat8-2 reveal that the TFs interfere the operation of the others in the cell. In the methanol grown adr1Δ mutant strain, transcription of *ACA1*, *CAT8-1*, and *CAT8-2* increased 8-, 3- and 24- fold respectively. In ethanol *ADR1* transcription of *cat8-1*Δ, *cat8-2*Δ, and *cat8-1*Δ*cat8-2*Δ mutants increased 1.54-, 1.60-, and 2.72-fold, respectively. Furthermore, a regulation between Cat8-1 and Cat8-2 was identified. Deletion of either Cat8-1 or Cat8-2, led to a ca. 50% increase in the transcription of the other, in the ethanol grown cells; but the same effect was not observed in the methanol grown cells. Although the transcriptional engineering results obtained with the knock-out strains clarified the

regulatory potential of the TFs to an extent, a genome-wide transcriptome data is required for gaining a deeper insight.

4.4.3.3 Regulation of the Transcription Factor Levels by Carbon Sources

In order to gain a deeper understanding about the interactions between the modified TFs and the engineered P_{ADH2} and P_{AOX1} variants, fermentation experiments with five carbon sources were carried out. The mTRs of Aca1, Adr1, Cat8-1, and Cat8-2 were determined in response to the carbon sources, as follows: *i)* 2% (v/v) ethanol, *ii)* 1% (v/v) methanol, *iii)* limited glucose, *iv)* excess glycerol, and *v)* excess glucose. The variations in the mTRs were determined based on the transcription of wild-type *P. pastoris* cells grown in ethanol.

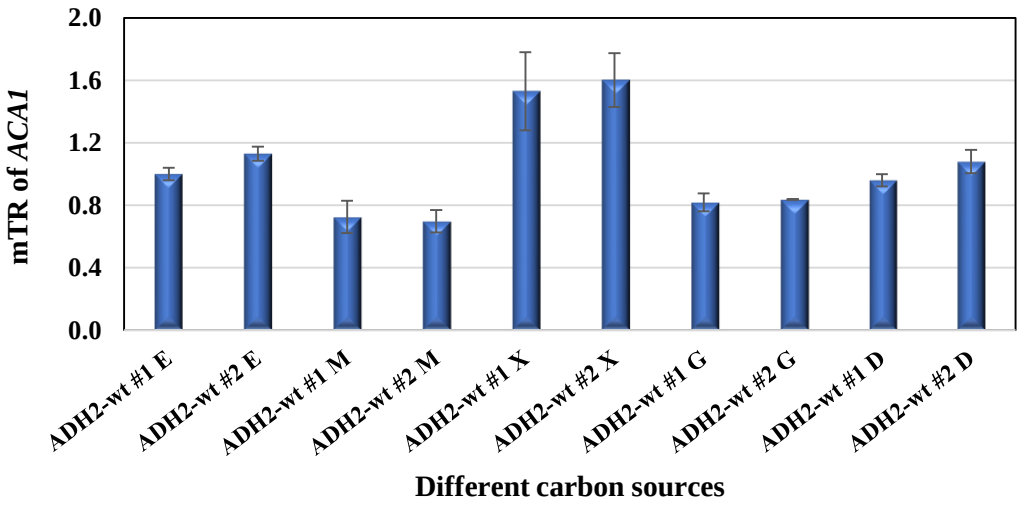


Figure 4.42 mRNA transcription ratios (mTRs) of *ACA1* in *P. pastoris* strains grown in different carbon sources at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. Error bars represent the standard deviation among three technical replicates (±). Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol.

Table 4.33 mRNA transcription ratios (mTRs) of *ACA1* in *P. pastoris* strains grown in different carbon sources at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol. StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	Carbon Source	mTR of *ACA1*	StDev
ADH2-wt #1	E	1.0	0.0
ADH2-wt #2	E	1.1	0.0
ADH2-wt #1	M	0.7	0.1
ADH2-wt #2	M	0.7	0.1
ADH2-wt #1	X	1.5	0.3
ADH2-wt #2	X	1.6	0.2
ADH2-wt #1	G	0.8	0.1
ADH2-wt #2	G	0.8	0.0
ADH2-wt #1	D	1.0	0.0
ADH2-wt #2	D	1.1	0.1

The variations in *ACA1* transcription on different carbon sources are given in Figure 4.42 and Table 4.33. *ACA1* was expressed in *P. pastoris* with robust transcriptions in response to the carbon sources. Among the tested conditions, the highest *ACA1* transcript level was determined with the limited glucose condition.

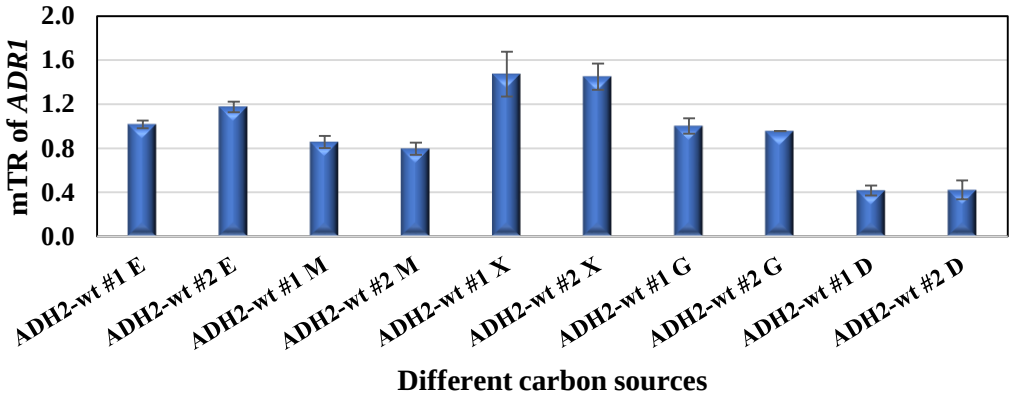


Figure 4.43 mRNA transcription ratios (mTRs) of *ADR1* in *P. pastoris* strains grown in different carbon sources at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to e thanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. Error bars represent the standard deviation among three technical replicates (±). Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol

Table 4.34 mRNA transcription ratios (mTRs) of *ADR1* in *P. pastoris* strains grown in different carbon sources at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol. StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	Carbon Source	mTR of *ADR1*	StDev
ADH2-wt #1	E	1.0	0.0
ADH2-wt #2	E	1.1	0.1
ADH2-wt #1	M	0.9	0.1
ADH2-wt #2	M	0.8	0.1
ADH2-wt #1	X	1.5	0.2
ADH2-wt #2	X	1.5	0.1
ADH2-wt #1	G	1.0	0.1
ADH2-wt #2	G	1.0	0.0
ADH2-wt #1	D	0.4	0.0
ADH2-wt #2	D	0.4	0.1

The variations in *ADR1* transcription on the carbon sources are given in Figure 4.43 and Table 4.34. The expression of *ADR1* gene also performed a similar trend with that of the *ACA1*. *ADR1* was expressed in *P. pastoris* with robust transcription performances in response to the carbon sources. The highest *ADR1* transcription was determined in the fermentation with limited glucose while the lowest was with the excess glucose. The fermentation with limited glucose resulted in a ca. 3.5-fold increase in the ADR1 transcription compared to that with the excess glucose, which is a consistent result related with the function of Adr1 which regulates the derepression of glucose-repressed genes.

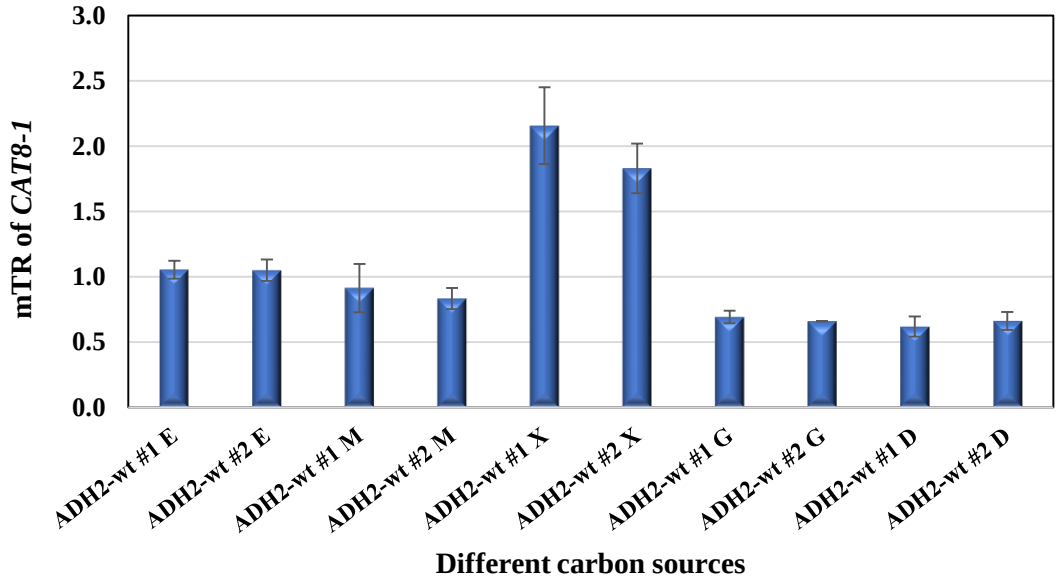


Figure 4.44 mRNA transcription ratios (mTRs) of *CAT8-1* in *P. pastoris* strains grown in different carbon sources at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented is relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. Error bars represent the standard deviation among three technical replicates (±). Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol.

Table 4.35 mRNA transcription ratios (mTRs) of *CAT8-1* in *P. pastoris* strains grown in different carbon sources at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol. StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	Carbon Source	mTR of *CAT8-1*	StDev
ADH2-wt #1	E	1.1	0.1
ADH2-wt #2	E	1.1	0.1
ADH2-wt #1	M	0.9	0.2
ADH2-wt #2	M	0.8	0.1
ADH2-wt #1	X	2.2	0.3
ADH2-wt #2	X	1.8	0.2
ADH2-wt #1	G	0.7	0.1
ADH2-wt #2	G	0.7	0.0
ADH2-wt #1	D	0.6	0.1
ADH2-wt #2	D	0.7	0.1

The variations in the *CAT8-1* transcription in response to the carbon sources are given in Figure 4.44 and Table 4.35. *CAT8-1* transcriptions were similar in the ethanol and methanol-grown cells. The transcription of *CAT8-1* were reduced 0.68- and 0.64-fold, respectively, in the fermentations with excess glycerol and excess glucose. At limited glucose condition, *CAT8-1* transcription was increased ca. 2-fold.

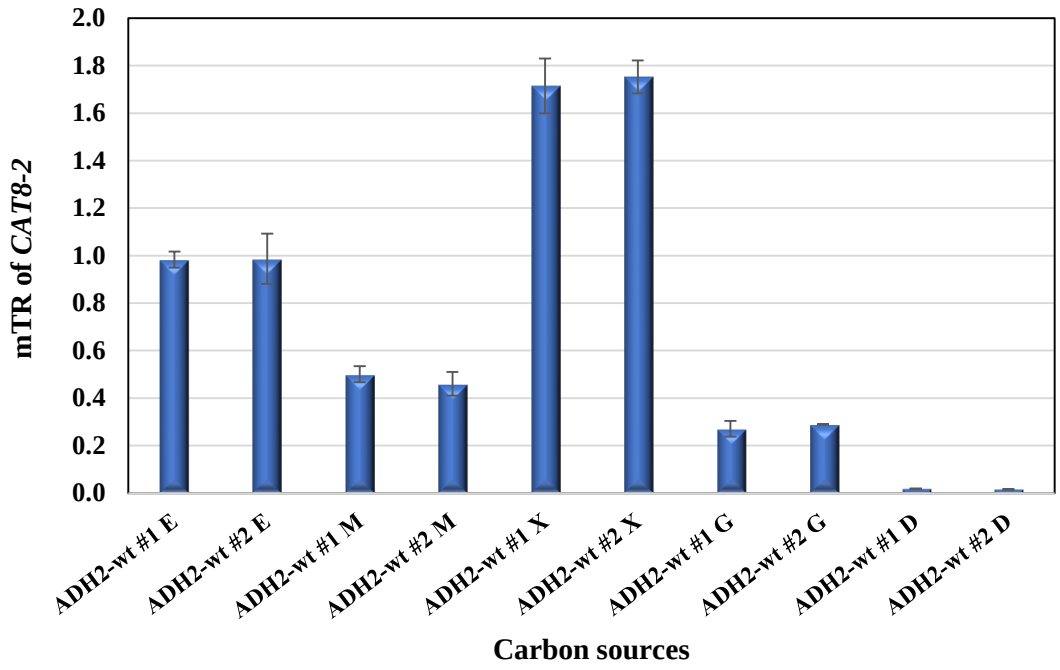


Figure 4. 45 mRNA transcription ratios (mTRs) of *CAT8-2* in *P. pastoris* strains grown in different carbon sources at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. Error bars represent the standard deviation among three technical replicates (±). Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol

Table 4. 36 mRNA transcription ratios (mTRs) of *CAT8-2* in *P. pastoris* strains grown in different carbon sources at t = 20 h. Transcription levels were normalized to *ACT1* mRNA expression and presented relative to ethanol grown wild-type *P. pastoris* constructed with $P_{ADH2\text{-}wt}$. Screening conditions: E: 2% (v/v) ethanol, M: 1% (v/v) methanol, X: limited glucose, D: excess glucose, G: excess glycerol. StDev: standard deviation of three technical replicates.

TECs of *P. pastoris*	**Carbon Source**	**mTR of *CAT8-2***	**StDev**
ADH2-wt #1	E	0.98	0.03
ADH2-wt #2	E	0.99	0.11
ADH2-wt #1	M	0.50	0.03
ADH2-wt #2	M	0.46	0.05
ADH2-wt #1	X	1.71	0.12
ADH2-wt #2	X	1.75	0.07
ADH2-wt #1	G	0.27	0.03
ADH2-wt #2	G	0.29	0.00
ADH2-wt #1	D	0.02	0.00
ADH2-wt #2	D	0.02	0.00

The variations in the *CAT8-2* transcription in response to the carbon sources are given in Figure 4.45 and Table 4.36. A different transcription pattern was observed with *CAT8-2* gene. With the excess glucose *CAT8-2* transcription did not occur. Whereas, the glycerol and methanol grown cells performed ca. 0.28- and 0.50-fold transcriptions, respectively. With the limited glucose *CAT8-2* transcription was 1.73-fold higher.

The overexpression results reveal that Cat8-2 is the only TF which cannot increase transcription in ethanol. Therefore we can conclude that Cat8-2 is the relatively abundant TF in the wild-type *P. pastoris*.

To conclude, all evaluated TFs in *P. pastoris* exhibited their highest transcriptions in the fermentations with limited-glucose which clarifies their functions in the derepression of glucose-repressed genes. Aca1, Adr1, and Cat8-1 exhibited

constitutive expression patterns while Cat8-2 had a regulated expression pattern in response to the carbon sources tested.

CHAPTER 5

CONCLUSIONS

The aim of the PhD thesis was to design and construct novel engineered promoter variants (NEPVs) of *P. pastoris* for recombinant protein (r-protein) production with strong and ethanol-inducible expression systems. In this context using promoter engineering approaches, the NEPVs: ***i)*** of ADH2 promoter (P_{ADH2}) in the first part, and ***ii)*** of AOX1 promoter (P_{AOX1}) in the second part, of the research programme were designed and constructed. The strength of the novel engineered promoter variants (NEPVs) was initially investigated with the intracellular synthesis of the enhanced green fluorescence protein (eGFP) in different carbon sources. Thereafter extracellular human serum albumin (HSA) production also was performed with the promising NEPVs to demonstrate their potentials.

ADH2 promoter (P_{ADH2}) variants were designed and constructed through nucleosome optimization and targeted addition and/or sequence optimization of the putative transcription factor binding sites (TFBSs). AOX1 promoter (P_{AOX1}) variants were engineered to change its induction mechanism from methanol to ethanol, which is originally a repressor carbon source for native P_{AOX1}. Adr1 (Mxr1), Aca2, Cat8, and TATA-binding motifs were used to design and construct P_{ADH2} and P_{AOX1} promoter libraries.

5.1 Novel Engineered Promoter Variants of P_{ADH2}

Based on the putative TFBSs, eight (8) P_{ADH2} variants were designed by adding or modifying the binding sites of the Adr1, Aca2, Cat8, and TATA. Computational redesign of nucleosome optimized P_{ADH2} was also performed using MATLAB algorithms that employ NuPoP. Taken together, two promising promotor engineering tools, *i.e.*, nucleosome optimization and optimization of the Cat8 binding motif, were combined in the design of $P_{ADH2\text{-}NucOpt\text{-}OptCat}$ and constructed. Overview of the NEPVs of P_{ADH2} is as follows:

a) With the NEPV $P_{ADH2\text{-}AddCat8\text{-}1(AddCat)}$ the cells in ethanol enhanced the eGFP synthesis 2.36-fold compared to that with $P_{ADH2\text{-}wt}$ in ethanol.
b) With the NEPV $P_{ADH2\text{-}AddCat2(OptCat)}$ the cells in ethanol and methanol enhanced the eGFP synthesis 4.76- and 3.80- fold compared to that with $P_{ADH2\text{-}wt}$ in ethanol and methanol, respectively.
c) With the NEPV $P_{ADH2\text{-}NucOpt}$ the cells in ethanol and methanol enhanced the eGFP synthesis 1.5- and 3.2-fold compared to that with $P_{ADH2\text{-}wt}$ in ethanol and methanol, respectively.
d) With the NEPV $P_{ADH2\text{-}NucOpt\text{-}OptCat}$ the cells in ethanol and methanol enhanced the eGFP synthesis 3.68- and 8.8-fold compared to that with $P_{ADH2\text{-}wt}$ in ethanol and methanol, respectively.
e) With the NEPVs $P_{ADH2\text{-}OptCat}$, $P_{ADH2\text{-}NucOpt}$, $P_{ADH2\text{-}AddCat}$, and $P_{ADH2\text{-}NucOpt\text{-}OptCat}$ the cells in ethanol enhanced transcription of the *eGFP* gene ca. 4.2-, 1.5-, 1.9-, and 3.6-fold compared to $P_{ADH2\text{-}wt}$ in ethanol, respectively.
f) With the NEPV $P_{ADH2\text{-}AddTATA}$ which was designed and constructed by integration of a TATA-binding motif, reduced the eGFP synthesis in the cells to 70% of the synthesis performed by the cells with $P_{ADH2\text{-}wt}$, in ethanol.
g) The promoter architecture designs which targeted Adr1 and Aca2 binding site modifications and/or additions, did not increase the eGFP synthesis in the cells with the NEPVs of P_{ADH2}.

h) In extracellular recombinant human serum albumin (r-HSA) production, with $P_{ADH2\text{-}wt}$, and with the NEPVs $P_{ADH2\text{-}NucOpt}$, and $P_{ADH2\text{-}OptCat}$, the r-HSA overall yield per gram of wet-cell weight ($Y_{P/X}$), respectively, were $Y_{P/X}$ = 0.26, 0.38, and 1.13 mg/g. Therefore, $Y_{P/X}$ values with $P_{ADH2\text{-}NucOpt}$ and $P_{ADH2\text{-}OptCat}$, respectively, were 1.43- and 4.30-fold higher than that with $P_{ADH2\text{-}wt}$.

5.2 Novel Engineered Promoter Variants of P_{AOX1}

The NEPVs of P_{AOX1} were designed by replacing Mxr1 response elements (Mxre) by *S. cerevisiae* Adr1 and Cat8 TFBS, besides the Aca2 TFBS; and ten NEPVs for P_{AOX1} were constructed. Overview of the ten NEPVs of P_{AOX1} is as follows:

a) A Cat8 binding motif was introduced onto the Mxre5 position of naturally occurring P_{AOX1} and the NEPV $P_{AOX1\text{-}Cat3}$ was constructed, which transformed ethanol repressed P_{AOX1} into the ethanol inducible NEPV $P_{AOX1\text{-}Cat3}$.

b) With the NEPV $P_{AOX1\text{-}Cat3}$ eGFP synthesis in the cells in ethanol was ca. 74% of expression in the cells constructed with $P_{AOX1\text{-}wt}$ in methanol. Besides, with the NEPV $P_{AOX1\text{-}Cat3}$ the expression in the cells in methanol enhanced 1.69-fold, compared to the cells with $P_{AOX1\text{-}wt}$ in methanol.

c) With the NEPV $P_{AOX1\text{-}Cat2}$ which was constructed by replacing the binding motif of Mxre3 in $P_{AOX1\text{-}wt}$ with a Cat8 binding motif, increased the expression strength 1.8-fold in methanol but did not change the repression in ethanol.

d) With the NEPV $P_{AOX1\text{-}Aca}$ which was constructed by integrating an Aca2 binding motif into $P_{AOX1\text{-}wt}$, increased the expression strength 1.38- fold in methanol, while there was no change in ethanol.

e) With the NEPV $P_{AOX1\text{-}Cat3Adr3}$ the cells in ethanol and methanol enhanced the eGFP synthesis 0.85- and 1.33-fold, respectively, compared to the cells with $P_{AOX1\text{-}wt}$ in methanol.

f) With the integration of an Aca2 binding motif into the NEPV $P_{AOX1\text{-}Cat3Adr3}$, the cells in ethanol reduced the eGFP synthesis from 85% to 60% compared to the cells with

$P_{AOX1\text{-}wt}$ in methanol; while the expression in methanol was increased from 133% to 165%.

g) The NEPV $P_{AOX1\text{-}mod}$ was designed by seven (7) simultaneous TFBS modifications. With the NEPV $P_{AOX1\text{-}mod}$ the cells in ethanol and methanol enhanced the eGFP synthesis 1.3- and 1.97-fold, respectively, compared to the cells with $P_{AOX1\text{-}wt}$ in methanol.

h) Introduction of Cat8 TFBS into a rationally designed position in P_{AOX1} created strong ethanol-inducible NEPVs denoted by $P_{AOX1\text{-}v}$, on a par with the methanol inducible $P_{AOX1\text{-}wt}$.

i) With the NEPVs $P_{AOX1\text{-}Cat3}$, $P_{AOX1\text{-}Aca}$, and $P_{AOX1\text{-}mod}$ the cells in methanol enhanced the transcription of the *eGFP* gene 1.57-, 1.45-, and 1.80-fold, compared to that with $P_{AOX1\text{-}wt}$ in methanol. Besides in ethanol, with the NEPVs $P_{AOX1\text{-}Cat3}$ and $P_{AOX1\text{-}mod}$ the cells enhanced transcription of the *eGFP* gene 0.70- and 1.34-fold, compared to the cells with $P_{AOX1\text{-}wt}$ in methanol.

j) In extracellular recombinant human serum albumin (HSA) production, with the NEPV $P_{AOX1\text{-}mod}$ the cells in ethanol produced 27.83 µg/ml, which is a close to 32.05 µg/ml which was produced with $P_{AOX1\text{-}wt}$ in methanol.

k) In extracellular recombinant human serum albumin (HSA) production, with the NEPV $P_{AOX1\text{-}mod}$ the r-HSA overall yield per gram of wet-cell weight ($Y_{P/X}$) were $Y_{P/X}$ = 0.476 mg/g and $Y_{P/X}$ = 1.341 mg/g in ethanol and methanol, respectively. Whereas, with the naturally occurring $P_{AOX1\text{-}wt}$,the $Y_{P/X}$ value in methanol was $Y_{P/X}$ = 0.848 mg/g. Thus, the NEPV $P_{AOX1\text{-}mod}$ performed r-HSA production in ethanol with a considerable yield which is %55 of the produced with $P_{AOX1\text{-}wt}$ in methanol; however, in methanol with a 1.59-fold increased yield compared to the produced with $P_{AOX1\text{-}wt}$ in methanol.

l) Addition an Adr1(Mxr1) binding site into P_{AOX1} and also Adr1(Mxr1) binding site optimization neither increased the promoter strength nor altered the regulation of P_{AOX1}, is indeed important

5.3 Transcriptional Engineering with Naturally Occuring Promoters and with a NEPV

In order to identify the functions of the TFs on the regulation of the naturally occuring promoters used in the design of the NEPVs, and of a NEPV of P_{AOX1}, using the tools of transcriptional engineering, transcription factor overexpression (TF-OE) and transcription factor knock-out (TF-KO) strains of Adr1 (Mxr1), Aca1, Cat8-1, and Cat8-2 were designed and constructed and tested in the synthesis of eGFP. Influences of the deletion and overexpression of the TFs on the strength of the promoters were tested with five carbon sources.

Overview of the transcriptionally engineered cells (TECs) by TF-OE and TF-KO with naturally occuring promoters P_{ADH2} and P_{AOX1} which are denoted as $P_{ADH2\text{-}wt}$ and $P_{AOX1\text{-}wt}$, respectively, are as follows:

5.3.1 TECs with $P_{ADH2\text{-}wt}$

a) The TEC with $P_{ADH2\text{-}wt}$ designed with *cat8-2*Δ enhanced eGFP synthesis 1.35- fold compared to that with $P_{ADH2\text{-}wt}$ in ethanol.

b) The TEC with $P_{ADH2\text{-}wt}$ designed with Cat8-2-OE demonstrated the highest derepression with the limited glucose and excess glycerol; where, the TECs with limited glucose and excess glycerol synthesized, respectively, 37% and 33% of the cells with $P_{ADH2\text{-}wt}$ in ethanol.

c) The TECs with $P_{ADH2\text{-}wt}$ designed with Aca1-OE, Adr1-OE, Cat8-1-OE, and Cat8-2-OE did not lead to any change in the expression in ethanol, however, enhanced the eGFP synthesis in methanol, respectively, to 57, 83, 29, and 61% of the cells with $P_{ADH2\text{-}wt}$ in ethanol.

d) The TEC with $P_{ADH2\text{-}wt}$ designed with deletion of *ADR1* gene reduced the eGFP expression in ethanol to 67 % of the wild-type strain. However, in methanol, the TEC with $P_{ADH2\text{-}wt}$ designed with *adr1Δ* enhanced the expression 1.83-fold of the wild-type

but failed in growing. This indicates that the TEC maintains the cellular functions with a low methanol uptake rate; however, other enzymes such as yeast ADH2 or AOX2 can catalyze methanol utilization to some extent.

e) The TECs with $P_{ADH2\text{-}wt}$ designed with deletions of either *cat8-1*Δ or *cat8-2*Δ increased the eGFP expressions in ethanol, respectively, 1.28- and 1.35-fold compared to the wild-type strain. Deletion of either *CAT8-1* or *CAT8-2* leads to an increase in the transcription of the other gene in ethanol. The single-knock out deletions revealed that the TEC having Cat8-1 can increase the expression higher in ethanol; whereas, the other TEC having Cat8-2 OE can increase the expression higher in methanol. Furthermore, in order to clarify the regulatory functions of Cat8-1 and Cat8-2 on the promoter, with a special design, the TEC with $P_{ADH2\text{-}wt}$ constructed with *cat8-1*Δ*cat8-2*Δ double knock-out reduced the eGFP expressions in ethanol to 71% of the cells with $P_{ADH2\text{-}wt}$. Moreover *P. pastoris* lost its ethanol utilization ability. Therefore, it is conclusively demonstrated that both Cat8-1 and Cat8-2 TFs are transcriptional activators of $P_{ADH2\text{-}wt}$.

f) The TEC with $P_{ADH2\text{-}wt}$ designed with the deletion of *adr1*Δ showed growth impairment in minimal media, containing methanol, ethanol, limited glucose, excess glucose or excess glycerol, which could be due to genome-wide integrated regulatory effects of Adr1(Mxr1).

5.3.2 TECs with $P_{AOX1\text{-}wt}$

a) In general, the TECs with $P_{AOX1\text{-}wt}$ enhanced expression strength of $P_{AOX1\text{-}wt}$ further.

b) The TEC with $P_{AOX1\text{-}wt}$ designed with *ADR1*-OE enhanced the eGFP synthesis in methanol 3.45-fold than that of the cells with $P_{AOX1\text{-}wt}$. The TECs with $P_{AOX1\text{-}wt}$ designed with Aca1-, Cat8-1-, and Cat8-2- OEs enhanced the eGFP syntheses in methanol, respectively, 1.53-, 1.18-, and 1.18-fold than that of the cells with $P_{AOX1\text{-}wt}$.

c) The TEC with $P_{AOX1\text{-}wt}$ designed with *ADR1*-OE created derepression in production of r-proteins in fermentation with glycerol; where, the synthesis in the TEC was 23% of the expression in the cells with $P_{AOX1\text{-}wt}$ in methanol.

d) The TECs with $P_{AOX1\text{-}wt}$ designed with deletions of either *cat8-1*Δ or *cat8-2*Δ influenced intracellular regulations. Deletion of *CAT8-1* enhanced the expression in the TEC in methanol 1.17-fold, and in ethanol enhanced to 22% of the cells with $P_{AOX1\text{-}wt}$ in methanol. Besides, deletion of *CAT8-2* enhanced the expression in the TEC in methanol 1.32-fold, and in ethanol enhanced to 21% of the cells with $P_{AOX1\text{-}wt}$ in methanol. These results conclusively demonstrate that merely with either *CAT8-1* or *CAT8-2*, the promoter architecture in the TEC becomes stronger; as was observed in $P_{ADH2\text{-}wt}$.

e) The TECs with $P_{AOX1\text{-}wt}$ designed with deletions of either *cat8-1*Δ or *cat8-2*Δ showed that Cat8-1 is sufficient and superior to Cat8-2 in methanol.

f) P_{AOX1} promoter- and transcriptional- engineering results reveal that *P. pastoris* Aca1 TF operates as an activator on the regulation of $P_{AOX1\text{-}wt}$ in methanol. The mRNA transcription measurements and the calculated mTRs reinforces this result as an 8-fold increase was observed in the transcription of *ACA1* in methanol in the TEC designed by *adr1*Δ, while no change was observed in ethanol. Therefore, Aca1 should be contributing to the activation of Adr1-dependent genes in methanol. Since the widely used carbon source of *P. pastoris* is methanol, it is plausible for Aca1to be active on the MUT pathway genes.

g) The TEC with $P_{AOX1\text{-}wt}$ designed by *ACA1*- KO, or in other words by *ACA1* deletion, could not be constructed despite the deletion trials guided with two different RNAs, which drives us to the conclusion that *ACA1* gene can be an essential gene for *P. pastoris*.

h) The TEC with $P_{AOX1\text{-}wt}$ designed by deletion of *ADR1* gene led to a total loss of $P_{AOX1\text{-}wt}$ activity in all the carbon sources tested. The deletion of *ADR1* reduced the expression of the TEC designed with $P_{ADH2\text{-}wt}$ in ethanol to 67% that of the cells with $P_{ADH2\text{-}wt}$. whereas, in fermentations with methanol, limited glucose, excess glycerol, and excess glucose $P_{ADH2\text{-}wt}$ expressions were increased by *adr1*Δ. These results conclusively demonstrate that Adr1 is not the main regulator of *P. pastoris* P_{ADH2}; furthermore, even in the absence of Adr1, $P_{ADH2\text{-}wt}$ performs remarkable expression

strength. Although, in *S. cerevisiae* Adr1 was annotated as an alcohol dehydrogenase 2 synthesis regulator due to its regulatory function, our findings demonstrated that in *P. pastoris* Adr1 is the main activator of P_{AOX1} in response to methanol; therefore its annotation as "methanol expression regulator 1" (Mxr1) represents its role in *P. pastoris* accurately.

5.3.3 Transcriptional Engineering with a NEPV: TECs of $P_{AOX1\text{-}mod}$

As the NEPV $P_{AOX1\text{-}mod}$ altered the well known regulations of the P_{AOX1}-wt, the TECs were designed with $P_{AOX1\text{-}mod}$. Integration of novel *cis*-acting elements into P_{AOX1} generated NEPVs that have distinct transcriptional strengths and regulations. Overexpression of all TFs increased the activity of $P_{AOX1\text{-}mod}$ in methanol and ethanol, as follows:

a) The TEC with $P_{AOX1\text{-}mod}$ designed by Adr1-OE in ethanol enhanced eGFP expression 4.44-fold compared to the wild-type cells with $P_{AOX1\text{-}wt}$ in methanol. Besides, the TEC with $P_{AOX1\text{-}mod}$ designed by Adr1-OE in methanol enhanced the expression 3.83-fold, which was 2.26-fold with the NEPV $P_{AOX1\text{-}mod}$ in wild-type (wt) *P. pastoris* compared to the wt-cells with $P_{AOX1\text{-}wt}$ in methanol.

b) The TEC with $P_{AOX1\text{-}mod}$ designed by Aca1-OE in methanol enhanced the expression 3.12-fold compared to the wt-cells with $P_{AOX1\text{-}wt}$ in methanol.

c) The TECs with $P_{AOX1\text{-}mod}$ designed by Cat8-1-OE and Cat8-2-OE in ethanol and methanol enhanced the expression 2.14- and 3.13-fold, compared to the wt-cells with $P_{AOX1\text{-}wt}$ in methanol.

d) The TECs with $P_{AOX1\text{-}mod}$ designed with deletions of either *cat8-1*Δ or *cat8-2*Δ showed that Cat8-1 is sufficient and superior to Cat8-2 in ethanol and methanol.

e) The TEC with $P_{AOX1\text{-}mod}$ designed with deletion of *ADR1* gene in methanol enhanced the expression 6.33-fold compared to the wt-cells with $P_{AOX1\text{-}wt}$. The TEC could not grow on methanol but our results confirmed that methanol was consumed through the less-efficient pathways and the cell metabolism was operative at a

sustainable level. ADR1 deletion exhibited very different consequences with $P_{AOX1\text{-}wt}$ where all the promoter activity was lost; however, in *adr1*Δ strain $P_{AOX\text{-}mod}$ demonstrates remarkable activity in all carbon sources tested.

f) A very potent derepression based eGFP production system was generated with the TEC with $P_{AOX1\text{-}mod}$ designed by deletions of *ADR1* gene in fermentations with limited glucose enhanced expression 2.09-fold compared to the wt-cells with $P_{AOX1\text{-}wt}$ in methanol.

Summary of Conclusion

In this study, promoter engineering of P_{ADH2} led to stronger NEPVs that reached 4.76-fold higher intracellular eGFP synthesis and 4.30-fold stronger extracellular HSA production levels than that of $P_{ADH2\text{-}wt}$. As a second approach, the induction mechanism of P_{AOX1} was redesigned for the use of ethanol as sole carbon and energy source. Regulation of ethanol repressible P_{AOX1} was changed to be ethanol-inducible by a single TFBS modification, which showed P_{AOX1} was mainly controlled by its actual *cis*-acting elements. NEPV $P_{AOX1\text{-}mod}$ exhibited 1.97-fold higher intracellular eGFP synthesis potential and 1.59-fold higher extracellular HSA production level than that of $P_{AOX1\text{-}wt}$ in methanol. At the same time, in ethanol, $P_{AOX1\text{-}mod}$ reached 1.30-fold higher eGFP synthesis level and similar extracellular HSA titer compared to methanol-induced $P_{AOX1\text{-}wt.}$

Transcriptional engineering of *P. pastoris* created novel r-protein expression systems by $P_{ADH2\text{-}wt}$ and $P_{AOX1\text{-}mod}$ *i.e.* stronger induction by ethanol and methanol or derepression in limited glucose and excess glycerol. Overexpression of Cat8-2 provided a derepressible r-protein production system in limited-glucose; in that activity of P_{ADH2} reached 37% of ethanol-induced P_{ADH2} activity. A very potent activation by derepression was observed by $P_{AOX1\text{-}mod}$ in *adr1*Δ strain, in limited glucose 206% activity of methanol-induced $P_{AOX1\text{-}wt}$ was obtained with the TEC. Cat8-1 and Cat8-2 are both activators of P_{ADH2} and P_{AOX1}; and *cat8-1cat8-2*Δ double knock-

out strain lost its ethanol utilization ability. Among *P. pastoris* Cat8 TFs, merely Cat8-1 was sufficient for the activation of P_{ADH2} and $P_{AOX1\text{-}v}$. Also, Cat8-1 was superior to Cat8-2 in terms of effective transcriptional activation. Aca1 was mainly effective on $P_{AOX1\text{-}v}$ in methanol. Adr1 was found as the main regulator of P_{AOX1} but not P_{ADH2}. Adr1 deletion resulted in impaired *P. pastoris* growth in minimal media including limited glucose, excess glucose, excess glycerol; and the cell lost its ability to grow on methanol. The highest ethanol induction activity by $P_{AOX1\text{-}mod}$ was obtained with Adr1 overexpression. In methanol containing minimal media, the strongest expression potential by $P_{ADH2\text{-}wt}$ and $P_{AOX1\text{-}mod}$ was obtained in *adr1*Δ strains, although the strain lost its ability to grow on methanol; which makes the TECs a potential candidate for retentostat bioreactor applications.

The results of this Ph.D. research demonstrated that the NEPVs and transcriptionally engineered r-protein expression platforms are promising candidates for industrial r-protein production processes.

REFERENCES

Abate, G., Bastonini, E., Braun, K.A., Verdone, L., Young, E.T., Caserta, M., 2012. Snf1/AMPK regulates Gcn5 occupancy, H3 acetylation and chromatin remodelling at S. cerevisiae ADY2 promoter. Biochim. Biophys. Acta 1819, 419–427. https://doi.org/10.1016/j.bbagrm.2012.01.009

Afek, A., Lukatsky, D.B., 2012. Nonspecific Protein-DNA Binding Is Widespread in the Yeast Genome. Biophys. J. 102, 1881–1888. https://doi.org/10.1016/j.bpj.2012.03.044

Afek, A., Sela, I., Musa-Lempel, N., Lukatsky, D.B., 2011. Nonspecific Transcription-Factor-DNA Binding Influences Nucleosome Occupancy in Yeast. Biophys. J. 101, 2465–2475. https://doi.org/10.1016/j.bpj.2011.10.012

Ahn, J., Hong, J., Lee, H., Park, M., Lee, E., Kim, C., Choi, E., Jung, J., Lee, H., 2007. Translation elongation factor 1-alpha gene from Pichia pastoris: molecular cloning, sequence, and use of its promoter. Appl. Microbiol. Biotechnol. 74, 601–608. https://doi.org/10.1007/s00253-006-0698-6

Ahn, J., Hong, J., Park, M., Lee, H., Lee, E., Kim, C., Lee, J., Choi, E., Jung, J., Lee, H., 2009. Phosphate-responsive promoter of a Pichia pastoris sodium phosphate symporter. Appl. Environ. Microbiol. 75, 3528–3534. https://doi.org/10.1128/AEM.02913-08

Alper, H., Fischer, C., Nevoigt, E., Stephanopoulos, G., 2005. Tuning genetic control through promoter engineering. Proc. Natl. Acad. Sci. U. S. A. 102, 12678–12683. https://doi.org/10.1073/pnas.0504604102

Archer, D.B., MacKenzie, D.A., Jeenes, D.J., 2006. Genetic engineering: yeasts and filamentous fungi, in: Kristiansen, B., Ratledge, C. (Eds.), Basic Biotechnology. Cambridge University Press, Cambridge, pp. 119–154.

https://doi.org/10.1017/CBO9780511802409.007

Ata, Ö., Prielhofer, R., Gasser, B., Mattanovich, D., Çalık, P., 2017. Transcriptional engineering of the glyceraldehyde-3-phosphate dehydrogenase promoter for improved heterologous protein production in Pichia pastoris. Biotechnol. Bioeng. https://doi.org/10.1002/bit.26363

Bajhaiya, A.K., Ziehe Moreira, J., Pittman, J.K., 2017. Transcriptional Engineering of Microalgae: Prospects for High-Value Chemicals. Trends Biotechnol. 35, 95–99. https://doi.org/10.1016/j.tibtech.2016.06.001

Basehoar, A.D., Zanton, S.J., Pugh, B.F., 2004. Identification and Distinct Regulation of Yeast TATA Box-Containing Genes. Cell 116, 699–709. https://doi.org/10.1016/S0092-8674(04)00205-3

Baumann, K., Carnicer, M., Dragosits, M., Graf, A.B., Stadlmann, J., Jouhten, P., Maaheimo, H., Gasser, B., Albiol, J., Mattanovich, D., Ferrer, P., 2010. A multi-level study of recombinant Pichia pastoris in different oxygen conditions. BMC Syst. Biol. 4, 141. https://doi.org/10.1186/1752-0509-4-141

Beier, D., Sledziewski, A., Young, E.T., 1985. Deletion analysis identifies a region, upstream of the ADH2 gene of Saccharomyces cerevisiae, which is required for ADR1- mediated derepression. Mol. Cell. Biol. 5, 1743–1749. https://doi.org/10.1128/MCB.5.7.1743

Berg, L., Strand, T.A., Valla, S., Brautaset, T., 2013. Combinatorial mutagenesis and selection to understand and improve yeast promoters. Biomed Res. Int. 2013, 926985. https://doi.org/10.1155/2013/926985

Berg, O.G., Winter, R.B., Von Hippel, P.H., 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. Biochemistry 20, 6929–6948. https://doi.org/10.1021/bi00527a028

Bhat, P.J., Murthy, T.V.S., 2001. Transcriptional control of the GAL/MEL regulon of yeast Saccharomyces cerevisiae: mechanism of galactose-mediated signal transduction. Mol. Microbiol. 40, 1059–1066. https://doi.org/10.1046/j.1365-2958.2001.02421.x

Blake, W.J., Balazsi, G., Kohanski, M.A., Isaacs, F.J., Murphy, K.F., Kuang, Y., Cantor, C.R., Walt, D.R., Collins, J.J., 2006. Phenotypic consequences of promoter-mediated transcriptional noise. Mol. Cell 24, 853–865. https://doi.org/10.1016/j.molcel.2006.11.003

Blanchin-Roland, S., Cordero Otero, R.R., Gaillardin, C., 1994. Two upstream activation sequences control the expression of the XPR2 gene in the yeast Yarrowia lipolytica. Mol. Cell. Biol. 14, 327–338. https://doi.org/10.1128/MCB.14.1.327

Blazeck, J., Garg, R., Reed, B., Alper, H.S., 2012. Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. Biotechnol. Bioeng. 109, 2884–2895. https://doi.org/10.1002/bit.24552

Blazeck, J., Liu, L., Redden, H., Alper, H., 2011. Tuning gene expression in Yarrowia lipolytica by a hybrid promoter approach. Appl. Environ. Microbiol. 77, 7905–7914. https://doi.org/10.1128/AEM.05763-11

Blumberg, H., Hartshorne, T.A., Young, E.T., 1988. Regulation of expression and activity of the yeast transcription factor ADR1. Mol. Cell. Biol. 8, 1868–1876. https://doi.org/10.1128/MCB.8.5.1868

Boer, M. d., 2000. Stp1p, Stp2p and Abf1p are involved in regulation of expression of the amino acid transporter gene BAP3 of Saccharomyces cerevisiae. Nucleic Acids Res. 28, 974–981. https://doi.org/10.1093/nar/28.4.974

Bowers, P.M., Schaufler, L.E., Klevit, R.E., 1999. A folding transition and novel zinc

finger accessory domain in the transcription factor ADR1. Nat. Struct. Biol. 6, 478. https://doi.org/10.1038/8283

Braun, K.A., Parua, P.K., Dombek, K.M., Miner, G.E., Young, E.T., 2013. 14-3-3 (Bmh) Proteins Regulate Combinatorial Transcription following RNA Polymerase II Recruitment by Binding at Adr1- Dependent Promoters in Saccharomyces cerevisiae. Mol. Cell. Biol. 33, 712–724. https://doi.org/10.1128/MCB.01226-12

Braun, K.A., Vaga, S., Dombek, K.M., Fang, F., Palmisano, S., Aebersold, R., Young, E.T., 2014. Phosphoproteomic analysis identifies proteins involved in transcription-coupled mRNA decay as targets of Snf1 signaling. Sci. Signal. 7, ra64. https://doi.org/10.1126/scisignal.2005000

Çalık, P., Ata, Ö., Güneş, H., Massahi, A., Boy, E., Keskin, A., Öztürk, S., Zerze, G.H., Özdamar, T.H., 2015. Recombinant protein production in Pichia pastoris under glyceraldehyde-3-phosphate dehydrogenase promoter: From carbon source metabolism to bioreactor operation parameters. Biochem. Eng. J. 95, 20–36. https://doi.org/10.1016/j.bej.2014.12.003

Cámara, E., Landes, N., Albiol, J., Gasser, B., Mattanovich, D., Ferrer, P., 2017. Increased dosage of AOX1 promoter-regulated expression cassettes leads to transcription attenuation of the methanol metabolism in Pichia pastoris. Sci. Rep. 7, 44302. https://doi.org/10.1038/srep44302

Camattari, A., Bianchi, M.M., Branduardi, P., Porro, D., Brambilla, L., 2007. Induction by Hypoxia of Heterologous-Protein Production with the KlPDC1 Promoter in Yeasts . Appl. Environ. Microbiol. 73, 922–929. https://doi.org/10.1128/AEM.01764-06

Carlson, M., 1999. Glucose repression in yeast. Curr. Opin. Microbiol. 2, 202–207.

https://doi.org/10.1016/S1369-5274(99)80035-6

Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M., Werner, T., 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 21, 2933–2942. https://doi.org/10.1093/bioinformatics/bti473

Caspary, F., Hartig, A., Schüller, H.-J., 1997. Constitutive and carbon source-responsive promoter elements are involved in the regulated expression of the Saccharomyces cerevisiae malate synthase gene MLS1. Mol. Gen. Genet. MGG 255, 619–627. https://doi.org/10.1007/s004380050536

Çelik, E., Çalık, P., 2012. Production of recombinant proteins by yeast cells. Biotechnol. Adv. 30, 1108–1118. https://doi.org/10.1016/j.biotechadv.2011.09.011

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

Chang, D.T.-H., Huang, C.-Y., Wu, C.-Y., Wu, W.-S., 2011. YPA: an integrated repository of promoter features in Saccharomyces cerevisiae. Nucleic Acids Res. 39, D647-52. https://doi.org/10.1093/nar/gkq1086

Chen, Y., Rai, R., Zhou, Z.-R., Kanoh, J., Ribeyre, C., Yang, Y., Zheng, H., Damay, P., Wang, F., Tsujii, H., Hiraoka, Y., Shore, D., Hu, H.-Y., Chang, S., Lei, M., 2011. A Conserved Motif within RAP1 Plays Diversified Roles in Telomere Protection and Regulation in Different Organisms. Nat. Struct. Mol. Biol. 18, 213–221. https://doi.org/10.1038/nsmb.1974

Cheng, C., Kacherovsky, N., Dombek, K.M., Camier, S., Thukral, S.K., Rhim, E., Young, E.T., 1994. Identification of potential target genes for Adr1p through

characterization of essential nucleotides in UAS1. Mol. Cell. Biol. 14, 3842–52.

Cherstvy, A.G., 2011. Electrostatic interactions in biological DNA-related systems. Phys. Chem. Chem. Phys. 13, 9942–9968. https://doi.org/10.1039/c0cp02796k

Cherstvy, A.G., Kolomeisky, A.B., Kornyshev, A.A., 2008. Protein−DNA Interactions: Reaching and Recognizing the Targets. J. Phys. Chem. B 112, 4741–4750. https://doi.org/10.1021/jp076432e

Chiang, Y.-C., Komarnitsky, P., Chase, D., Denis, C.L., 1996. ADR1 Activation Domains Contact the Histone Acetyltransferase GCN5 and the Core Transcriptional Factor TFIIB. J. Biol. Chem. 271, 32359–32365. https://doi.org/10.1074/jbc.271.50.32359

Chien, L.-J., Lee, C.-K., 2005. Expression of bacterial hemoglobin in the yeast, Pichia pastoris, with a low O2-induced promoter. Biotechnol. Lett. 27, 1491–1497. https://doi.org/10.1007/s10529-005-1324-x

Choi, B.-K., Bobrowicz, P., Davidson, R.C., Hamilton, S.R., Kung, D.H., Li, H., Miele, R.G., Nett, J.H., Wildt, S., Gerngross, T.U., 2003. Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast Pichia pastoris. Proc. Natl. Acad. Sci. 100, 5022–5027. https://doi.org/10.1073/pnas.0931263100

Ciofalo, V., Barton, N., Kreps, J., Coats, I., Shanahan, D., 2006. Safety evaluation of a lipase enzyme preparation, expressed in Pichia pastoris, intended for use in the degumming of edible vegetable oil. Regul. Toxicol. Pharmacol. 45, 1–8. https://doi.org/10.1016/j.yrtph.2006.02.001

Cook, W.J., Chase, D., Audino, D.C., Denis, C.L., 1994. Dissection of the ADR1 protein reveals multiple, functionally redundant activation domains interspersed with inhibitory regions: evidence for a repressor binding to the ADR1c region.

Mol. Cell. Biol. 14, 629–640. https://doi.org/10.1128/MCB.14.1.629

Cottier, F., Raymond, M., Kurzai, O., Bolstad, M., Leewattanapasuk, W., Jiménez-López, C., Lorenz, M.C., Sanglard, D., Váchová, L., Pavelka, N., Palková, Z., Mühlschlegel, F.A., 2012. The bZIP Transcription Factor Rca1p Is a Central Regulator of a Novel CO(2) Sensing Pathway in Yeast. PLoS Pathog. 8, e1002485. https://doi.org/10.1371/journal.ppat.1002485

Couderc, R., Baratti, J., 1980. Oxidation of methanol by the yeast, Pichia pastoris. Purification and properties of the alcohol oxidase. Agric. Biol. Chem. 44, 2279–2289. https://doi.org/10.1080/00021369.1980.10864320

Cregg, J.M., Barringer, K.J., Hessler, A.Y., Madden, K.R., 1985. Pichia pastoris as a host system for transformations. Mol. Cell. Biol. 5, 3376–3385. https://doi.org/10.1128/MCB.5.12.3376

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. Mol. Cell. Biol. 9, 1316–1323. https://doi.org/10.1128/MCB.9.3.1316

Cregg, J.M., Tolstorukov, I.I., 2012. P. pastoris ADH promoter and use thereof to direct expression of proteins. US 8222386 B2.

Curran, K.A., Crook, N.C., Karim, A.S., Gupta, A., Wagman, A.M., Alper, H.S., 2014. Design of synthetic yeast promoters via tuning of nucleosome architecture. Nat. Commun. 5, 1–8. https://doi.org/10.1038/ncomms5002

de Almeida, J.R.M., de Moraes, L.M.P., Torres, F.A.G., 2005. Molecular characterization of the 3-phosphoglycerate kinase gene (PGK1) from the methylotrophic yeast Pichia pastoris. Yeast 22, 725–737. https://doi.org/10.1002/yea.1243

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

Delic, M., Mattanovich, D., Gasser, B., 2013. Repressible promoters - a novel tool to generate conditional mutants in Pichia pastoris. Microb. Cell Fact. 12, 6. https://doi.org/10.1186/1475-2859-12-6

Denis, C.L., Fontaine, S.C., Chase, D., Kemp, B.E., Bemis, L.T., 1992. ADR1c mutations enhance the ability of ADR1 to activate transcription by a mechanism that is independent of effects on cyclic AMP-dependent protein kinase phosphorylation of Ser-230. Mol. Cell. Biol. 12, 1507–1514. https://doi.org/10.1128/MCB.12.4.1507

Dombek, K.M., Young, E.T., 1997. Cyclic AMP-dependent protein kinase inhibits ADH2 expression in part by decreasing expression of the transcription factor gene ADR1. Mol. Cell. Biol. 17, 1450–1458. https://doi.org/10.1128/MCB.17.3.1450

Dossani, Z.Y., Reider Apel, A., Szmidt-Middleton, H., Hillson, N.J., Deutsch, S., Keasling, J.D., Mukhopadhyay, A., 2018. A combinatorial approach to synthetic transcription factor-promoter combinations for yeast strain engineering. Yeast 35, 273–280. https://doi.org/10.1002/yea.3292

Eisen, A., Taylor, W.E., Blumberg, H., Young, E.T., 1988. The yeast regulatory protein ADR1 binds in a zinc-dependent manner to the upstream activating sequence of ADH2. Mol. Cell. Biol. 8, 4552–4556. https://doi.org/10.1128/MCB.8.10.4552

Ellis, S.B., Brust, P.F., Koutz, P.J., Waters, A.F., Harpold, M.M., Gingeras, T.R.,

1985. Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast Pichia pastoris. Mol. Cell. Biol. 5, 1111–1121. https://doi.org/10.1128/MCB.5.5.1111

Engler, C., Kandzia, R., Marillonnet, S., 2008. A one pot, one step, precision cloning method with high throughput capability. PLoS One 3, e3647. https://doi.org/10.1371/journal.pone.0003647

Featherstone, M., 2002. Coactivators in transcription initiation: here are your orders. Curr. Opin. Genet. Dev. 12, 149–155. https://doi.org/10.1016/S0959-437X(02)00280-0

Feller, A., Dubois, E., Ramos, F., Piérard, A., 1994. Repression of the genes for lysine biosynthesis in Saccharomyces cerevisiae is caused by limitation of Lys14-dependent transcriptional activation. Mol. Cell. Biol. 14, 6411–6418. https://doi.org/10.1128/MCB.14.10.6411

Friden, P., Schimmel, P., 1988. LEU3 of Saccharomyces cerevisiae activates multiple genes for branched-chain amino acid biosynthesis by binding to a common decanucleotide core sequence. Mol. Cell. Biol. 8, 2690–2697. https://doi.org/10.1128/MCB.8.7.2690

Fry, C.J., Peterson, C.L., 2001. Chromatin remodeling enzymes: who's on first? Curr. Biol. 11, R185–R197. https://doi.org/10.1016/S0960-9822(01)00090-2

Garcia-Gimeno, M.A., Struhl, K., 2000. Aca1 and Aca2, ATF/CREB Activators in Saccharomyces cerevisiae, Are Important for Carbon Source Utilization but Not the Response to Stress. Mol. Cell. Biol. 20, 4340–4349. https://doi.org/10.1128/MCB.20.12.4340-4349.2000

Gasser, B., Prielhofer, R., Marx, H., Maurer, M., Nocon, J., Steiger, M., Puxbaum, V., Sauer, M., Mattanovich, D., 2013. Pichia pastoris: protein production host

and model organism for biomedical research. Future Microbiol. 8, 191–208. https://doi.org/10.2217/fmb.12.133

Georis, I., Krijger, J.-J., Breunig, K.D., Vandenhaute, J., 2000. Differences in regulation of yeast gluconeogenesis revealed by Cat8p-independent activation of PCK1 and FBP1 genes in Kluyveromyces lactis. Mol. Gen. Genet. MGG 264, 193–203. https://doi.org/10.1007/s004380000314

Gündüz-Ergün, B., Çalık, P., 2016. Lignocellulose degrading extremozymes produced by Pichia pastoris: current status and future prospects. Bioprocess Biosyst. Eng. 39, 1–36. https://doi.org/10.1007/s00449-015-1476-6

Güneş, H., Çalık, P., 2016. Oxygen transfer as a tool for fine-tuning recombinant protein production by Pichia pastoris under glyceraldehyde-3-phosphate dehydrogenase promoter. Bioprocess Biosyst. Eng. 39, 1061–1072. https://doi.org/10.1007/s00449-016-1584-y

Hahn, S., Young, E.T., 2011. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics 189, 705–736. https://doi.org/10.1534/genetics.111.127019

Hamilton, S.R., Bobrowicz, P., Bobrowicz, B., Davidson, R.C., Li, H., Mitchell, T., Nett, J.H., Rausch, S., Stadheim, T.A., Wischnewski, H., Wildt, S., Gerngross, T.U., 2003. Production of complex human glycoproteins in yeast. Science 301, 1244–1246. https://doi.org/10.1126/science.1088166

Hampsey, M., 1998. Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery. Microbiol. Mol. Biol. Rev. 62, 465–503.

Hartner, F.S., Ruth, C., Langenegger, D., Johnson, S.N., Hyka, P., Lin-Cereghino, G.P., Lin-Cereghino, J., Kovar, K., Cregg, J.M., Glieder, A., 2008. Promoter

library designed for fine-tuned gene expression in Pichia pastoris. Nucleic Acids Res. 36, e76–e76. https://doi.org/10.1093/nar/gkn369

Hasslacher, M., Schall, M., Hayn, M., Bona, R., Rumbold, K., Luckl, J., Griengl, H., Kohlwein, S.D., Schwab, H., 1997. High-level intracellular expression of hydroxynitrile lyase from the tropical rubber tree Hevea brasiliensis in microbial hosts. Protein Expr. Purif. 11, 61–71. https://doi.org/10.1006/prep.1997.0765

Haurie, V., Perrot, M., Mini, T., Jeno, P., Sagliocco, F., Boucherie, H., 2001. The transcriptional activator Cat8p provides a major contribution to the reprogramming of carbon metabolism during the diauxic shift in Saccharomyces cerevisiae. J. Biol. Chem. 276, 76–85. https://doi.org/10.1074/jbc.M008752200

Havas, K., Flaus, A., Phelan, M., Kingston, R., Wade, P.A., Lilley, D.M.J., Owen-Hughes, T., 2000. Generation of Superhelical Torsion by ATP-Dependent Chromatin Remodeling Activities. Cell 103, 1133–1142. https://doi.org/10.1016/S0092-8674(00)00215-4

Hedges, D., Proft, M., Entian, K.D., 1995. CAT8, a new zinc cluster-encoding gene necessary for derepression of gluconeogenic enzymes in the yeast Saccharomyces cerevisiae. Mol. Cell. Biol. 15, 1915–1922. https://doi.org/10.1128/MCB.15.4.1915

Hellauer, K., Rochon, M.H., Turcotte, B., 1996. A novel DNA binding motif for yeast zinc cluster proteins: the Leu3p and Pdr3p transcriptional activators recognize everted repeats. Mol. Cell. Biol. 16, 6096–6102. https://doi.org/10.1128/MCB.16.11.6096

Hiesinger, M., Roth, S., Meissner, E., Schüller, H.-J., 2001. Contribution of Cat8 and Sip4 to the transcriptional activation of yeast gluconeogenic genes by carbon source-responsive elements. Curr. Genet. 39, 68–76.

https://doi.org/10.1007/s002940000182

Hohenblum, H., Borth, N., Mattanovich, D., 2003. Assessing viability and cell-associated product of recombinant protein producing Pichia pastoris with flow cytometry. J. Biotechnol. 102, 281–290. https://doi.org/10.1016/S0168-1656(03)00049-X

Huisinga, K.L., Pugh, B.F., 2004. A Genome-Wide Housekeeping Role for TFIID and a Highly Regulated Stress-Related Role for SAGA in Saccharomyces cerevisiae. Mol. Cell 13, 573–585. https://doi.org/10.1016/S1097-2765(04)00087-5

Inan, M., Meagher, M.M., 2001. Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of Pichia pastoris. J. Biosci. Bioeng. 92, 585–589. https://doi.org/10.1016/S1389-1723(01)80321-2

Istrail, S., Davidson, E.H., 2005. Logic functions of the genomic cis-regulatory code. Proc. Natl. Acad. Sci. 102, 4954–4959. https://doi.org/10.1073/pnas.0409624102

Iyer, V., Struhl, K., 1995. Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. EMBO J. 14, 2570–2579. https://doi.org/10.1002/j.1460-2075.1995.tb07255.x

Jansen, A., Verstrepen, K.J., 2011. Nucleosome Positioning in Saccharomyces cerevisiae. Microbiol. Mol. Biol. Rev. 75, 301–320. https://doi.org/10.1128/MMBR.00046-10

Jeppsson, M., Johansson, B., Jensen, P.R., Hahn-Hagerdal, B., Gorwa-Grauslund, M.F., 2003. The level of glucose-6-phosphate dehydrogenase activity strongly influences xylose fermentation and inhibitor sensitivity in recombinant Saccharomyces cerevisiae strains. Yeast 20, 1263–1272.

https://doi.org/10.1002/yea.1043

Kacherovsky, N., Tachibana, C., Amos, E., Fox III, D., Young, E.T., 2008. Promoter Binding by the Adr1 Transcriptional Activator May Be Regulated by Phosphorylation in the DNA-Binding Region. PLoS One 3, e3213.

Karaoglan, M., Karaoglan, F.E., Inan, M., 2016a. Functional analysis of alcohol dehydrogenase (ADH) genes in Pichia pastoris. Biotechnol. Lett. 38, 463–469. https://doi.org/10.1007/s10529-015-1993-z

Karaoglan, M., Karaoglan, F.E., Inan, M., 2016b. Comparison of ADH3 promoter with commonly used promoters for recombinant protein production in Pichia pastoris. Protein Expr. Purif. 121, 112–117. https://doi.org/10.1016/j.pep.2016.01.017

Karpichev, I. V, Luo, Y., Marians, R.C., Small, G.M., 1997. A complex containing two transcription factors regulates peroxisome proliferation and the coordinate induction of beta-oxidation enzymes in Saccharomyces cerevisiae. Mol. Cell. Biol. 17, 69–80. https://doi.org/10.1128/MCB.17.1.69

Kern, A., Hartner, F.S., Freigassner, M., Spielhofer, J., Rumpf, C., Leitner, L., Frohlich, K.-U., Glieder, A., 2007. Pichia pastoris “just in time” alternative respiration. Microbiology 153, 1250–1260. https://doi.org/10.1099/mic.0.2006/001404-0

Kim, J., Iyer, V.R., 2004. Global role of TATA box-binding protein recruitment to promoters in mediating gene expression profiles. Mol. Cell. Biol. 24, 8104–8112. https://doi.org/10.1128/MCB.24.18.8104-8112.2004

Koller, A., Valesco, J., Subramani, S., 2000. The CUP1 promoter of Saccharomyces cerevisiae is inducible by copper in Pichia pastoris. Yeast 16, 651–656. https://doi.org/10.1002/(SICI)1097-0061(200005)16:7<651::AID-

YEA580>3.0.CO;2-F

Komarnitsky, P.B., Klebanow, E.R., Weil, P.A., Denis, C.L., 1998. ADR1-Mediated Transcriptional Activation Requires the Presence of an Intact TFIID Complex. Mol. Cell. Biol. 18, 5861–5867. https://doi.org/10.1128/MCB.18.10.5861

Kostrewa, D., Zeller, M.E., Armache, K.-J., Seizl, M., Leike, K., Thomm, M., Cramer, P., 2009. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. Nature 462, 323–330. https://doi.org/10.1038/nature08548

Koutz, P., Davis, G.R., Stillman, C., Barringer, K., Cregg, J., Thill, G., 1989. Structural comparison of the Pichia pastoris alcohol oxidase genes. Yeast 5, 167–177. https://doi.org/10.1002/yea.320050306

Kranthi, B.V., Kumar, R., Kumar, N.V., Rao, D.N., Rangarajan, P.N., 2009. Identification of key DNA elements involved in promoter recognition by Mxr1p , a master regulator of methanol utilization pathway in Pichia pastoris. Biochim. Biophys. Acta 1789, 460–468. https://doi.org/10.1016/j.bbagrm.2009.05.004

Kuberl, A., Schneider, J., Thallinger, G.G., Anderl, I., Wibberg, D., Hajek, T., Jaenicke, S., Brinkrolf, K., Goesmann, A., Szczepanowski, R., Puhler, A., Schwab, H., Glieder, A., Pichler, H., 2011. High-quality genome sequence of Pichia pastoris CBS7435. J. Biotechnol. 154, 312–320. https://doi.org/10.1016/j.jbiotec.2011.04.014

Kulkarni, M.M., Arnosti, D.N., 2005. cis-Regulatory Logic of Short-Range Transcriptional Repression in Drosophila melanogaster. Mol. Cell. Biol. 25, 3411–3420. https://doi.org/10.1128/MCB.25.9.3411-3420.2005

Kumar, N.V., Rangarajan, P.N., 2012. The Zinc Finger Proteins Mxr1p and Repressor of Phosphoenolpyruvate Carboxykinase (ROP) Have the Same DNA Binding

Specificity but Regulate Methanol Metabolism Antagonistically in Pichia pastoris. J. Biol. Chem. 287, 34465–34473. https://doi.org/10.1074/jbc.M112.365304

Kumar, N.V., Rangarajan, P.N., 2011. Catabolite repression of phosphoenolpyruvate carboxykinase by a zinc finger protein under biotin- and pyruvate carboxylase-deficient conditions in Pichia pastoris. Microbiology 157, 3361–3369. https://doi.org/10.1099/mic.0.053488-0

Kurtzman, C.P., 2009. Biotechnological strains of Komagataella (Pichia) pastoris are Komagataella phaffii as determined from multigene sequence analysis. J. Ind. Microbiol. Biotechnol. 36, 1435–1438. https://doi.org/10.1007/s10295-009-0638-4

Lam, F.H., Steger, D.J., O'Shea, E.K., 2008. Chromatin decouples promoter threshold from dynamic range. Nature 453, 246–250. https://doi.org/10.1038/nature06867

Landes, N., Gasser, B., Vorauer-Uhl, K., Lhota, G., Mattanovich, D., Maurer, M., 2016. The vitamin-sensitive promoter PTHI11 enables pre-defined autonomous induction of recombinant protein production in Pichia pastoris. Biotechnol. Bioeng. 113, 2633–2643. https://doi.org/10.1002/bit.26041

Leavitt, J.M., Tong, A., Tong, J., Pattie, J., Alper, H.S., 2016. Coordinated transcription factor and promoter engineering to establish strong expression elements in Saccharomyces cerevisiae. Biotechnol. J. 11, 866–876. https://doi.org/10.1002/biot.201600029

Lesage, P., Yang, X., Carlson, M., 1996. Yeast SNF1 protein kinase interacts with SIP4, a C6 zinc cluster transcriptional activator: a new role for SNF1 in the glucose response. Mol. Cell. Biol. 16, 1921–1928. https://doi.org/10.1128/MCB.16.5.1921

Li, B., Carey, M., Workman, J.L., 2007. The role of chromatin during transcription. Cell 128, 707–719. https://doi.org/10.1016/j.cell.2007.01.015

Li, Y., Flanagan, P., Tschochner, H., Kornberg, R., 1994. RNA polymerase II initiation factor interactions and transcription start site selection. Science (80-.). 263, 805–807. https://doi.org/10.1126/science.8303296

Liang, S., Zou, C., Lin, Y., Zhang, X., Ye, Y., 2013. Identification and characterization of P GCW14 : a novel, strong constitutive promoter of Pichia pastoris. Biotechnol. Lett. 35, 1865–1871. https://doi.org/10.1007/s10529-013-1265-8

Lin-Cereghino, G.P., Godfrey, L., de la Cruz, B.J., Johnson, S., Khuongsathiene, S., Tolstorukov, I., Yan, M., Lin-Cereghino, J., Veenhuis, M., Subramani, S., Cregg, J.M., 2006. Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in Pichia pastoris. Mol. Cell. Biol. 26, 883–897. https://doi.org/10.1128/MCB.26.3.883-897.2006

Liu, G., Bergenholm, D., Nielsen, J., 2016. Genome-Wide Mapping of Binding Sites Reveals Multiple Biological Functions of the Transcription Factor Cst6p in Saccharomyces cerevisiae. MBio 7, e00559-16. https://doi.org/10.1128/mBio.00559-16

Liu, H., Tan, X., Russell, K.A., Veenhuis, M., Cregg, J.M., 1995. PER3 , a Gene Required for Peroxisome Biogenesis in Pichia pastoris , Encodes a Peroxisomal Membrane Protein Involved in Protein Import. J. Biol. Chem. 270, 10940–10951. https://doi.org/10.1074/jbc.270.18.10940

Love, K.R., Shah, K.A., Whittaker, C.A., Wu, J., Bartlett, M.C., Ma, D., Leeson, R.L., Priest, M., Borowsky, J., Young, S.K., Love, J.C., 2016. Comparative genomics and transcriptomics of Pichia pastoris. BMC Genomics 17, 550.

https://doi.org/10.1186/s12864-016-2876-y

Macauley-Patrick, S., Fazenda, M.L., McNeil, B., Harvey, L.M., 2005. Heterologous protein production using the Pichia pastoris expression system. Yeast 22, 249–270. https://doi.org/10.1002/yea.1208

Madzak, C., 2015. Yarrowia lipolytica: recent achievements in heterologous protein expression and pathway engineering. Appl. Microbiol. Biotechnol. 99, 4559–4577. https://doi.org/10.1007/s00253-015-6624-z

Madzak, C., Treton, B., Blanchin-Roland, S., 2000. Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast Yarrowia lipolytica. J. Mol. Microbiol. Biotechnol. 2, 207–216.

Malik, S., Roeder, R.G., 2000. Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. Trends Biochem. Sci. 25, 277–283. https://doi.org/10.1016/S0968-0004(00)01596-6

Massahi, A., Çalık, P., 2018. Naturally occurring novel promoters around pyruvate branch-point for recombinant protein production in Pichia pastoris (Komagataella phaffii): Pyruvate decarboxylase- and pyruvate kinase-promoters. Biochem. Eng. J. 138, 111–120. https://doi.org/10.1016/j.bej.2018.07.012

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

Mattanovich, D., Graf, A., Stadlmann, J., Dragosits, M., Redl, A., Maurer, M., Kleinheinz, M., Sauer, M., Altmann, F., Gasser, B., 2009. Genome, secretome and glucose transport highlight unique features of the protein production host

Pichia pastoris. Microb. Cell Fact. 8, 29. https://doi.org/10.1186/1475-2859-8-29

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., Pugh, B.F., 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res. 18, 1073–1083. https://doi.org/10.1101/gr.078261.108

Mehrotra, R., Renganaath, K., Kanodia, H., Loake, G.J., Mehrotra, S., 2017. Towards combinatorial transcriptional engineering. Biotechnol. Adv. 35, 390–405. https://doi.org/10.1016/j.biotechadv.2017.03.006

Menendez, J., Valdes, I., Cabrera, N., 2003. The ICL1 gene of Pichia pastoris, transcriptional regulation and use of its promoter. Yeast 20, 1097–1108. https://doi.org/10.1002/yea.1028

Mogno, I., Vallania, F., Mitra, R.D., Cohen, B.A., 2010. TATA is a modular component of synthetic promoters. Genome Res. 20, 1391–1397. https://doi.org/10.1101/gr.106732.110

Naatsaari, L., Mistlberger, B., Ruth, C., Hajek, T., Hartner, F.S., Glieder, A., 2012. Deletion of the Pichia pastoris KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One 7, e39720. https://doi.org/10.1371/journal.pone.0039720

Nevoigt, E., Fischer, C., Mucha, O., Matthaus, F., Stahl, U., Stephanopoulos, G., 2007. Engineering promoter regulation. Biotechnol. Bioeng. 96, 550–558. https://doi.org/10.1002/bit.21129

Nevoigt, E., Kohnke, J., Fischer, C.R., Alper, H., Stahl, U., Stephanopoulos, G., 2006. Engineering of promoter replacement cassettes for fine-tuning of gene expression in Saccharomyces cerevisiae. Appl. Environ. Microbiol. 72, 5266–

5273. https://doi.org/10.1128/AEM.00530-06

Ogata, K., Nishikawa, H., Ohsugi, M., 1969. A yeast capable of utilizing methanol. Agric. Biol. Chem. 33, 1519–1520. https://doi.org/10.1080/00021369.1969.10859497

Özcan, S., Leong, T., Johnston, M., 1996. Rgt1p of Saccharomyces cerevisiae, a key regulator of glucose-induced genes, is both an activator and a repressor of transcription. Mol. Cell. Biol. 16, 6419–6426. https://doi.org/10.1128/MCB.16.11.6419

Parua, P.K., Dombek, K.M., Young, E.T., 2014. Yeast 14-3-3 Protein Functions as a Comodulator of Transcription by Inhibiting Coactivator Functions. J. Biol. Chem. 289, 35542–35560. https://doi.org/10.1074/jbc.M114.592287

Parua, P.K., Ratnakumar, S., Braun, K.A., Dombek, K.M., Arms, E., Ryan, P.M., Young, E.T., 2010. 14-3-3 (Bmh) proteins inhibit transcription activation by Adr1 through direct binding to its regulatory domain. Mol. Cell. Biol. 30, 5273–5283. https://doi.org/10.1128/MCB.00715-10

Passoth, V., Hahn–Hägerdal, B., 2000. Production of a heterologous endo-1,4-β-xylanase in the yeast Pichia stipitis with an O2-regulated promoter. Enzyme Microb. Technol. 26, 781–784. https://doi.org/10.1016/S0141-0229(00)00171-X

Payne, W.E., Gannon, P.M., Kaiser, C.A., 1995. An inducible acid phosphatase from the yeast Pichia pastoris: characterization of the gene and its product. Gene 163, 19–26. https://doi.org/10.1016/0378-1119(95)00379-K

Periyasamy, S., Govindappa, N., Sreenivas, S., Sastry, K., 2013. Isolation, characterization and evaluation of the Pichia pastoris sorbitol dehydrogenase promoter for expression of heterologous proteins. Protein Expr. Purif. 92, 128–

133. https://doi.org/10.1016/j.pep.2013.09.008

Petrenko, N., Jin, Y., Wong, K.H., Struhl, K., 2017. Evidence that Mediator is essential for Pol II transcription, but is not a required component of the preinitiation complex in vivo. Elife 6, e28447. https://doi.org/10.7554/eLife.28447

Polupanov, A.S., Nazarko, V.Y., Sibirny, A.A., 2012. Gss1 protein of the methylotrophic yeast Pichia pastoris is involved in glucose sensing, pexophagy and catabolite repression. Int. J. Biochem. Cell Biol. 44, 1906–1918. https://doi.org/10.1016/j.biocel.2012.07.017

Porro, D., Sauer, M., Branduardi, P., Mattanovich, D., 2005. Recombinant protein production in yeasts. Mol. Biotechnol. 31, 245–259. https://doi.org/10.1385/MB:31:3:245

Portela, R.M.C., Vogl, T., Kniely, C., Fischer, J.E., Oliveira, R., Glieder, A., 2017. Synthetic core promoters as universal parts for fine-tuning expression in different yeast species. ACS Synth. Biol. 6, 471–484. https://doi.org/10.1021/acssynbio.6b00178

Prielhofer, R., Barrero, J.J., Steuer, S., Gassler, T., Zahrl, R., Baumann, K., Sauer, M., Mattanovich, D., Gasser, B., Marx, H., 2017. GoldenPiCS: a Golden Gate-derived modular cloning system for applied synthetic biology in the yeast Pichia pastoris. BMC Syst. Biol. 11, 123. https://doi.org/10.1186/s12918-017-0492-3

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

Proft, M., Grzesitza, D., Entian, K.D., 1995. Identification and characterization of

regulatory elements in the phosphoenolpyruvate carboxykinase gene PCK1 of Saccharomyces cerevisiae. Mol. Gen. Genet. 246, 367–373.

Ptashne, M., Gann, A., 1997. Transcriptional activation by recruitment. Nature 386, 569–577. https://doi.org/10.1038/386569a0

Pugh, B.F., Tjian, R., 1991. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes Dev. 5, 1935–1945.

Qi, K., Zhong, J.-J., Xia, X.-X., 2014. Triggering Respirofermentative Metabolism in the Crabtree-Negative Yeast Pichia guilliermondii by Disrupting the CAT8 Gene. Appl. Environ. Microbiol. 80, 3879–3887. https://doi.org/10.1128/AEM.00854-14

Qin, X., Qian, J., Yao, G., Zhuang, Y., Zhang, S., Chu, J., 2011. GAP Promoter Library for Fine-Tuning of Gene Expression in Pichia pastoris . Appl. Environ. Microbiol. 77, 3600–3608. https://doi.org/10.1128/AEM.02843-10

Rahner, A., Hiesinger, M., Schuller, H.-J., 1999. Deregulation of gluconeogenic structural genes by variants of the transcriptional activator Cat8p of the yeast Saccharomyces cerevisiae. Mol. Microbiol. 34, 146–156. https://doi.org/10.1046/j.1365-2958.1999.01588.x

Rahner, A., Scholer, A., Martens, E., Gollwitzer, B., Schuller, H.-J., 1996. Dual Influence of the Yeast Catlp (Snflp) Protein Kinase on Carbon Source-Dependent Transcriptional Activation of Gluconeogenic Genes by the Regulatory Gene CAT8. Nucleic Acids Res. 24, 2331–2337. https://doi.org/10.1093/nar/24.12.2331

Ramírez, M.A., Lorenz, M.C., 2009. The Transcription Factor Homolog CTF1 Regulates β-Oxidation in Candida albicans . Eukaryot. Cell 8, 1604–1614. https://doi.org/10.1128/EC.00206-09

Randez-Gil, F., Bojunga, N., Proft, M., Entian, K.D., 1997. Glucose derepression of gluconeogenic enzymes in Saccharomyces cerevisiae correlates with phosphorylation of the gene activator Cat8p. Mol. Cell. Biol. 17, 2502–2510. https://doi.org/10.1128/MCB.17.5.2502

Raser, J.M., 2004. Control of Stochasticity in Eukaryotic Gene Expression. Science (80-.). 304, 1811–1814. https://doi.org/10.1126/science.1098641

Ratnakumar, S., Kacherovsky, N., Arms, E., Young, E.T., 2009. Snf1 Controls the Activity of Adr1 Through Dephosphorylation of Ser230. Genetics 182, 735–745. https://doi.org/10.1534/genetics.109.103432

Ratnakumar, S., Young, E.T., 2010. Snf1 Dependence of Peroxisomal Gene Expression Is Mediated by Adr1. J. Biol. Chem. 285, 10703–10714. https://doi.org/10.1074/jbc.M109.079848

Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A., Segal, E., 2012. Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nat. Genet. 44, 743–750. https://doi.org/10.1038/ng.2305

Redden, H., Alper, H.S., 2015. The development and characterization of synthetic minimal yeast promoters. Nat. Commun. 6, 7810.

Reece, R.J., Ptashne, M., 1993. Determinants of binding-site specificity among yeast C6 zinc cluster proteins. Science 261, 909–911.

Reed, S.H., Akiyama, M., Stillman, B., Friedberg, E.C., 1999. Yeast autonomously replicating sequence binding factor is involved in nucleotide excision repair. Genes Dev. 13, 3052–3058.

Richmond, T.J., Davey, C.A., 2003. The structure of DNA in the nucleosome core.

Nature 423, 145.

Roth, S., Kumme, J., Schuller, H.-J., 2004. Transcriptional activators Cat8 and Sip4 discriminate between sequence variants of the carbon source-responsive promoter element in the yeast Saccharomyces cerevisiae. Curr. Genet. 45, 121–128. https://doi.org/10.1007/s00294-003-0476-2

Roth, S., Schüller, H.J., 2001. Cat8 and Sip4 mediate regulated transcriptional activation of the yeast malate dehydrogenase gene MDH2 by three carbon source-responsive promoter elements. Yeast 18, 151–162. https://doi.org/10.1002/1097-0061(20010130)18:2<151::AID-YEA662>3.0.CO;2-Q

Ruchala, J., Kurylenko, O.O., Soontorngun, N., Dmytruk, K. V, Sibirny, A.A., 2017. Transcriptional activator Cat8 is involved in regulation of xylose alcoholic fermentation in the thermotolerant yeast Ogataea (Hansenula) polymorpha. Microb. Cell Fact. 16, 36. https://doi.org/10.1186/s12934-017-0652-6

Rußmayer, H., Buchetics, M., Gruber, C., Valli, M., Grillitsch, K., Modarres, G., Guerrasio, R., Klavins, K., Neubauer, S., Drexler, H., Steiger, M., Troyer, C., Al Chalabi, A., Krebiehl, G., Sonntag, D., Zellnig, G., Daum, G., Graf, A.B., Altmann, F., Koellensperger, G., Hann, S., Sauer, M., Mattanovich, D., Gasser, B., 2015. Systems-level organization of yeast methylotrophic lifestyle. BMC Biol. 13, 80. https://doi.org/10.1186/s12915-015-0186-5

Sahu, U., Krishna Rao, K., Rangarajan, P.N., 2014. Trm1p, a Zn(II)(2)Cys(6)-type transcription factor, is essential for the transcriptional activation of genes of methanol utilization pathway, in Pichia pastoris. Biochem. Biophys. Res. Commun. 451, 158–164. https://doi.org/10.1016/j.bbrc.2014.07.094

Saleem, R.A., Rogers, R.S., Ratushny, A. V, Dilworth, D.J., Shannon, P.T.,

Shteynberg, D., Wan, Y., Moritz, R.L., Nesvizhskii, A.I., Rachubinski, R.A., Aitchison, J.D., 2010. Integrated Phosphoproteomics Analysis of a Signaling Network Governing Nutrient Response and Peroxisome Induction. Mol. Cell. Proteomics 9, 2076–2088. https://doi.org/10.1074/mcp.M000116-MCP201

Schjerling, P., Holmberg, S., 1996. Comparative amino acid sequence analysis of the C6 zinc cluster family of transcriptional regulators. Nucleic Acids Res. 24, 4599–4607. https://doi.org/10.1093/nar/24.23.4599

Schöler, A., Schüller, H.J., 1994. A carbon source-responsive promoter element necessary for activation of the isocitrate lyase gene ICL1 is common to genes of the gluconeogenic pathway in the yeast Saccharomyces cerevisiae. Mol. Cell. Biol. 14, 3613–3622. https://doi.org/10.1128/MCB.14.6.3613

Schotte, P., Dewerte, I., De Groeve, M., De Keyser, S., De Brabandere, V., Stanssens, P., 2016. Pichia pastoris Mut(S) strains are prone to misincorporation of O-methyl-l-homoserine at methionine residues when methanol is used as the sole carbon source. Microb. Cell Fact. 15, 98. https://doi.org/10.1186/s12934-016-0499-2

Sears, I.B., O'Connor, J., Rossanese, O.W., Glick, B.S., 1998. A versatile set of vectors for constitutive and regulated gene expression in Pichia pastoris. Yeast 14, 783–790. https://doi.org/10.1002/(SICI)1097-0061(19980615)14:8<783::AID-YEA272>3.0.CO;2-Y

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

Sela, I., Lukatsky, D.B., 2011. DNA Sequence Correlations Shape Nonspecific Transcription Factor-DNA Binding Affinity. Biophys. J. 101, 160–166.

https://doi.org/10.1016/j.bpj.2011.04.037

Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., Segal, E., 2012. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat. Biotechnol. 30, 521. https://doi.org/10.1038/nbt.2205

Shen, S., Sulter, G., Jeffries, T.W., Cregg, J.M., 1998. A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast Pichia pastoris. Gene 216, 93–102. https://doi.org/10.1016/S0378-1119(98)00315-1

Shen, W.-C., 2003. Systematic analysis of essential yeast TAFs in genome-wide transcription and preinitiation complex assembly. EMBO J. 22, 3395–3402. https://doi.org/10.1093/emboj/cdg336

Shen, W., Xue, Y., Liu, Y., Kong, C., Wang, X., Huang, M., Cai, M., Zhou, X., Zhang, Y., Zhou, M., 2016. A novel methanol-free Pichia pastoris system for recombinant protein expression. Microb. Cell Fact. 15, 178. https://doi.org/10.1186/s12934-016-0578-4

Shi, L., Wang, X., Wang, J., Zhang, P., Qi, F., Cai, M., Zhang, Y., Zhou, X., 2018. Transcriptome analysis of Δmig1Δmig2 mutant reveals their roles in methanol catabolism, peroxisome biogenesis and autophagy in methylotrophic yeast Pichia pastoris. Genes Genomics 40, 399–412. https://doi.org/10.1007/s13258-017-0641-5

Sinha, J., Plantz, B.A., Inan, M., Meagher, M.M., 2005. Causes of proteolytic degradation of secreted recombinant proteins produced in methylotrophic yeast Pichia pastoris: case study with recombinant ovine interferon-tau. Biotechnol. Bioeng. 89, 102–112. https://doi.org/10.1002/bit.20318

Sloan, J.S., Dombek, K.M., Young, E.T., 1999. Post-translational Regulation of Adr1

Activity Is Mediated by Its DNA Binding Domain. J. Biol. Chem. 274, 37575–37582. https://doi.org/10.1074/jbc.274.53.37575

Smale, S.T., Kadonaga, J.T., 2003. The RNA Polymerase II Core Promoter. Annu. Rev. Biochem. 72, 449–479. https://doi.org/10.1146/annurev.biochem.72.121801.161520

Soontorngun, N., Larochelle, M., Drouin, S., Robert, F., Turcotte, B., 2007. Regulation of gluconeogenesis in Saccharomyces cerevisiae is mediated by activator and repressor functions of Rds2. Mol. Cell. Biol. 27, 7895–7905. https://doi.org/10.1128/MCB.01055-07

Soutourina, J., 2017. Transcription regulation by the Mediator complex. Nat. Rev. Mol. Cell Biol. 19, 262–274. https://doi.org/10.1038/nrm.2017.115

Stadlmayr, G., Mecklenbrauker, A., Rothmuller, M., Maurer, M., Sauer, M., Mattanovich, D., Gasser, B., 2010. Identification and characterisation of novel Pichia pastoris promoters for heterologous protein production. J. Biotechnol. 150, 519–529. https://doi.org/10.1016/j.jbiotec.2010.09.957

Stewart, J.J., Stargell, L.A., 2001. The stability of the TFIIA-TBP-DNA complex is dependent on the sequence of the TATAAA element. J. Biol. Chem. 276, 30078–30084. https://doi.org/10.1074/jbc.M105276200

Struhl, K., 1995. Yeast transcriptional regulatory mechanisms. Annu. Rev. Genet. 29, 651–674. https://doi.org/10.1146/annurev.ge.29.120195.003251

Struhl, K., Segal, E., 2013. Determinants of nucleosome positioning. Nat. Struct. Mol. Biol. 20, 267–273. https://doi.org/10.1038/nsmb.2506

Sturmberger, L., Chappell, T., Geier, M., Krainer, F., Day, K.J., Vide, U., Trstenjak, S., Schiefer, A., Richardson, T., Soriaga, L., Darnhofer, B., Birner-Gruenberger,

R., Glick, B.S., Tolstorukov, I., Cregg, J., Madden, K., Glieder, A., 2016. Refined Pichia pastoris reference genome sequence. J. Biotechnol. 235, 121–131. https://doi.org/10.1016/j.jbiotec.2016.04.023

Tachibana, C., Biddick, R., Law, G.L., Young, E.T., 2007. A poised initiation complex is activated by SNF1. J. Biol. Chem. 282, 37308–37315. https://doi.org/10.1074/jbc.M707363200

Tachibana, C., Yoo, J.Y., Tagne, J.-B., Kacherovsky, N., Lee, T.I., Young, E.T., 2005. Combined Global Localization Analysis and Transcriptome Data Identify Genes That Are Directly Coregulated by Adr1 and Cat8. Mol. Cell. Biol. 25, 2138–2146. https://doi.org/10.1128/MCB.25.6.2138

Teixeira, M.C., Monteiro, P.T., Palma, M., Costa, C., Godinho, C.P., Pais, P., Cavalheiro, M., Antunes, M., Lemos, A., Pedreira, T., Sá-Correia, I., 2018. YEASTRACT: An upgraded database for the analysis of transcription regulatory networks in Saccharomyces cerevisiae. Nucleic Acids Res. 46, D348–D353. https://doi.org/10.1093/nar/gkx842

Thomas, M.C., Chiang, C.-M., 2006. The general transcription machinery and general cofactors. Crit. Rev. Biochem. Mol. Biol. 41, 105–178. https://doi.org/10.1080/10409230600648736

Thompson, C.A., 2010. FDA approves kallikrein inhibitor to treat hereditary angioedema. Am. J. Health. Syst. Pharm. https://doi.org/10.2146/news100005

Thukral, S.K., Eisen, A., Young, E.T., 1991. Two monomers of yeast transcription factor ADR1 bind a palindromic sequence symmetrically to activate ADH2 expression. Mol. Cell. Biol. 11, 1566–1577. https://doi.org/10.1128/MCB.11.3.1566

Thukral, S.K., Tavianini, M.A., Blumberg, H., Young, E.T., 1989. Localization of a

minimal binding domain and activation regions in yeast regulatory protein ADR1. Mol. Cell. Biol. 9, 2360–2369. https://doi.org/10.1128/MCB.9.6.2360

Tirosh, I., Weinberger, A., Carmi, M., Barkai, N., 2006. A genetic signature of interspecies variations in gene expression. Nat. Genet. 38, 830.

Tschopp, J.F., Brust, P.F., Cregg, J.M., Stillman, C.A., Gingeras, T.R., 1987. Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris. Nucleic Acids Res. 15, 3859–3876. https://doi.org/10.1093/nar/15.9.3859

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

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

van Heusden, G.P.H., Steensma, H.Y., 2006. Yeast 14-3-3 proteins. Yeast 23, 159–171. https://doi.org/10.1002/yea.1338

van Werven, F.J., van Teeffelen, H.A.A.M., Holstege, F.C.P., Timmers, H.T.M., 2009. Distinct promoter dynamics of the basal transcription factor TBP across the yeast genome. Nat. Struct. Mol. Biol. 16, 1043–1048. https://doi.org/10.1038/nsmb.1674

Verdone, L., 2002. Hyperacetylation of chromatin at the ADH2 promoter allows Adr1 to bind in repressed conditions. EMBO J. 21, 1101–1111. https://doi.org/10.1093/emboj/21.5.1101

Vervecken, W., Kaigorodov, V., Callewaert, N., Geysens, S., De Vusser, K., Contreras, R., 2004. In Vivo Synthesis of Mammalian-Like, Hybrid-Type N-Glycans in Pichia pastoris. Appl. Environ. Microbiol. 70, 2639–2646. https://doi.org/10.1128/AEM.70.5.2639-2646.2004

Vogl, T., Ruth, C., Pitzer, J., Kickenweiz, T., Glieder, A., 2014. Synthetic core promoters for Pichia pastoris. ACS Synth. Biol. 3, 188–191. https://doi.org/10.1021/sb400091p

Vogl, T., Sturmberger, L., Fauland, P.C., Hyden, P., Fischer, J.E., Schmid, C., Thallinger, G.G., Geier, M., Glieder, A., 2018. Methanol independent induction in Pichia pastoris by simple derepressed overexpression of single transcription factors. Biotechnol. Bioeng. 115, 1037–1050. https://doi.org/10.1002/bit.26529

Vogl, T., Sturmberger, L., Kickenweiz, T., Wasmayer, R., Schmid, C., Hatzl, A.-M., Gerstmann, M.A., Pitzer, J., Wagner, M., Thallinger, G.G., Geier, M., Glieder, A., 2016. A Toolbox of Diverse Promoters Related to Methanol Utilization: Functionally Verified Parts for Heterologous Pathway Expression in Pichia pastoris. ACS Synth. Biol. 5, 172–186. https://doi.org/10.1021/acssynbio.5b00199

von Hippel, P.H., 2007. From “Simple” DNA-Protein Interactions to the Macromolecular Machines of Gene Expression. Annu. Rev. Biophys. Biomol. Struct. 36, 79–105. https://doi.org/10.1146/annurev.biophys.34.040204.144521

von Hippel, P.H., Berg, O.G., 1986. On the specificity of DNA-protein interactions. Proc. Natl. Acad. Sci. 83, 1608–1612. https://doi.org/10.1073/pnas.83.6.1608

Walther, K., Schüller, H., 2001. Adr1 and Cat8 synergistically activate the glucose-regulated alcohol dehydrogenase gene ADH2 of the yeast Saccharomyces cerevisiae. Microbiology 147, 2037–2044. https://doi.org/10.1099/00221287-

147-8-2037

Wang, J.-P., Fondufe-Mittendorf, Y., Xi, L., Tsai, G.-F., Segal, E., Widom, J., 2008. Preferentially quantized linker DNA lengths in Saccharomyces cerevisiae. PLoS Comput. Biol. 4, e1000175. https://doi.org/10.1371/journal.pcbi.1000175

Wang, J., Wang, X., Shi, L., Qi, F., Zhang, P., Zhang, Y., Zhou, X., Song, Z., Cai, M., 2017. Methanol-Independent Protein Expression by AOX1 Promoter with trans-Acting Elements Engineering and Glucose-Glycerol-Shift Induction in Pichia pastoris. Sci. Rep. 7, 41850. https://doi.org/10.1038/srep41850

Wang, X., Cai, M., Shi, L., Wang, Q., Zhu, J., Wang, J., Zhou, M., Zhou, X., Zhang, Y., 2016a. PpNrg1 is a transcriptional repressor for glucose and glycerol repression of AOX1 promoter in methylotrophic yeast Pichia pastoris. Biotechnol. Lett. 38, 291–298. https://doi.org/10.1007/s10529-015-1972-4

Wang, X., Wang, Q., Wang, J., Bai, P., Shi, L., Shen, W., Zhou, M., Zhou, X., Zhang, Y., Cai, M., 2016b. Mit1 Transcription Factor Mediates Methanol Signaling and Regulates the Alcohol Oxidase 1 (AOX1) Promoter in Pichia pastoris. J. Biol. Chem. 291, 6245–6261. https://doi.org/10.1074/jbc.M115.692053

Watanabe, D., Hashimoto, N., Mizuno, M., Zhou, Y., Akao, T., Shimoi, H., 2013. Accelerated alcoholic fermentation caused by defective gene expression related to glucose derepression in Saccharomyces cerevisiae. Biosci. Biotechnol. Biochem. 77, 2255–2262. https://doi.org/10.1271/bbb.130519

Waterham, H.R., Digan, M.E., Koutz, P.J., Lair, S. V, Cregg, J.M., 1997. Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene 186, 37–44. https://doi.org/10.1016/S0378-1119(96)00675-0

West, R.W., Yocum, R.R., Ptashne, M., 1984. Saccharomyces cerevisiae GAL1-

GAL10 divergent promoter region: location and function of the upstream activating sequence UASG. Mol. Cell. Biol. 4, 2467–2478. https://doi.org/10.1128/MCB.4.11.2467

Workman, J.L., Kingston, R.E., 1998. Alteration of nucleosome structure as a mechanism of transcriptional regulation. Annu. Rev. Biochem. 67, 545–579. https://doi.org/10.1146/annurev.biochem.67.1.545

Xi, L., Fondufe-Mittendorf, Y., Xia, L., Flatow, J., Widom, J., Wang, J.-P., 2010. Predicting nucleosome positioning using a duration Hidden Markov Model. BMC Bioinformatics 11, 346. https://doi.org/10.1186/1471-2105-11-346

Xuan, Y., Zhou, X., Zhang, W., Zhang, X., Song, Z., Zhang, Y., 2009. An upstream activation sequence controls the expression of AOX1 gene in Pichia pastoris. FEMS Yeast Res. 9, 1271–1282. https://doi.org/10.1111/j.1567-1364.2009.00571.x

Young, E.T., Dombek, K.M., Tachibana, C., Ideker, T., 2003. Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. J. Biol. Chem. 278, 26146–26158. https://doi.org/10.1074/jbc.M301981200

Young, E.T., Kacherovsky, N., Van Riper, K., 2002. Snf1 Protein Kinase Regulates Adr1 Binding to Chromatin but Not Transcription Activation. J. Biol. Chem. 277, 38095–38103. https://doi.org/10.1074/jbc.M206158200

Yudkovsky, N., Ranish, J.A., Hahn, S., 2000. A transcription reinitiation intermediate that is stabilized by activator. Nature 408, 225.

Zampar, G.G., Kummel, A., Ewald, J., Jol, S., Niebel, B., Picotti, P., Aebersold, R., Sauer, U., Zamboni, N., Heinemann, M., 2013. Temporal system-level organization of the switch from glycolytic to gluconeogenic operation in yeast.

Mol. Syst. Biol. 9, 651. https://doi.org/10.1038/msb.2013.11

Zaret, K.S., Carroll, J.S., 2011. Pioneer transcription factors: establishing competence for gene expression. Genes Dev. 25, 2227–2241. https://doi.org/10.1101/gad.176826.111

Zhan, C., Wang, S., Sun, Y., Dai, X., Liu, X., Harvey, L., McNeil, B., Yang, Y., Bai, Z., 2016. The Pichia pastoris transmembrane protein GT1 is a glycerol transporter and relieves the repression of glycerol on AOX1 expression. FEMS Yeast Res. 16, fow033. https://doi.org/10.1093/femsyr/fow033

Zhan, C., Yang, Y., Zhang, Z., Li, X., Liu, X., Bai, Z., 2017. Transcription factor Mxr1 promotes the expression of Aox1 by repressing glycerol transporter 1 in Pichia pastoris. FEMS Yeast Res. 17. https://doi.org/10.1093/femsyr/fox015

Zhang, P., Zhang, W., Zhou, X., Bai, P., Cregg, J.M., Zhang, Y., 2010. Catabolite repression of Aox in Pichia pastoris is dependent on hexose transporter PpHxt1 and pexophagy. Appl. Environ. Microbiol. 76, 6108–6118. https://doi.org/10.1128/AEM.00607-10

Zhang, Y., Liu, R., Wu, X., 2007. The proteolytic systems and heterologous proteins degradation in the methylotrophic yeastPichia pastoris . Ann. Microbiol. 57, 553. https://doi.org/10.1007/BF03175354

APPENDIX A

The Sequence of P_{ADH2} variants

Modified positions are represented with bold and underlined characters.

> $P_{ADH2\text{-}wt}$

TCCTTTTTACCACCCAAGTGCGAGTGAAACACCCCATGGCTGCTCTCCGA
TTGCCCCTCTACAGGCATAAGGGTGTGACTTTGTGGGCTTGAATTTTACA
CCCCCTCCAACTTTTCTCGCATCAATTGATCCTGTTACCAATATTGCATG
CCCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATCGCAG
GGTGAGACTGTAGAGACCCCACATAGTGACAATGATTATGTAAGAAGAG
GGGGGTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTGAACA
ATGCCCAGCAGTCCTCCCCACTCTTTGACAAATCAGTATCACCGATTAAC
ACCCCAAATCTTATTCTCAACGGTCCCTCATCCTTGCACCCCTCTTTGGA
CAAATGGCAGTTAGCATTGGTGCACTGACTGACTGCCCAACCTTAAACC
CAAATTTCTTAGAAGGGGCCCATCTAGTTAGCGAGGGGTGAAAAATTCC
TCCATCGGAGATGTATTGACCGTAAGTTGCTGCTTAAAAAAAATCAGTT
CAGATAGCGAGACTTTTTTGATTTCGCAACGGGAGTGCCTGTTCCATTCG
ATTGCAATTCTCACCCCTTCTGCCCAGTCCTGCCAATTGCCCATGAATCT
GCTAATTTCGTTGATTCCCACCCCCCTTTCCAACTCCACAAATTGTCCAA
TCTCGTTTTCCATTTGGGAGAATCTGCATGTCGACTACATAAAGCGACCG
GTGTCCGAAAAGATCTGTGTAGTTTTCAACATTTTGTGCTCCCCCCGCTG
TTTGAAAACGGGGGTGAGCGCTCTCCGGGGTGCGAATTCGTGCCCAATT
CCTTTCACCCTGCCTATTGTAGACGTCAACCCGCATCTGGTGCGAATATA
GCGCACCCCCAATGATCACACCAACAATTGGTCCACCCCTCCCCAATCTC
TAATATTCACAATTCACCTCACTATAAATACCCCTGTCCTGCTCCCAAAT
TCTTTTTTCCTTCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTACGA
AA

>P*ADH2-NucOpt*

TCCTTTTTACCACC**T**AAGTGCGAGTGAAACACCC**T**ATGGCTGCTCTCCGA
TTGCCCCTCTACAGGCATAAGGGTGTGA**T**TTT**T**T**TTTT**TT**T**AATTTTACA
CCCCCTCCAACTTTT**T**TCGC**G**T**A**AATTGATCCTGTTACCAATATTGCATG
CCCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATCGCAG
GGT**A**A**A**A**A**T**A**TA**T**AGACCCCACA**A**A**AAA**A**A**AATGATTATGTAAGAAGA
GGGGGGTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTGAA**A**
AATGCCCAGC**TT**T**TT**TCCC**T**A**T**TCTTTGACAAATCAGTATCAC**TT**ATTAA
CACCCCAAAT**T**TT**T**TTCTCAACGGTCCCTCATCCTTGCACCCCTCTTTGG
ACAAATGGCAGTTAG**T**ATT**A**GTGCACTGACTGACTGCC**T**AACCTTAAAC
CC**T**AATTTCTTAGAAGGGGCCCAT**A**TAGTTAGCGAGGGGTGAAAAATTC
CTCCATCGGAGATGTATT**A**ACCGTAA**T**TT**TT**T**TT**TTAAAAAAAAA**AA**A**A**T
TCAGATAGCGA**A**A**T**TTTTTTGATTTCGC**G**ACG**C**G**C**GT**TTT**T**T**TT**TTT**TT**T**
TTTT**TTTT**TTCTCACCCCTTCTGCCCAGT**T**CTGCCAATTGCCCATGAATCT
ACTAATTTCGTTGATTCCCACCCCCCTTTCCAACTCCA**A**AAATTT**T**T**TT**AA
T**T**T**TT**TTTTT**TTT**TTTGGGAGAATCTG**A**ATGT**AT**A**T**TACATAAAGCGACCG
GTGTCCGAAAA**A**AT**T**T**T**T**T**T**TT**TTTTT**T**AA**TT**TTTTT**T**T**TT**T**T**CCCCCGCT**T**T
TT**A**AAAACGGGGGT**A**AGCGCTCTCCGGGGTGCGAATTCG**C**GCCC**T**ATTC
CTTTCACCCTGCCTATTGTAGACGTCAACCCGCATCTGGTGCGAATATAG
CGCACCCCCAATGATCACACCAACAATTGGTCCACCCCTCCCCAATCTCT
AATATTCACAATTCACCTCACTATAAATACCCCTGTCCTGCTCCCAAATT
CTTTTTTCCTTCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTACGA
AA

>P*ADH2-AddCat2(OptCat)*

TCCTTTTTACCACCCAAGTGCGAGTGAAACACCCCATGGCTGCTCTCCGA
TTGCCCCTCTACAGGCATAAGGGTGTGACTTTGTGGGCTTGAATTTTACA
CCCCCTCCAACTTTTCTCGCATCAATTGATCCTGTTACCAATATTGCATG
CCCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATCGCAG
GGTGAGACTGTAGAGACCCCACATAGTGACAATGATTATGTAAGAAGAG
GGGGGTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTGAACA
ATGCCCAGCAGTCCTCCCCACTCTTTGACAAATCAGTATCACCGATTAAC
ACCCCAAATCTTATTCTCAACGGTCCCTCATCCTTGCACCCCTCTTTGGA
CAAATGGCAGTTAGCATTGGTGCACTGACTGACTGCCCAACCTTAAACC
CAAATTTCTTAGAAGGGGCCCATCTAGTTAGCGAGGGGTGAAAAATTCC
TCCATCGGAGATGTATTGACCGTAAGTTGCTGCTTAAAAAAAATCAGTT
CAGATAGCGAGACTTTTTTGATTTCGCAACGGGAGTGCCTGTTCCATTCG
ATTGCAATTCTCACCCCTTCTGCCCAGTCCTGCCAATTGCCCATGAATCT
GCTAATTTCGTTGATTCCCACCCCCCTTTCCAACTCCACAAATTGTCCAA
TCTCGTTTTCCATTTGGGAGAATCTGCATGTCGACTACATAAAG**TTC**C**GT**
TCGTCCGAAAAGATCTGTGTAGTTTTCAACATTTTGTGCTCCCCCCGCTG
TTTGAAAACGGGGGTGAGCGCTCTCCGGGGTGCGAATTCGTGCCCAATT
CCTTTCACCCTGCCTATTGTAGACGTCAACCCGCATCTGGTGCGAATATA
GCGCACCCCCAATGATCACACCAACAATTGGTCCACCCCTCCCCAATCTC
TAATATTCACAATTCACCTCACTATAAATACCCCTGTCCTGCTCCCAAAT
TCTTTTTTCCTTCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTACGA
AA

>P*ADH2-AddCat1(AddCat)*

TCCTTTTTACCACCCAAGTGCGAGTGAAACACCCCATGGCTGCTCTCCGA
TTGCCCCTCTACAGGCATAAGGGTGTGACTTTGTGGGCTTGAATTTTACA
CCCCCTCCAACTTTTCTCGCATCAATTGATCCTGTTACCAATATTGCATG
CCCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATCGCAG
GGTGAGACTGTAGAGACCCCACATAGTGACAATGATTATGTAAGAAGAG
GGGGGTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTGAACA
ATGCCCAGCAGTCCTCCCCACTCTTTGACAAATCAGTATCACCGATTAAC
ACCCCAAATCTTATTCTCAACGGTCCCTCATCCTTGCACCCCTCTTTGGA
CAAATGGCAGTTAGCATTGGTGCACTGACTGACTGCCCAACCTTAAACC
CAAATTTCTTAGAAGGGGCCCATCTAGTTAGCGAGGGGTGAAAAATTCC
TCCATCGGAGATGTATTGACCGTAAGTTGCTGCTTAAAAAAAATCAGTT
CAGATAGCGAGACTTTTTTGATTTCGCAACGGGAGTGCCTGTTCCATTCG
ATTGCAATTCTCACCCCTTCTGCCCAGTCCTGCCAATTGCCCATGAATCT
GCT**<u>TCCG</u>**T**<u>TCG</u>**T**<u>CCGA</u>**CCCACCCCCCTTTCCAACTCCACAAATTGTCCA
ATCTCGTTTTCCATTTGGGAGAATCTGCATGTCGACTACATAAAGCGACC
GGTGTCCGAAAAGATCTGTGTAGTTTTCAACATTTTGTGCTCCCCCCGCT
GTTTGAAAACGGGGGTGAGCGCTCTCCGGGGTGCGAATTCGTGCCCAAT
TCCTTTCACCCTGCCTATTGTAGACGTCAACCCGCATCTGGTGCGAATAT
AGCGCACCCCCAATGATCACACCAACAATTGGTCCACCCCTCCCCAATC
TCTAATATTCACAATTCACCTCACTATAAATACCCCTGTCCTGCTCCCAA
ATTCTTTTTTCCTTCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTAC
GAAA

>P*ADH2-AddAdr*

TCCTTTTTACCACCCAAGTGCGAGTGAAACACCCCATGGCTGCTCTCCGA
TTGCCCCTCTACAGGCATAAGGGTGTGACTTTGTGGGCTTGAATTTTACA
CCCCCTCCAACTTTTCTCGCATCAATTGATCCTGTTACCAATATTGCATG
CCCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATCGCAG
GGTGAGACTGTAGAGACCCCACATAGTGACAATGATTATGTAAGAAGAG
GGGGGTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTGAACA
ATGCCCAGCAGTCCTCCCCACTCTTTGACAAATCAGTATCACCGATTAAC
ACCCCAAATCTTATTCTCAACGGTCCCTCATCCTTGCACCCCTCTTTGGA
CAAATGGCAGTTAGCATTGGTGCACTGACTGACTGCCCAACCTTAAACC
CAAATTTCTTAGAAGGGGCCCATCTAGTTAGCGAGGGGTGAAAAATTCC
TCCATCGGAGATGTATTGACCGTAAGTTGCTGCTTAAAAAAAATCAGTT
CAGATAGCGAGACTTTTTTGATTTCGCAACGGGAGTGCCTGT**C**CC**CAA**C
GATTGCA**T**T**GGGG**ACCCCTTCTGCCCAGTCCTGCCAATTGCCCATGAAT
CTGCTAATTTCGTTGATTCCCACCCCCCTTTCCAACTCCACAAATTGTCC
AATCTCGTTTTCCATTTGGGAGAATCTGCATGTCGACTACATAAAGCGAC
CGGTGTCCGAAAAGATCTGTGTAGTTTTCAACATTTTGTGCTCCCCCCGC
TGTTTGAAAACGGGGGTGAGCGCTCTCCGGGGTGCGAATTCGTGCCCAA
TTCCTTTCACCCTGCCTATTGTAGACGTCAACCCGCATCTGGTGCGAATA
TAGCGCACCCCCAATGATCACACCAACAATTGGTCCACCCCTCCCCAAT
CTCTAATATTCACAATTCACCTCACTATAAATACCCCTGTCCTGCTCCCA
AATTCTTTTTTCCTTCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTA
CGAAA

>P*ADH2-OptAdr1-1*

TCCTTTTTACCACCCAAGTGCGAGTGAAACACCCCATGGCTGCTCTCCGA
TTGCCCCTCTACAGGCATAAGGGTGTGACTTTGTGGGCTTGAATTTTACA
CCCCCTCCAACTTTTCTCGCATCAATTGATCCTGTTACCAATATTGCATG
CCCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATCGCAG
GGTGAGACTGTAGAGACCCCA**A**ATAGTGACAATGATTATGTAAGAA**TT**G
GGG**A**GTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTGAACA
ATGCCCAGCAGTCCTCCCCACTCTTTGACAAATCAGTATCACCGATTAAC
ACCCCAAATCTTATTCTCAACGGTCCCTCATCCTTGCACCCCTCTTTGGA
CAAATGGCAGTTAGCATTGGTGCACTGACTGACTGCCCAACCTTAAACC
CAAATTTCTTAGAAGGGGCCCATCTAGTTAGCGAGGGGTGAAAAATTCC
TCCATCGGAGATGTATTGACCGTAAGTTGCTGCTTAAAAAAAATCAGTT
CAGATAGCGAGACTTTTTTGATTTCGCAACGGGAGTGCCTGTTCCATTCG
ATTGCAATTCTCACCCCTTCTGCCCAGTCCTGCCAATTGCCCATGAATCT
GCTAATTTCGTTGATTCCCACCCCCCTTTCCAACTCCACAAATTGTCCAA
TCTCGTTTTCCATTTGGGAGAATCTGCATGTCGACTACATAAAGCGACCG
GTGTCCGAAAAGATCTGTGTAGTTTTCAACATTTTGTGCTCCCCCCGCTG
TTTGAAAACGGGGGTGAGCGCTCTCCGGGGTGCGAATTCGTGCCCAATT
CCTTTCACCCTGCCTATTGTAGACGTCAACCCGCATCTGGTGCGAATATA
GCGCACCCCCAATGATCACACCAACAATTGGTCCACCCCTCCCCAATCTC
TAATATTCACAATTCACCTCACTATAAATACCCCTGTCCTGCTCCCAAAT
TCTTTTTTCCTTCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTACGA
AA

>P*ADH2-OptAdr1-2*

TCCTTTTTACCACCCAAGTGCGAGTGAAACACCCCATGGCTGCTCTCCGA
TTGCCCCTCTACAGGCATAAGGGTGTGACTTTGTGGGCTTGAATTTTACA
CCCCCTCCAACTTTTCTCGCATCAATTGATCCTGTTACCAATATTGCATG
CCCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATCGCAG
GGTGAGACTGTAGAGACCCCACATAGTGACAATGATTATGTAAGAAGAG
GGGGGTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTGAACA
ATGCCCAGCAGTCCTCCCCACTCTTTGACAAATCAGTATCACCGATTAAC
ACCCCAAATCTTATTCTCAACGGTCCCTCATCCTTGCACCCCTCTTTGGA
CAAATGGCAGTTAGCATTGGTGCACTGACTGACTGCCCAACCTTAAACC
CAAATTTCTTAGAAGGGGCCCATCTAGTTAGCGAGGGGTGAAAAATTCC
TCCATCGGAGATGTATTGACCGTAAGTTGCTGCTTAAAAAAAATCAGTT
CAGATAGCGAGACTTTTTTGATTTCGCAACGGGAGTGCCTGTTCCATTCG
ATTGCAATTCTCACCCCTTCTGCCCAGTCCTGCCAATTGCCCATGAATCT
GCTAATTTCGTTGATTCCCACCCCCCTTTCCA**T**CTCC**A**AAAATTGTCCAA
TCTCGTTTTCCATTT**T**GGAGAATCTGCATGTCGACTACATAAAGCGACCG
GTGTCCGAAAAGATCTGTGTAGTTTTCAACATTTTGTGCTCCCCCCGCTG
TTTGAAAACGGGGGTGAGCGCTCTCCGGGGTGCGAATTCGTGCCCAATT
CCTTTCACCCTGCCTATTGTAGACGTCAACCCGCATCTGGTGCGAATATA
GCGCACCCCCAATGATCACACCAACAATTGGTCCACCCCTCCCCAATCTC
TAATATTCACAATTCACCTCACTATAAATACCCCTGTCCTGCTCCCAAAT
TCTTTTTTCCTTCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTACGA
AA

>P*ADH2-OptAdr1-3*

TCCTTTTTACCACCCAAGTGCGAGTGAAACACCCCATGGCTGCTCTCCGA
TTGCCCCTCTACAGGCATAAGGGTGTGACTTTGTGGGCTTGAATTTTACA
CCCCCTCCAACTTTTCTCGCATCAATTGATCCTGTTACCAATATTGCATG
CCCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATCGCAG
GGTGAGACTGTAGAGACCCCACATAGTGACAATGATTATGTAAGAAGAG
GGGGGTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTGAACA
ATGCCCAGCAGTCCTCCCCACTCTTTGACAAATCAGTATCACCGATTAAC
ACCCCAAATCTTATTCTCAACGGTCCCTCATCCTTGCACCCCTCTTTGGA
CAAATGGCAGTTAGCATTGGTGCACTGACTGACTGCCCAACCTTAAACC
CAAATTTCTTAGAAGGGGCCCATCTAGTTAGCGAGGGGTGAAAAATTCC
TCCATCGGAGATGTATTGACCGTAAGTTGCTGCTTAAAAAAAATCAGTT
CAGATAGCGAGACTTTTTTGATTTCGCAACGGGAGTGCCTGTTCCATTCG
ATTGCAATTCTCACCCCTTCTGCCCAGTCCTGCCAATTGCCCATGAATCT
GCTAATTTCGTTGATTCCCACCCCCCTTTCCAACTCCACAAATTGTCCAA
TCTCGTTTTCCATTTGGGAGAATCTGCATGTCGACTACATAAAGCGACCG
GTGTCCGAAAAGATCTGTGTAGTTTTCAACATTTTGTGCTC**T**CCCC**AA**TG
TTTGAAA**TT**GGGG**A**TGAGCGCTCTCCGGGGTGCGAATTCGTGCCCAATT
CCTTTCACCCTGCCTATTGTAGACGTCAACCCGCATCTGGTGCGAATATA
GCGCACCCCCAATGATCACACCAACAATTGGTCCACCCCTCCCCAATCTC
TAATATTCACAATTCACCTCACTATAAATACCCCTGTCCTGCTCCCAAAT
TCTTTTTTCCTTCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTACGA
AA

>P*ADH2-AddAca*

TCCTTTTTACCACCCAAGTGCGAGTGAAACACCCCATGGCTGCTCTCCGA
TTGCCCCTCTACAGGCATAAGGGTGTGACTTTGTGGGCTTGAATTTTACA
CCCCCTCCAACTTTTCTCGCATCAATTGATCCTGTTACCAATATTGCATG
CCCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATCGCAG
GGTGAGACTGTAGAGACCCCACATAGTGACAATGATTATGTAAGAAGAG
GGGGGTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTGAACA
ATGCCCAGCAGTCCTCCCCACTCTTTGACAAATCAGTATCACCGATTAAC
ACCCCAAATCTTATTCTCAACGGTCCCTCATCCTTGCACCCCTCTTTGGA
CAAATGGCAGTTAGCATTGGTGCACTGACTGACTGCCCAACCTTAAACC
CAAATTTCTTAGAAGGGGCCCATCTAGTTAGCGAGGGGTGAAAAATTCC
TCCATCGGAGATGTATTGACCGTAAGTTGCTGCTTAAAAAAAATCAGTT
CAGATAGCGAGACTTTTTTGATTTCGCAACGGGAGTGCCTGTTCCATTCG
ATTGCAATTCTCACCCCTTCTGCCCAGTCCTGCCAATTGCCCATGAATCT
GCTAATTTCGTTGATTCCCACCCCCCTTTCCAACTCCACAAATTGTCCAA
TCTCGTTTTCCATTTGGGAGAATCTGCATGTCGACTACATAAAGCGACCG
GTGTCCGAAA**GCC**T**A**T**TGTAGACG**T**CA**A**C**C**TTTTGTGCTCCCCCCGCT
GTTTGAAAACGGGGGTGAGCGCTCTCCGGGGTGCGAATTCGTGCCCAAT
TCCTTTCACCCTGCCTATTGTAGACGTCAACCCGCATCTGGTGCGAATAT
AGCGCACCCCCAATGATCACACCAACAATTGGTCCACCCCTCCCCAATC
TCTAATATTCACAATTCACCTCACTATAAATACCCCTGTCCTGCTCCCAA
ATTCTTTTTTCCTTCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTAC
GAAA

>P*ADH2-AddTATA*

TCCTTTTTACCACCCAAGTGCGAGTGAAACACCCCATGGCTGCTCTCCGA
TTGCCCCTCTACAGGCATAAGGGTGTGACTTTGTGGGCTTGAATTTTACA
CCCCCTCCAACTTTTCTCGCATCAATTGATCCTGTTACCAATATTGCATG
CCCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATCGCAG
GGTGAGACTGTAGAGACCCCACATAGTGACAATGATTATGTAAGAAGAG
GGGGGTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTGAACA
ATGCCCAGCAGTCCTCCCCACTCTTTGACAAATCAGTATCACCGATTAAC
ACCCCAAATCTTATTCTCAACGGTCCCTCATCCTTGCACCCCTCTTTGGA
CAAATGGCAGTTAGCATTGGTGCACTGACTGACTGCCCAACCTTAAACC
CAAATTTCTTAGAAGGGGCCCATCTAGTTAGCGAGGGGTGAAAAATTCC
TCCATCGGAGATGTATTGACCGTAAGTTGCTGCTTAAAAAAAATCAGTT
CAGATAGCGAGACTTTTTTGATTTCGCAACGGGAGTGCCTGTTCCATTCG
ATTGCAATTCTCACCCCTTCTGCCCAGTCCTGCCAATTGCCCATGAATCT
GCTAATTTCGTTGATTCCCACCCCCCTTTCCAACTCCACAAATTGTCCAA
TCTCGTTTTCCATTTGGGAGAATCTGCATGTCGACTACATAAAGCGACCG
GTGTCCGAAAAGATCTGTGTAGTTTTCAACATTTTGTGCTCCCCCCGCTG
TTTGAAAACGGGGGTGAGCGCTCTCCGGGGTGCGAATTCGTGCCCAATT
CCTTTCACCCTGCCTATTGTAGACGTCAACCCGCA**<u>A</u>**C**<u>A</u>**G**<u>CAATAT</u>**ATA**<u>A</u>**
A**<u>CA</u>**G**<u>A</u>**ACCCCCAATGATCACACCAACAATTGGTCCACCCCTCCCCAATC
TCTAATATTCACAATTCACCTCACTATAAATACCCCTGTCCTGCTCCCAA
ATTCTTTTTTCCTTCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTAC
GAAA

>P*ADH2-NucOpt-OptCat*

TCCTTTTTACCACC**T**AAGTGCGAGTGAAACACCC**T**ATGGCTGCTCTCCGA
TTGCCCCTCTACAGGCATAAGGGTGTGA**T**TTT**T**T**TTTT**TT**T**AATTTTACA
CCCCCTCCAACTTTT**T**TCGC**G**T**A**AATTGATCCTGTTACCAATATTGCATG
CCCGGAGGAGACTTGCCCCCTAATTTCGCGGCGTCGTCCCGGATCGCAG
GGT**A**A**A**A**A**T**A**TA**T**AGACCCCACA**A**A**AAA**A**A**AATGATTATGTAAGAAGA
GGGGGGTGATTCGGCCGGCTATCGAACTCTAACAACTAGGGGGGTGAA**A**
AATGCCCAGC**TT**T**TT**TCCC**T**A**T**TCTTTGACAAATCAGTATCAC**TT**ATTAA
CACCCCAAAT**T**TT**T**TTCTCAACGGTCCCTCATCCTTGCACCCCTCTTTGG
ACAAATGGCAGTTAG**T**ATT**A**GTGCACTGACTGACTGCC**T**AACCTTAAAC
CC**T**AATTTCTTAGAAGGGGCCCAT**A**TAGTTAGCGAGGGGTGAAAAATTC
CTCCATCGGAGATGTATT**A**ACCGTAA**T**TT**TT**T**TT**TTAAAAAAAA**AA**A**AT
TCAGATAGCGA**A**A**T**TTTTTTGATTTCGC**G**ACG**C**G**C**GT**TTT**T**T**TT**TTT**TT**T**
TTTT**TTTT**TTCTCACCCCTTCTGCCCAGT**T**CTGCCAATTGCCCATGAATCT
ACTAATTTCGTTGATTCCCACCCCCCTTTCCAACTCCA**A**AAATT**T**T**TT**AA
T**T**T**TT**TTTTT**TTT**TTTGGGAGAATCTG**A**ATGT**AT**A**T**TACATAAAG**TTC**C**GT**
TCGTCCGAAAA**A**AT**T**T**T**T**T**T**TT**TTTTT**T**AA**TT**TTTT**T**T**TT**T**T**CCCCCGCT**T**T
TT**A**AAAACGGGGGT**A**AGCGCTCTCCGGGGTGCGAATTCG**C**GCCC**T**ATTC
CTTTCACCCTGCCTATTGTAGACGTCAACCCGCATCTGGTGCGAATATAG
CGCACCCCCAATGATCACACCAACAATTGGTCCACCCCTCCCCAATCTCT
AATATTCACAATTCACCTCACTATAAATACCCCTGTCCTGCTCCCAAATT
CTTTTTTCCTTCTTCCATCAGCTACTAGCTTTTATCTTATTTACTTTACGA
AA

APPENDIX B

The Sequence of $P_{AOX\ 1}$ variants

Modified positions are presented with bold and underlined characters.

> $P_{AOX1\text{-}wt}$

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCC
GACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAGGAGG
GGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTT
CTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGCT
TGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCAT
GACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTGT
TTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCAG
ATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGG
CCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAAG
CGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACG
GCCAGTTGGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTCTT
GTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAGC
GCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGCAA
ATGGGGAAACACCCGCTTTTTGGATGATTATGCATTGTCTCCACATTGTA
TGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCATGAT
CAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATAAACAGAA
GGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATTAGCTTACT
TTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACTT
TTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAACG

> $P_{AOX1\text{-}Adr1}$

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCC
GACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAGGAGG
GGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCAC**ACCCCA**
ATATTATTTGGGGTACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGG
CTTGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACC
ATGACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTT
GTTTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCA
GATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATG
GCCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAA
GCGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAAC
GGCCAGTTGGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTCT
TGTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAG
CGCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGCA
AATGGGGAAACACCCGCTTTTTGGATGATTATGCATTGTCTCCACATTGT
ATGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCATGA
TCAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATAAACAGA
AGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATTAGCTTAC
TTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACT
TTTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAACG

> P*AOX1-Adr2*

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCC
GACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAGGAGG
GGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTT
CTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGCT
TGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCAT
GACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTGT
TTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCAG
ATGAGGGC**GACCCCACATTTTTTTTTTGACCCCA**CATGTTCCCCAAAT
GGCCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAA
AGCGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAA
CGGCCAGTTGGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTC
TTGTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTA
GCGCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGC
AAATGGGGAAACACCCGCTTTTTGGATGATTATGCATTGTCTCCACATTG
TATGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCATG
ATCAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATAAACAG
AAGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATTAGCTTA
CTTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGAC
TTTTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAACG

> P*AOX1-Adr3*

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCC
GACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAGGAGG
GGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTT
CTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGCT
TGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCAT
GACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTGT
TTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCAG
ATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGG
CCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAAG
CGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACG
GCCAGTTGGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTCTT
GTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAGC
GCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGCAA
ATGGGGAAACACCCGCTTTTTGGATGATTA**ACCCCAATACATTTTGGG**
GTTGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCATG
ATCAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATAAACAG
AAGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATTAGCTTA
CTTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGAC
TTTTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAACG

> P*AOX1-Cat1*

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCC
GACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACA**ATTCC**
GTTCGTCCGATTAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTT
CTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGCT
TGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCAT
GACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTGT
TTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCAG
ATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGG
CCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAAG
CGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACG
GCCAGTTGGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTCTT
GTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAGC
GCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGCAA
ATGGGGAAACACCCGCTTTTTGGATGATTATGCATTGTCTCCACATTGTA
TGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCATGAT
CAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATAAACAGAA
GGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATTAGCTTACT
TTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACTT
TTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAACG

> $P_{AOX1\text{-}Cat2}$

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCC
GACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAGGAGG
GGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTT
CTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGCT
TGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCAT
GACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTGT
TTATTTCCGAATGC**<u>CCTCTCGTCCGGGCTTTTT</u>**CCGAACATCACTCCAG
ATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGG
CCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAAG
CGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACG
GCCAGTTGGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTCTT
GTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAGC
GCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGCAA
ATGGGGAAACACCCGCTTTTTGGATGATTATGCATTGTCTCCACATTGTA
TGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCATGAT
CAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATAAACAGAA
GGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATTAGCTTACT
TTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACTT
TTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAACG

> $P_{AOX1\text{-}Cat3}$

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCC
GACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAGGAGG
GGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTT
CTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGCT
TGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCAT
GACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTGT
TTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCAG
ATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGG
CCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAAG
CGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACG
GCCAGTTGGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTCTT
GTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAGC
GCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAAC**ATAT**
TCCGTTCGTCCGAATCTTTTTGGATGATTATGCATTGTCTCCACATTGTA
TGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCATGAT
CAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATAAACAGAA
GGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATTAGCTTACT
TTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACTT
TTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAACG

> $P_{AOX1\text{-}Aca}$

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCC
GACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAGGAGG
GGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTT
CTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGCT
TGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCAT
GACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTGT
TTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCAG
ATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGG
CCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAAG
CGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACG
GCCAGT**<u>GCCTATTGTAGACGTCAACCC</u>**AAGTCGGCATACCGTTTGTCT
TGTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAG
CGCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGCA
AATGGGGAAACACCCGCTTTTTGGATGATTATGCATTGTCTCCACATTGT
ATGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCATGA
TCAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATAAACAGA
AGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATTAGCTTAC
TTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACT
TTTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAACG

> P*AOX1-Cat3Adr3*

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCC
GACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAGGAGG
GGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTT
CTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGCT
TGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCAT
GACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTGT
TTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCAG
ATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGG
CCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAAG
CGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACG
GCCAGTTGGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTCTT
GTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAGC
GCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAAC**ATAT**
TCCGTTCGTCCGAATCTTTTTGGATGATTA**ACCCCAATACATTTTGGG**
GTTGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCATG
ATCAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATAAACAG
AAGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATTAGCTTA
CTTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGAC
TTTTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAACG

> P*AOX1-AcaCat3Adr3*

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCC
GACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAGGAGG
GGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTT
CTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGCT
TGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCAT
GACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTGT
TTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCAG
ATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGG
CCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAAG
CGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACG
GCCAGT**GCCTATTGTAGACGTCAACCC**AAGTCGGCATACCGTTTGTCT
TGTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAG
CGCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAAC**ATA
TTCCGTTCGTCCGAAT**CTTTTTGGATGATTA**ACCCCAATACATTTTGG
GGT**TGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCA
TGATCAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATAAACA
GAAGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATTAGCTT
ACTTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGA
CTTTTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAAC
G

> P*AOX1-mod*

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCC
GACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACA**ATTCC**
GTTCGTCCGATTAGCAGACCGTTGCAAACGCAGGACCTCCAC**ACCCCA**
ATATTATTTGGGGTACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGG
CTTGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACC
ATGACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTT
GTTTATTTCCGAATGC**CCTCTCGTCCGGGCTTTTT**CCGAACATCACTCC
AGATGAGGGC**GACCCCACATTTTTTTTTTGACCCCA**CATGTTCCCCAA
ATGGCCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACA
AAAGCGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCT
AACGGCCAGT**GCCTATTGTAGACGTCAACCC**AAGTCGGCATACCGTTT
GTCTTGTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGC
TTAGCGCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAA
C**ATATTCCGTTCGTCCGAAT**CTTTTTGGATGATTA**ACCCCAATACATT**
TTGGGGTTGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACG
TTCATGATCAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATA
AACAGAAGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATT
AGCTTACTTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTT
AACGACTTTTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTC
GAAACG

APPENDIX C

Gene sequence of pPICZα-A*::egfp* (4286 bp)

‘’AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATC
CGACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAGGAG
GGGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCT
TCTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGC
TTGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCA
TGACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTG
TTTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCAG
ATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGG
CCCAAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAAG
CGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACG
GCCAGTTGGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTCTT
GTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAGC
GCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGCAA
ATGGGGAAACACCCGCTTTTTGGATGATTATGCATTGTCTCCACATTGTA
TGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCATGAT
CAAAATTTAACTGTTCTAACCCCTACTTGACAGCAATATATAAACAGAA
GGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATTAGCTTACT
TTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACTT
TTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAACGAT
GAGATTTCCTTCAATTTTTACTGCTGTTTTATTCGCAGCATCCTCCGCATT
AGCTGCTCCAGTCAACACTACAACAGAAGATGAAACGGCACAAATTCCG
GCTGAAGCTGTCATCGGTTACTCAGATTTAGAAGGGGATTTCGATGTTGC
TGTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTTTATAAATA
CTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGAA
AAGAGAGGCTGAAGCTGAATTCATGGTGAGCAAGGGCGAGGAGCTGTT
CACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCA
AGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTG
GCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCT
ACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGA

AGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTAC
AAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
ATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGG
CACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACA
TCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCC
CATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACC
CAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCC
TGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCT
GTACAAGTAAGGTACCTCGAGCCGCGGCGGCCGCCAGCTTTCTAGAACA
AAAACTCATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCAT
CATCATCATTGAGTTTGTAGCCTTAGACATGACTGTTCCTCAGTTCAAGT
TGGGCACTTACGAGAAGACCGGTCTTGCTAGATTCTAATCAAGAGGATG
TCAGAATGCCATTTGCCTGAGAGATGCAGGCTTCATTTTTGATACTTTTT
TATTTGTAACCTATATAGTATAGGATTTTTTTTGTCATTTTGTTTCTTCTC
GTACGAGCTTGCTCCTGATCAGCCTATCTCGCAGCTGATGAATATCTTGT
GGTAGGGGTTTGGGAAAATCATTCGAGTTTGATGTTTTTCTTGGTATTTC
CCACTCCTCTTCAGAGTACAGAAGATTAAGTGAGACCTTCGTTTGTGCGG
ATCCCCCACACACCATAGCTTCAAAATGTTTCTACTCCTTTTTTACTCTTC
CAGATTTTCTCGGACTCCGCGCATCGCCGTACCACTTCAAAACACCCAA
GCACAGCATACTAAATTTTCCCTCTTTCTTCCTCTAGGGTGTCGTTAATTA
CCCGTACTAAAGGTTTGGAAAAGAAAAAAGAGACCGCCTCGTTTCTTTT
TCTTCGTCGAAAAAGGCAATAAAAATTTTTATCACGTTTCTTTTTCTTGA
AATTTTTTTTTTTAGTTTTTTTCTCTTTCAGTGACCTCCATTGATATTTAAG
TTAATAAACGGTCTTCAATTTCTCAAGTTTCAGTTTCATTTTTCTTGTTCT
ATTACAACTTTTTTTACTTCTTGTTCATTAGAAAGAAAGCATAGCAATCT
AATCTAAGGGGCGGTGTTGACAATTAATCATCGGCATAGTATATCGGCA
TAGTATAATACGACAAGGTGAGGAACTAAACCATGGCCAAGTTGACCAG
TGCCGTTCCGGTGCTCACCGCGCGCGACGTCGCCGGAGCGGTCGAGTTC
TGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCG
CCGGTGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGA
CCAGGTGGTGCCGGACAACACCCTGGCCTGGGTGTGGGTGCGCGGCCTG
GACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCCGGG
ACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGC
GGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGC
CGAGGAGCAGGACTGACACGTCCGACGGCGGCCCACGGGTCCCAGGCC
TCGGAGATCCGTCCCCCTTTTCCTTTGTCGATATCATGTAATTAGTTATGT
CACGCTTACATTCACGCCCTCCCCCCACATCCGCTCTAACCGAAAAGGA
AGGAGTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTTATAGTTA
TGTTAGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGT
ACAGACGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAA
GGTTTTGGGACGCTCGAAGGCTTTAATTTGCAAGCTGGAGACCAACATG
TGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTG

CTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCG
ACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCA
GGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGC
CGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTT
TCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTC
CAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCC
TTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATC
GCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTA
GGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
GAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGG
AAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC
GGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGAT
CTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAC
GAAAACTCACGTTAAGGGATTTTGGTCATGAGATC’’

*eGFP gene start and end positions were underlined.

APPENDIX D

Putative TFBSs analysis of P_{ADH2} by MatInspector

List of predicted putative TFBSs by MatInspector. Matching score is calculated based on the similarity of the binding site sequence found on P_{ADH2} and the matrix and gives a score of 1.0 in case of 100% similarity. Start and end positions refer to the location of the respective binding site on P_{ADH2} sequence.

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
F$PRES	Pheromone response elements	F$STE12.01	Transcription factor activated by a MAP kinase signaling cascade, activates genes involved in mating or pseudohyphal/invasive growth pathways	0.86	22	34	gagtgaAACAccc
F$IRTF	Iron-responsive transcriptional activators	F$AFT2.01	Activator of Fe (iron) transcription 2, iron-regulated transcriptional activator	0.98	23	37	agtgaaaCACCccat
F$CYTO	Activator of cytochrome C	F$HAP1.01	HAP1, S. cerevisiae member of GAL family, regulates heme	0.76	45	59	gaggggcaATCGgag

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
			dependent cytochrome expression				
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	1	50	64	ctgtagaGGGGcaat
F$IRTF	Iron-responsive transcriptional activators	F$AFT2.01	Activator of Fe (iron) transcription 2, iron-regulated transcriptional activator	0.98	68	82	aaagtcaCACCctta
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type l	F$RAP1.06	RAP1 (TUF1), activator or repressor depending on context	0.94	65	87	cccacaaagtcaCACCctatgc
F$YMSE	Yeast middle sporulation elements	F$NDT80.01	Meiosis-specific transcription factor, binds to promoters containing middle sporulation elements (MSE)	0.85	77	91	caagcccaCAAAgtc
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type l	F$RAP1.02	RAP1 (TUF1), activator or repressor depending on context	0.82	83	105	gtgggcttgaattttaCACCccc
F$RRPE	Ribosomal RNA processing element	F$STB3.01	Ribosomal RNA processing element (RRPE)-binding protein	0.82	88	104	cttgaattTTACacccc

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
F$IRTF	Iron-responsive transcriptional activators	F$AFT2.01	Activator of Fe (iron) transcription 2, iron-regulated transcriptional activator	0.98	92	106	aattttaCACCccct
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	1	98	112	agttggaGGGGgtgt
F$YABF	Yeast ABF factors	F$ABF1.04	ARS (autonomously replicating sequence)-binding factor I	0.77	114	132	ggATCAattgatgcgagaa
F$RFXP	Regulatory factor X protein, homologous to mammalian RFX1-5	F$RFX1.01	RFX1 (CRT1) is a DNA-binding protein that acts by recruiting Ssn6 and Tup1, general repressors to the promoters of damage-inducible genes	0.92	132	146	caatattgGTAAcag
F$BZIP	Fungal basic leucine zipper family	F$YAP1.02	Yeast activator protein of the basic leucine zipper (bZIP) family	0.93	131	151	gcatgcaatattgGTAAcagg
F$FBAS	Fungi branched amino acid biosynthesis	F$LEU3.02	LEU3, S. cerevisiae, zinc cluster protein	0.82	149	165	tgcCCGGaggagacttg
F$YGAL	Yeast GAL4 factor	F$GAL4.01	GAL4 transcriptional activator in response to galactose induction	0.75	148	172	taggggcaagtctcctCCGGgcat
F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$ICGG_N10.01	Inverted CGG motifs separated by a spacer of 10 bp; Put3, Tea1 and Cha4 binding sites	0.76	152	172	cCGGAggagacttgcccccta

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	1	163	177	gaaattaGGGGgcaa
F$PDRE	Pleiotropic drug resistance responsive elements	F$PDR1.01	Zinc cluster protein, master regulator involved in recruiting other zinc cluster proteins to pleiotropic drug response elements (PDREs)	0.88	174	182	gCCGCgaaa
F$YMCB	Yeast Mlu I cell cycle box	F$SWI4.01	DNA binding component of the SBF(SCB binding factor) complex (Swi4p-Swi6p)	0.84	172	184	acgccgCGAAatt
F$ASG1	Activator of stress genes	F$ASG1.01	Fungal zinc cluster transcription factor Asg1	0.79	177	193	cGCGGcgtcgtcccgga
F$CSRE	Carbon source-responsive elements	F$CAT8.01	Zinc cluster transcriptional activator binding to carbon source responsive elements (CSRE)	0.86	187	201	ccctgcgatCCGGga
O$MTEN	Core promoter motif ten elements	O$DMTE.01	Drosophila motif ten element	0.77	186	206	gtcccggATCGcagggtgaga
F$YADR	Yeast metabolic regulator	F$ADR1.01	Alcohol Dehydrogenase Regulator, carbon source-responsive zinc-finger transcription factor	0.94	214	222	gaCCCCaca

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
F$YMIG	Yeast GC-Box Proteins	F$MIG3.01	Zinc finger transcriptional repressor MIG3	0.81	212	230	tgtcactatgtGGGGtctc
F$BZIP	Fungal basic leucine zipper family	F$CIN5.01	bZIP transcriptional factor of the yAP-1 family that mediates pleiotropic drug resistance and salt tolerance	0.89	226	246	tgacaatgattatgTAAGaag
F$BZIP	Fungal basic leucine zipper family	F$CIN5.01	bZIP transcriptional factor of the yAP-1 family that mediates pleiotropic drug resistance and salt tolerance	0.89	231	251	cccctcttcttacaTAATcat
O$VTBP	Vertebrate TATA binding protein factor	O$ATATA.01	Avian C-type LTR TATA box	0.78	233	249	gattatgTAAGaagagg
F$YMIG	Yeast GC-Box Proteins	F$MIG1.02	MIG1, zinc finger protein mediates glucose repression	0.83	237	255	atgtaagaagaGGGGggtg
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	1	241	255	aagaagaGGGGgtg
F$CSRE	Carbon source-responsive elements	F$CSRE.01	Carbon source-responsive element (yeast)	0.79	253	267	gtgattcgGCCGgct
F$RDS1	Regulator of Drug Sensitivity 1	F$RDS1.01	Regulator of Drug Sensitivity 1, zinc cluster transcription factor involved in	0.82	255	267	agcCGGCcgaatc

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
			conferring resistance to cycloheximide				
F$RDS1	Regulator of Drug Sensitivity 1	F$RDS1.01	Regulator of Drug Sensitivity 1, zinc cluster transcription factor involved in conferring resistance to cycloheximide	0.82	256	268	attCGGCcggcta
F$SKN7	Skn7 response regulator of Saccharomyces cerevisiae	F$SKN7.01	SKN7, a transcription factor contributing to the oxidative stress response	0.92	259	269	CGGCcggctat
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	1	279	293	acaactaGGGGggtg
F$ROX1	Repressor of hypoxic genes	F$ROX1.01	Heme-dependent transcriptional repressor of hypoxic genes	0.82	290	302	ggcaTTGTtcacc
F$YMAT	Yeast mating factors	F$MATALPHA2.02	Homeodomain transcriptional repressor Matalpha2	0.86	291	303	gggcaTTGTtcac
F$YMIG	Yeast GC-Box Proteins	F$MIG1.01	MIG1, zinc finger protein mediates glucose repression	0.84	309	327	tgtcaaagagtGGGGagga
F$TALE	Fungal TALE homeodomain class	F$TOS8.01	Homeodomain-containing transcription factor	0.96	321	333	ctgatttGTCAaa
F$BZIP	Fungal basic leucine	F$YAP1.02	Yeast activator protein of the basic leucine zipper (bZIP) family	0.93	319	339	gtgatactgatttGTCAaaga

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
	zipper family						
F$YORE	Yeast oleate response elements	F$OAF1.01	Oleate-activated transcription factor, acts alone and as a heterodimer with Pip2p	0.83	328	352	aatcagtatcACCGattaacacccc
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type l	F$RAP1.06	RAP1 (TUF1), activator or repressor depending on context	0.94	335	357	atcaccgattaaCACCccaaatc
F$IRTF	Iron-responsive transcriptional activators	F$AFT2.01	Activator of Fe (iron) transcription 2, iron-regulated transcriptional activator	0.98	340	354	cgattaaCACCccaa
F$YADR	Yeast metabolic regulator	F$ADR1.01	Alcohol Dehydrogenase Regulator, carbon source-responsive zinc-finger transcription factor	0.94	347	355	caCCCCaaa
F$YMIG	Yeast GC-Box Proteins	F$MIG1.01	MIG1, zinc finger protein mediates glucose repression	0.84	345	363	gaataagatttGGGGtgtt
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type l	F$RAP1.05	RAP1 (TUF1), activator or repressor depending on context	0.85	368	390	cggtccctcatcctTGCAcccct

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
F$IRTF	Iron-responsive transcriptional activators	F$AFT1.01	Transcription factor involved in iron utilization and homeostasis (iron-sensing transcription factor AFT1)	0.87	377	391	atccttgCACCcctc
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	1	382	396	ccaaagaGGGGtgca
F$YMIG	Yeast GC-Box Proteins	F$MIG1.02	MIG1, zinc finger protein mediates glucose repression	0.83	382	400	ttgtccaaagaGGGGtgca
F$CSRE	Carbon source-responsive elements	F$SIP4.01	Zinc cluster transcriptional activator, binds to the carbon source-responsive element (CSRE) of gluconeogenic genes	0.76	391	405	gCCATttgtccaaag
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type l	F$RAP1.02	RAP1 (TUF1), activator or repressor depending on context	0.82	413	435	gggcagtcagtcagtgCACCaat
F$TGRF	Telobox-containing general regulatory factor	F$TBF1.01	TTAGGG repeat binding factor 1	0.82	434	446	ccaacCTTAaacc
F$YRAP	Yeast activator of glycolyse	F$RAP1.04	RAP1 telomeric binding sites	0.78	435	457	caaccttaaACCCaaatttctta

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
	genes / repressor of mating type l						
F$TGRF	Telobox-containing general regulatory factor	F$TBF1.01	TTAGGG repeat binding factor 1	0.82	441	453	taaacCCAAattt
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	1	454	468	cttagaaGGGGccca
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	1	474	488	ttagcgaGGGGtgaa
F$RRPE	Ribosomal RNA processing element	F$STB3.01	Ribosomal RNA processing element (RRPE)-binding protein	0.82	481	497	gaggaattTTTCacccc
F$MGCM	Monomeric Gal4-class motifs	F$LYS14.01	Transcriptional activator involved in regulation of genes of the lysine biosynthesis pathway	0.8	485	501	gatggaGGAAtttttca
F$YMCM	Yeast cell cycle and metabolic regulator	F$MCM1.02	Yeast factor MCM1 cooperating with MATalpha factors	0.75	491	507	aTTCCtccatcggagat
F$YORE	Yeast oleate response elements	F$OAF1.01	Oleate-activated transcription factor, acts alone and as a heterodimer with Pip2p	0.83	490	514	tcaatacatcTCCGatggaggaatt

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
F$YGAL	Yeast GAL4 factor	F$GAL4.03	DNA-binding zinc cluster transcription factor required for activation of GAL genes	0.84	496	520	ttACGGtcaatacatctccgatgga
F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$ICGG_N10.01	Inverted CGG motifs separated by a spacer of 10 bp; Put3, Tea1 and Cha4 binding sites	0.76	500	520	tCGGAgatgtattgaccgtaa
F$YGAL	Yeast GAL4 factor	F$GAL4.03	DNA-binding zinc cluster transcription factor required for activation of GAL genes	0.84	498	522	caTCGGagatgtattgaccgtaagt
F$CYTO	Activator of cytochrome C	F$HAP1.01	HAP1, S. cerevisiae member of GAL family, regulates heme dependent cytochrome expression	0.76	513	527	cagcaactTACGgtc
F$RFXP	Regulatory factor X protein, homologous to mammalian RFX1-5	F$RFX1.02	RFX1 (CRT1), acts by recruiting Ssn6 and Tup1, general repressors to the promoters of damage-inducible genes	0.77	519	533	tttaagcagCAACtt
O$VTBP	Vertebrate TATA binding protein factor	O$ATATA.01	Avian C-type LTR TATA box	0.78	522	538	ttttttTAAGcagcaa
F$RFXP	Regulatory factor X	F$RFX1.01	RFX1 (CRT1) is a DNA-binding protein	0.92	562	576	ttgatttcGCAAcgg

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
	protein, homologous to mammalian RFX1-5		that acts by recruiting Ssn6 and Tup1, general repressors to the promoters of damage-inducible genes				
F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$ICGG_N10.01	Inverted CGG motifs separated by a spacer of 10 bp; Put3, Tea1 and Cha4 binding sites	0.76	571	591	aTGGAacaggcactcccgttg
F$YHSF	Yeast heat shock factors	F$HSF1.01	Trimeric heat shock transcription factor	0.95	573	605	agaattgcaatcgaatGGAAcaggcactcccgt
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	1	604	618	ggcagaaGGGGtgag
F$YHSF	Yeast heat shock factors	F$HSF1.02	Trimeric heat shock transcription factor	0.79	599	631	attggcaggactgggcagaaggggtgAGAAttg
F$YSTR	Yeast stress response elements	F$USV1.01	Usv (Up in StarVation), Nsf1 (nutrient and stress factor 1)	0.83	630	644	agattcATGGgcaat
F$YMIG	Yeast GC-Box Proteins	F$MIG1.02	MIG1, zinc finger protein mediates glucose repression	0.83	661	679	agttggaaaggGGGGtgg
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	1	663	677	ttggaaaGGGGggtg
F$YHSF	Yeast heat shock factors	F$HSF1.01	Trimeric heat shock transcription factor	0.95	689	721	gcagattctccccaaatGGA

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
							Aaacgagattggac
F$YMCM	Yeast cell cycle and metabolic regulator	F$MCM1.01	Yeast factor MCM1 cooperating with MATalpha factors	0.83	699	715	tctCCCAaatggaaaac
O$VTBP	Vertebrate TATA binding protein factor	O$MTATA.01	Muscle TATA box	0.84	729	745	ctacaTAAAgcgaccgg
F$FBAS	Fungi branched amino acid biosynthesis	F$LEU3.02	LEU3, S. cerevisiae, zinc cluster protein	0.82	739	755	cgaCCGGtgtccgaaaa
F$MGCM	Monomeric Gal4-class motifs	F$CEP3.01	Essential kinetochore protein, component of the CBF3 complex that binds the CDEIII region of the centromere	0.86	744	760	ggtgTCCGaaaagatct
F$GATC	GATA-class proteins binding to GATC motifs	F$GAT4.01	GATA-type Zn finger protein Gat4	0.8	749	763	cacaGATCttttcgg
F$GATC	GATA-class proteins binding to GATC motifs	F$GAT4.01	GATA-type Zn finger protein Gat4	0.8	752	766	aaaaGATCtgtgtag
F$YMSE	Yeast middle	F$MSE.01	Middle sporulation element (MSE)	0.92	774	788	ggggagcaCAAAatg

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
	sporulation elements						
F$YMIG	Yeast GC-Box Proteins	F$MIG1.02	MIG1, zinc finger protein mediates glucose repression	0.83	783	801	tttcaaacagcGGGGggag
F$YMIG	Yeast GC-Box Proteins	F$MIG1.01	MIG1, zinc finger protein mediates glucose repression	0.84	794	812	gtttgaaaacgGGGGtgag
F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$ICGG_N10.01	Inverted CGG motifs separated by a spacer of 10 bp; Put3, Tea1 and Cha4 binding sites	0.76	802	822	cCGGAgagcgctcacccccgt
F$YMCB	Yeast Mlu I cell cycle box	F$SWI4.01	DNA binding component of the SBF(SCB binding factor) complex (Swi4p-Swi6p)	0.84	828	840	tgggcaCGAAttc
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type l	F$RAP1.02	RAP1 (TUF1), activator or repressor depending on context	0.82	833	855	cgtgcccaattccttCACCctg
F$CSRE	Carbon source-responsive elements	F$SIP4.01	Zinc cluster transcriptional activator, binds to the carbon source-responsive element (CSRE) of gluconeogenic genes	0.76	843	857	tCCTTtcaccctgcc
F$MMAT	M-box interacting	F$MAT1MC.01	HMG-BOX protein interacts with M-box	0.87	856	866	cctATTGtaga

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
	with Mat1-Mc		site, cooperativity with HMG-Box STE11 protein				
F$YMAT	Yeast mating factors	F$MATALPHA2.02	Homeodomain transcriptional repressor Matalpha2	0.86	855	867	gcctaTTGTagac
F$BZIP	Fungal basic leucine zipper family	F$CST6.01	Chromosome stability, bZIP transcription factor of the ATF/CREB family (ACA2)	0.85	855	875	gcctattgtagACGTcaacc
F$IRTF	Iron-responsive transcriptional activators	F$AFT1.01	Transcription factor involved in iron utilization and homeostasis (iron-sensing transcription factor AFT1)	0.87	890	904	tatagcgCACCccca
F$YCAT	Yeast CCAAT binding factors	F$HAP234.01	Yeast factor complex HAP2/3/5, homolog to vertebrate NF-Y/CP1/CBF	0.86	897	909	cacccCCAAtgat
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type l	F$RAP1.06	RAP1 (TUF1), activator or repressor depending on context	0.94	900	922	ccccaatgatcaCACCaacaatt
F$YCAT	Yeast CCAAT binding factors	F$HAP234.01	Yeast factor complex HAP2/3/5, homolog to vertebrate NF-Y/CP1/CBF	0.86	909	921	tcacaCCAAcaat

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
F$ROX1	Repressor of hypoxic genes	F$ROX1.01	Heme-dependent transcriptional repressor of hypoxic genes	0.82	912	924	ccaaTTGTtggtg
F$MMAT	M-box interacting with Mat1-Mc	F$MAT1MC.01	HMG-BOX protein interacts with M-box site, cooperativity with HMG-Box STE11 protein	0.87	914	924	ccaATTGttgg
F$YMAT	Yeast mating factors	F$MATALPHA2.02	Homeodomain transcriptional repressor Matalpha2	0.86	913	925	accaaTTGTtggt
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type l	F$RAP1.02	RAP1 (TUF1), activator or repressor depending on context	0.82	911	933	acaccaacaattggtcCACCcct
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	1	925	939	ttggggaGGGGtgga
O$PTBP	Plant TATA binding protein factor	O$PTATA.02	Plant TATA box	0.9	962	976	tcacTATAaataccc
O$VTBP	Vertebrate TATA binding protein factor	O$VTATA.01	Cellular and viral TATA box elements	0.9	963	979	cactaTAAAtacccctg
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	1	970	984	caggacaGGGGtatt

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Sequence
F$MGCM	Monomeric Gal4-class motifs	F$RGT1.02	Glucose-responsive transcription factor involved in regulation of glucose transporters	0.87	993	1009	gaagaAGGAaaaaagaa
F$YHSF	Yeast heat shock factors	F$HSF1.01	Trimeric heat shock transcription factor	0.95	994	1026	aaagctagtagctgatGGAAgaaggaaaaaaga
F$GATA	Fungal GATA binding factors	F$GATA.01	GATA binding factor (yeast)	0.89	1020	1034	ataaGATAaaagcta
O$VTBP	Vertebrate TATA binding protein factor	O$ATATA.01	Avian C-type LTR TATA box	0.78	1024	1040	aagtaaaTAAGataaaa
F$FKHD	Fungal fork head transcription factors	F$FKH2.01	Fork head transcription factor Fkh2	0.86	1028	1044	cgtaaagTAAAtaagat

APPENDIX E

Differences in Putative TFBSs Between P_{ADH2} and Its Variants

Table E1: Differences in Putative TFBSs Between P_{ADH2} and $P_{ADH2\text{-}OptAdr1\text{-}1}$

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$YMIG	Yeast GC-Box ProteIns	F$MIG3.01	Zinc finger transcriptional repressor MIG3	0.81	212	230	221	tgtcactatgtGGGGtctc	Del
F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$ICGG_N10.01	Inverted CGG motifs separated by a spacer of 10 bp; Put3, Tea1 and Cha4 binding sites	0.76	248	268	258	gGGGAgtgattcggccggcta	Ins
F$YMIG	Yeast GC-Box ProteIns	F$MIG1.02	MIG1, zinc finger protein mediates glucose repression	0.83	237	255	246	atgtaagaagaGGGGggtg	Del
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	1	241	255	248	aagaagaGGGGgtg	Del

Table E2: Differences in Putative TFBSs Between P_{ADH2} and $P_{ADH2\text{-}OptAdr1\text{-}2}$

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$YHSF	Yeast heat shock factors	F$HSF1.01	Trimeric heat shock transcription factor	0.95	659	691	675	gacaatttttggagatGGAAaggggggtgggaa	Ins

Table E3: Differences in Putative TFBSs Between P_{ADH2} and $P_{ADH2\text{-}OptAdr1\text{-}3}$

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$YQA1	Neurospora crassa QA1 gene activator	F$QA1F.01	qa-1F, required for quinic acid induction of transcription in the qa gene cluster	0.75	803	823	813	cccggagagcgcTCATcccca	Ins
F$YMIG	Yeast GC-Box ProteIns	F$MIG1.02	MIG1, zinc finger protein mediates glucose repression	0.83	783	801	792	tttcaaacagcGGGGggag	Del
F$YMIG	Yeast GC-Box ProteIns	F$MIG1.01	MIG1, zinc finger protein mediates glucose repression	0.84	794	812	803	gtttgaaaacgGGGGtgag	Del
F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$ICGG_N10.01	Inverted CGG motifs separated by a spacer of 10 bp; Put3, Tea1 and Cha4 binding sites	0.76	802	822	812	cCGGAgagcgctcacccccgt	Del

Table E4: Differences in Putative TFBSs Between P_{ADH2} and $P_{ADH2\text{-}AddAdr1}$

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$ICGG_N10.01	Inverted CGG motifs separated by a spacer of 10 bp; Put3, Tea1 and Cha4 binding sites	0.76	571	591	581	aTGGAacaggcactcccgttg	Del
F$YHSF	Yeast heat shock factors	F$HSF1.01	Trimeric heat shock transcription factor	0.95	573	605	589	agaattgcaatcgaatGGAAcaggcactccgt	Del
F$YHSF	Yeast heat shock factors	F$HSF1.02	Trimeric heat shock transcription factor	0.79	599	631	615	attggcaggactgggcagaagggtgAGAAttg	Del

Table E5: Differences in Putative TFBSs Between P_{ADH2} and $P_{ADH2\text{-}AddCat\text{-}1st(AddCat)}$

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
O$MTEN	Core promoter motif ten elements	O$DMTE.01	Drosophila motif ten element	0.77	640	660	650	tcggacgAACGgaagcagatt	Ins
F$YGAL	Yeast GAL4 factor	F$GAL4.01	GAL4 transcriptional activator in response to galactose induction	0.75	640	664	652	aatctgcttccgttcgtCCGAccca	Ins
F$CSRE	Carbon source-responsive elements	F$CAT8.01	Zinc cluster transcriptional activator binding to carbon source responsive elements (CSRE)	0.86	648	662	655	tccgttcgtCCGAcc	Ins
F$MGCM	Monomeric Gal4-class motifs	F$RGT1.02	Glucose-responsive transcription factor involved in regulation of glucose transporters	0.87	648	664	656	tgggtCGGAcgaacgga	Ins
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type l	F$RAP1.04	RAP1 telomeric binding sites	0.78	651	673	662	gttcgtccgACCCacccccctttt	Ins

Table E6: Differences in Putative TFBSs Between P_{ADH2} and $P_{ADH2\text{-}Add\text{-}Cat2nd(OptCat)}$

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
O$VTBP	Vertebrate TATA binding protein factor	O$MTATA.01	Muscle TATA box	0.84	729	745	737	ctacaTAAAgcgaccgg	Del
F$FBAS	Fungi branched amino acid biosynthesis	F$LEU3.02	LEU3, S. cerevisiae, zinc cluster protein	0.82	739	755	747	cgaCCGGtgtccgaaaa	Del
F$MGCM	Monomeric Gal4-class motifs	F$LYS14.01	Transcriptional activator involved in regulation of genes of the lysine biosynthesis pathway	0.8	732	748	740	acgaacGGAActttatg	Ins
O$MTEN	Core promoter motif ten elements	O$DMTE.01	Drosophila motif ten element	0.77	732	752	742	tcggacgAACGgaacttttatg	Ins
F$CSRE	Carbon source-responsive elements	F$CAT8.01	Zinc cluster transcriptional activator binding to carbon source responsive elements (CSRE)	0.86	740	754	747	tccgttcgtCCGAaa	Ins
F$MGCM	Monomeric Gal4-class motifs	F$RGT1.02	Glucose-responsive transcription factor involved in regulation of glucose transporters	0.87	740	756	748	cttttCGGAcgaacgga	Ins
F$YORE	Yeast oleate response elements	F$OAF1.01	Oleate-activated transcription factor, acts alone and as a heterodimer with Pip2p	0.83	738	762	750	gttccgttcgTCCGaaaagatctgt	Ins
F$YABF	Yeast ABF factors	F$ABF1.03	ARS (autonomously replicating sequence)-binding factor I	0.89	742	760	751	agatctttcggACGAacg	Ins

Table E7: Differences in Putative TFBSs Between P*ADH2* and P *ADH2-Add-TATA*

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$IRTF	Iron-responsive transcriptional activators	F$AFT1.01	Transcription factor involved in iron utilization and homeostasis (iron-sensing transcription factor AFT1)	0.87	890	904	897	tatagcgCACCccca	Del
O$YTBP	Yeast TATA binding protein factor	O$SPT15.01	TATA-binding protein, general transcription factor that interacts with other factors to form the preinitiation complex at promoters	0.83	879	895	887	acagcaaTATAtaaaca	Ins
O$PTBP	Plant TATA binding protein factor	O$PTATA.02	Plant TATA box	0.9	882	896	889	gcaaTATAtaaacag	Ins
O$YTBP	Yeast TATA binding protein factor	O$SPT15.01	TATA-binding protein, general transcription factor that interacts with other factors to form the preinitiation complex at promoters	0.83	882	898	890	ttctgttTATAtattgc	Ins
F$FKHD	Fungal fork head transcription factors	F$FKH1.01	Fork head transcription factor Fkh1	0.81	883	899	891	caatataTAAAcagaac	Ins
O$PTBP	Plant TATA binding protein factor	O$PTATA.01	Plant TATA box	0.88	884	898	891	aataTATAaacagaa	Ins
O$VTBP	Vertebrate TATA binding protein factor	O$VTATA.01	Cellular and viral TATA box elements	0.9	883	899	891	caataTATAaacagaac	Ins

Table E8: Differences in Putative TFBSs Between P_{ADH2} and $P_{ADH2\text{-}AddAca}$

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$GATC	GATA-class proteIns binding to GATC motifs	F$GAT4.01	GATA-type Zn finger protein Gat4	0.8	749	763	756	cacaGATCttttcgg	Del
F$GATC	GATA-class proteIns binding to GATC motifs	F$GAT4.01	GATA-type Zn finger protein Gat4	0.8	752	766	759	aaaaGATCtgtgtag	Del
F$MMAT	M-box interacting with Mat1-Mc	F$MAT1MC.01	HMG-BOX protein interacts with M-box site, cooperativity with HMG-Box STE11 protein	0.87	756	766	761	cctATTGtaga	Ins
F$YMAT	Yeast mating factors	F$MATALPHA2.02	Homeodomain transcriptional repressor Matalpha2	0.86	755	767	761	gcctaTTGTagac	Ins
F$BZIP	Fungal basic leucine zipper family	F$CST6.01	Chromosome stability, bZIP transcription factor of the ATF/CREB family (ACA2)	0.85	755	775	765	gcctattgtagACGTcaacc	Ins
F$TGRF	Telobox-containing general regulatory factor	F$TBF1.02	TTAGGG repeat binding factor 1	0.91	769	781	775	tcAACCcttttgt	Ins

Table E9: Differences in Putative TFBSs Between P_{ADH2} and $P_{ADH2\text{-}NucOpt}$

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$IRTF	Iron-responsive transcriptional activators	F$AFT2.01	Activator of Fe (iron) transcription 2, iron-regulated transcriptional activator	0.98	23	37	30	agtgaaaCACCccat	Del
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type l	F$RAP1.04	RAP1 telomeric binding sites	0.78	61	83	72	aaaaatcacACCCttatgcctgt	Ins
F$YMSE	Yeast middle sporulation elements	F$NDT80.01	Meiosis-specific transcription factor, binds to promoters containing middle sporulation elements (MSE)	0.85	77	91	84	caagccccaCAAAgtc	Del
F$YMCB	Yeast Mlu I cell cycle box	F$SWI4.01	DNA binding component of the SBF(SCB binding factor) complex (Swi4p-Swi6p)	0.84	113	125	119	tttacgCGAAaaaa	Ins
F$YMCB	Yeast Mlu I cell cycle box	F$STUAP.01	Aspergillus Stunted protein, (bHLH)-like structure, regulates multicellular complexity during asexual reproduction	0.95	114	126	120	ttttCGCGtaaat	Ins
F$RDNA	RDNA binding factor	F$REB1.02	rDNA enhancer binding protein 1, termination factor for RNA polymerase I and transcription factor for RNA polymerase II	0.85	195	207	201	tttTTACcctgcg	Ins
O$YTBP	Yeast TATA binding protein factor	O$SPT15.01	TATA-binding protein, general transcription factor that interacts with other factors to form the preinitiation complex at promoters	0.83	201	217	209	gtaaaaaTATAtagacc	Ins
O$PTBP	Plant TATA binding protein factor	O$PTATA.02	Plant TATA box	0.9	203	217	210	ggtcTATAtatttt	Ins
O$YTBP	Yeast TATA binding	O$SPT15.01	TATA-binding protein, general transcription factor that interacts with other factors to form the	0.83	202	218	210	gggtctaTATAttttta	Ins

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
	protein factor		preinitiation complex at promoters						
O$YTBP	Yeast TATA binding protein factor	O$SPT15.01	TATA-binding protein, general transcription factor that interacts with other factors to form the preinitiation complex at promoters	0.83	204	220	212	tggggtcTATAtatttt	Ins
F$YMSE	Yeast middle sporulation elements	F$MSE.01	Middle sporulation element (MSE)	0.92	213	227	220	agaccccaCAAAaaa	Ins
F$RRPE	Ribosomal RNA processing element	F$STB3.01	Ribosomal RNA processing element (RRPE)-binding protein	0.82	288	304	296	tgggcattTTTCacccc	Ins
F$ROX1	Repressor of hypoxic genes	F$ROX1.01	Heme-dependent transcriptional repressor of hypoxic genes	0.82	290	302	296	ggcaTTGTtcacc	Del
F$YMAT	Yeast mating factors	F$MATALPHA2.02	Homeodomain transcriptional repressor Matalpha2	0.86	291	303	297	gggcaTTGTtcac	Del
F$MGCM	Monomeric Gal4-class motifs	F$RGT1.02	Glucose-responsive transcription factor involved in regulation of glucose transporters	0.87	304	320	312	gaataGGGAaaaaaagct	Ins
F$YMIG	Yeast GC-Box ProteIns	F$MIG1.01	MIG1, zinc finger protein mediates glucose repression	0.84	309	327	318	tgtcaaagagtGGGGagga	Del
F$YORE	Yeast oleate response elements	F$OAF1.01	Oleate-activated transcription factor, acts alone and as a heterodimer with Pip2p	0.83	328	352	340	aatcagtatcACCGattaacacccc	Del
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type l	F$RAP1.02	RAP1 (TUF1), activator or repressor depending on context	0.82	413	435	424	gggcagtcagtcagtgCACCaat	Del
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type l	F$RAP1.04	RAP1 telomeric binding sites	0.78	435	457	446	caaccttaaACCCaaatttctta	Del

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$YSTR	Yeast stress response elements	F$RPH1.01	Jumonji-like transcription factor	0.85	440	454	447	gaaattAGGGtttaa	Ins
F$CYTO	Activator of cytochrome C	F$HAP1.01	HAP1, S. cerevisiae member of GAL family, regulates heme dependent cytochrome expression	0.76	513	527	520	cagcaactTACGgtc	Del
F$RFXP	Regulatory factor X protein, homologous to mammalian RFX1-5	F$RFX1.02	RFX1 (CRT1), acts by recruiting Ssn6 and Tup1, general repressors to the promoters of damage-inducible genes	0.77	519	533	526	tttaagcagCAACtt	Del
O$VTBP	Vertebrate TATA binding protein factor	O$ATATA.01	Avian C-type LTR TATA box	0.78	522	538	530	ttttttTAAGcagcaa	Del
F$RFXP	Regulatory factor X protein, homologous to mammalian RFX1-5	F$RFX1.01	RFX1 (CRT1) is a DNA-binding protein that acts by recruiting Ssn6 and Tup1, general repressors to the promoters of damage-inducible genes	0.92	562	576	569	ttgatttcGCAAcgg	Del
F$YMCB	Yeast Mlu I cell cycle box	F$SWI4.01	DNA binding component of the SBF(SCB binding factor) complex (Swi4p-Swi6p)	0.84	564	576	570	gcgtcgCGAAatc	Ins
F$YRSC	Yeast transcription factors remodeling chromatin structure	F$RSC30.01	Component of the RSC chromatin remodeling complex	0.81	564	584	574	gatttcgcgaCGCGcgttttt	Ins
F$YMCB	Yeast Mlu I cell cycle box	F$MBP1.01	Transcription factor binding to the MluI cell cycle box (MCB)	0.8	569	581	575	aacgCGCGtcgcg	Ins
F$YMCB	Yeast Mlu I cell cycle box	F$MCB.01	Mlu I cell cycle box, activates G1/S-specific transcription (yeast)	0.84	570	582	576	gcgaCGCGcgttt	Ins
F$YRSC	Yeast transcription factors remodeling chromatin structure	F$RSC3.01	Component of the RSC chromatin remodeling complex	0.82	566	586	576	tttcgcgacgCGCGttttttt	Ins

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$YMCB	Yeast Mlu I cell cycle box	F$STUAP.01	Aspergillus Stunted protein, (bHLH)-like structure, regulates multicellular complexity during asexual reproduction	0.95	571	583	577	aaaaCGCGcgtcg	Ins
F$YRSC	Yeast transcription factors remodeling chromatin structure	F$RSC3.01	Component of the RSC chromatin remodeling complex	0.82	567	587	577	aaaaaaaacgCGCGtcgcgaa	Ins
F$YMCB	Yeast Mlu I cell cycle box	F$MCB.01	Mlu I cell cycle box, activates G1/S-specific transcription (yeast)	0.84	572	584	578	gacgCGCGttttt	Ins
F$YRSC	Yeast transcription factors remodeling chromatin structure	F$RSC30.01	Component of the RSC chromatin remodeling complex	0.81	569	589	579	aaaaaaaaaaCGCGcgtcgcg	Ins
F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$ICGG_N10.01	Inverted CGG motifs separated by a spacer of 10 bp; Put3, Tea1 and Cha4 binding sites	0.76	571	591	581	aTGGAacaggcactcccgttg	Del
F$YHSF	Yeast heat shock factors	F$HSF1.01	Trimeric heat shock transcription factor	0.95	573	605	589	agaattgcaatcgaatGGAAcaggcactcccgt	Del
F$YHSF	Yeast heat shock factors	F$HSF1.02	Trimeric heat shock transcription factor	0.79	609	641	625	ttcatgggcaattggcagaactgggcAGAAggg	Ins
F$HOMD	Homeodomain-containing transcriptional regulators	F$YOX1.01	Yeast homeobox 1, homeodomain-containing transcriptional repressor	0.86	687	701	694	aaaaaAATTaaaaaa	Ins
F$YHSF	Yeast heat shock factors	F$HSF1.01	Trimeric heat shock transcription factor	0.95	689	721	705	gcagattctcccaaatGGAAaacga	Del

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
								gattgga c	
F$YMCM	Yeast cell cycle and metabolic regulator	F$MCM1.01	Yeast factor MCM1 cooperating with MATalpha factors	0.83	699	715	707	tctCCC Aaatgg aaaac	Del
O$YTBP	Yeast TATA binding protein factor	O$SPT15.01	TATA-binding protein, general transcription factor that interacts with other factors to form the preinitiation complex at promoters	0.83	719	735	727	tatgtaa TATA cattca	Ins
F$BZIP	Fungal basic leucine zipper family	F$CIN5.01	bZIP transcriptional factor of the yAP-1 family that mediates pleiotropic drug resistance and salt tolerance	0.89	720	740	730	gaatgta tattaca TAAA gcg	Ins
F$YMAT	Yeast mating factors	F$HMRA2.01	Hidden Mat Right A2, a2 is one of two genes encoded by the a mating type cassette in S. cerevisiae	0.94	727	739	733	gctttaT GTAat a	Ins
F$BZIP	Fungal basic leucine zipper family	F$CIN5.01	bZIP transcriptional factor of the yAP-1 family that mediates pleiotropic drug resistance and salt tolerance	0.89	725	745	735	ccggtc gctttatg TAAT ata	Ins
F$RRPE	Ribosomal RNA processing element	F$STB3.01	Ribosomal RNA processing element (RRPE)-binding protein	0.82	746	762	754	aaaaattt TTTCg gaca	Ins
F$GATC	GATA-class proteIns binding to GATC motifs	F$GAT4.01	GATA-type Zn finger protein Gat4	0.8	749	763	756	cacaG ATCttt tcgg	Del
F$GATC	GATA-class proteIns binding to GATC motifs	F$GAT4.01	GATA-type Zn finger protein Gat4	0.8	752	766	759	aaaaG ATCtgt gtag	Del
F$HOMD	Homeodomain-containing transcriptional regulators	F$YOX1.01	Yeast homeobox 1, homeodomain-containing transcriptional repressor	0.86	766	780	773	aaaaaA ATTaa aaaa	Ins

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$YMSE	Yeast middle sporulation elements	F$MSE.01	Middle sporulation element (MSE)	0.92	774	788	781	ggggagcaCAAAatg	Del
O$VTBP	Vertebrate TATA binding protein factor	O$MTATA.01	Muscle TATA box	0.84	787	803	795	gttttTAAAaagcgggg	Ins
O$VTBP	Vertebrate TATA binding protein factor	O$MTATA.01	Muscle TATA box	0.84	792	808	800	cttttTAAAaacggggg	Ins
F$YORE	Yeast oleate response elements	F$OAF1.01	Oleate-activated transcription factor, acts alone and as a heterodimer with Pip2p	0.83	792	816	804	agcgcttaccCCCGttttt aaaaag	Ins
F$RDNA	RDNA binding factor	F$REB1.02	rDNA enhancer binding protein 1, termination factor for RNA polymerase I and transcription factor for RNA polymerase II	0.85	802	814	808	cgcTTACcccccgt	Ins
F$DUIS	DAL upstream induction sequence	F$DAL82.01	Transcriptional activator for allantoin catabolic genes	0.92	829	837	833	aattCGCGc	Ins
F$YMCB	Yeast Mlu I cell cycle box	F$STUAP.01	Aspergillus Stunted protein, (bHLH)-like structure, regulates multicellular complexity during asexual reproduction	0.95	829	841	835	aattCGCGccct a	Ins

Table E9: Differences in Putative TFBSs Between P*ADH2-NucOpt* and P *ADH2- NucOpt-OptCat*

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
O$VTBP	Vertebrate TATA binding protein factor	O$MTATA.01	Muscle TATA box	0.84	729	745	737	ttacaTAAAgcgaccgg	Del
F$FBAS	Fungi branched amino acid biosynthesis	F$LEU3.02	LEU3, S. cerevisiae, zinc cluster protein	0.82	739	755	747	cgaCCGGtgtccgaaaa	Del
F$MGCM	Monomeric Gal4-class motifs	F$LYS14.01	Transcriptional activator involved in regulation of genes of the lysine biosynthesis pathway	0.8	732	748	740	acgaacGGAActttatg	Ins
O$MTEN	Core promoter motif ten elements	O$DMTE.01	Drosophila motif ten element	0.77	732	752	742	tcggacgAACGgaacttttatg	Ins
F$CSRE	Carbon source-responsive elements	F$CAT8.01	Zinc cluster transcriptional activator binding to carbon source responsive elements (CSRE)	0.86	740	754	747	tccgttcgtCCGAaa	Ins
F$MGCM	Monomeric Gal4-class motifs	F$RGT1.02	Glucose-responsive transcription factor involved in regulation of glucose transporters	0.87	740	756	748	tttttCGGAcgaacgga	Ins
F$YORE	Yeast oleate response elements	F$OAF1.01	Oleate-activated transcription factor, acts alone and as a heterodimer with Pip2p	0.83	738	762	750	gttccgttcgTCCGaaaaaattttt	Ins

APPENDIX F

Differences in putative TFBSs between P_{AOX1} and Its Variants

Table F1: Differences in Putative TFBSs Between P_{AOX1} and $P_{ADH2\text{-}AOX1\text{-}Adr1}$

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$Y ADR	Yeast metabolic regulator	F$ADR1.01	Alcohol Dehydrogenase Regulator, carbon source-responsive zinc-finger transcription factor	0,94	140	148	144	caCCCCaat	Ins
F$Y MIG	Yeast GC-Box ProteIns	F$MIG1.01	MIG1, zinc finger protein mediates glucose repression	0,84	138	156	147	caaataatattGGGGtgtg	Ins
F$Y MIG	Yeast GC-Box ProteIns	F$MIG1.01	MIG1, zinc finger protein mediates glucose repression	0,84	145	163	154	caatattatttGGGGtact	Ins
F$Y ADR	Yeast metabolic regulator	F$ADR1.01	Alcohol Dehydrogenase Regulator, carbon source-responsive zinc-finger transcription factor	0,94	153	161	157	taCCCCaaa	Ins

Table F2: Differences in Putative TFBSs Between P_{AOX1} and $P_{ADH2\text{-}AOX1\text{-}Adr2}$

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$YMIG	Yeast GC-Box ProteIns	F$MIG2.01	MIG2, zinc finger protein mediates glucose repression	0,81	355	373	364	tttctgagtgtGGGGtcaa	Del
F$YMIG	Yeast GC-Box ProteIns	F$MIG1.01	MIG1, zinc finger protein mediates glucose repression	0,84	353	371	362	aaaaaaaatgtGGGGtcgc	Ins
F$YADR	Yeast metabolic regulator	F$ADR1.01	Alcohol Dehydrogenase Regulator, carbon source-responsive zinc-finger transcription factor	0,94	374	382	378	gaCCCCaca	Ins
F$YMIG	Yeast GC-Box ProteIns	F$MIG1.01	MIG1, zinc finger protein mediates glucose repression	0,84	372	390	381	ggggaacatgtGGGGtcaa	Ins

Table F3: Differences in Putative TFBSs Between P*AOX1* and P *ADH2- AOX1-Adr3*

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$PHD1	Pseudohyphal determinant 1	F$PHD1.01	Transcription factor involved in regulation of filamentous growth	0,89	669	681	675	acaaTGCAtaatc	Del
F$ICGG	Inverted CGG triplets spaced preferentially by 10 bp	F$ICGG_N10.01	Inverted CGG motifs separated by a spacer of 10 bp; Put3, Tea1 and Cha4 binding sites	0,76	663	683	673	tTGGAtgattaacccccaatac	Ins
F$TGRF	Telobox-containing general regulatory factor	F$TBF1.01	TTAGGG repeat binding factor 1	0,82	671	683	677	ttaacCCCAatac	Ins
F$YADR	Yeast metabolic regulator	F$ADR1.01	Alcohol Dehydrogenase Regulator, carbon source-responsive zinc-finger transcription factor	0,94	673	681	677	aaCCCCaat	Ins
F$TGRF	Telobox-containing general regulatory factor	F$TBF1.01	TTAGGG repeat binding factor 1	0,82	684	696	690	gcaacCCCAaaat	Ins
F$YADR	Yeast metabolic regulator	F$ADR1.01	Alcohol Dehydrogenase Regulator, carbon source-responsive zinc-finger transcription factor	0,94	686	694	690	aaCCCCaaa	Ins

Table F4: Differences in Putative TFBSs Between P*AOX1* and P *ADH2- AOX1-Cat1*

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$YSTR	Yeast stress response elements	F$MSN2.01	Transcriptional activator for genes in multistress response	1	90	104	97	aacaggaGGGGatac	Del
F$YMAT	Yeast mating factors	F$MATALPHA2.02	Homeodomain transcriptional repressor Matalpha2	0,86	87	99	93	cggaaTTGTtgcg	Ins
F$MGCM	Monomeric Gal4-class motifs	F$LYS14.01	Transcriptional activator involved in regulation of genes of the lysine biosynthesis pathway	0,8	88	104	96	acgaacGGAAttgttgc	Ins
O$MTEN	Core promoter motif ten elements	O$DMTE.01	Drosophila motif ten element	0,77	88	108	98	tcggacgAACGgaattgttgc	Ins
F$CSRE	Carbon source-responsive elements	F$CAT8.01	Zinc cluster transcriptional activator binding to carbon source responsive elements (CSRE)	0,86	96	110	103	tccgttcgtCCGAtt	Ins
F$MGCM	Monomeric Gal4-class motifs	F$RGT1.02	Glucose-responsive transcription factor involved in regulation of glucose transporters	0,87	96	112	104	ctaatCGGAcgaacgga	Ins
F$GATA	Fungal GATA binding factors	F$GAT1.01	GATA-type Zn finger protein Gat1	0,83	103	117	110	gtccgATTAgcagac	Ins

Table F5: Differences in Putative TFBSs Between P*AOX1* and P *ADH2- AOX1-Cat2*

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$RFXP	Regulatory factor X protein, homologous to mammalian RFX1-5	F$RFX1.02	RFX1 (CRT1), acts by recruiting Ssn6 and Tup1, general repressors to the promoters of damage-inducible genes	0,77	302	316	309	ttccgaatgCAACaa	Del
F$YMCM	Yeast cell cycle and metabolic regulator	F$MCM1.02	Yeast factor MCM1 cooperating with MATalpha factors	0,75	301	317	309	tTTCCgaatgcaacaag	Del
F$PRES	Pheromone response elements	F$STE12.01	Transcription factor activated by a MAP kinase signaling cascade, activates genes involved in mating or pseudohyphal/invasive growth pathways	0,86	306	318	312	gaatgcAACAagc	Del
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type l	F$RAP1.02	RAP1 (TUF1), activator or repressor depending on context	0,82	312	334	323	aacaagctccgcattaCACCcga	Del
F$IRTF	Iron-responsive transcriptional activators	F$AFT2.01	Activator of Fe (iron) transcription 2, iron-regulated transcriptional activator	0,98	321	335	328	cgcattaCACCcgaa	Del
F$CSRE	Carbon source-responsive elements	F$RDS2.01	Regulator of drug sensitivity 2, zinc cluster transcriptional activator involved in conferring resistance to ketoconazole	0,85	322	336	329	gcattacacCCGAac	Del
F$YMAT	Yeast mating factors	F$HMRA2.01	Hidden Mat Right A2, a2 is one of two genes encoded by the a mating type cassette in S. cerevisiae	0,94	323	335	329	ttcgggTGTAatg	Del

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$YRAP	Yeast activator of glycolyse genes / repressor of mating type l	F$RAP1.03	RAP1 (TUF1), activator or repressor depending on context	0,87	320	342	331	ccgcattacACCCgaacatcact	Del
F$YMCB	Yeast Mlu I cell cycle box	F$SWI4.01	DNA binding component of the SBF(SCB binding factor) complex (Swi4p-Swi6p)	0,84	326	338	332	tacaccCGAAcat	Del
F$CSRE	Carbon source-responsive elements	F$CAT8.01	Zinc cluster transcriptional activator binding to carbon source responsive elements (CSRE)	0,86	311	325	318	ccctctcgtCCGGgc	Ins
F$YGAL	Yeast GAL4 factor	F$GAL4.01	GAL4 transcriptional activator in response to galactose induction	0,75	314	338	326	tctcgtccgggcttttttCCGAacat	Ins
F$ASG1	Activator of stress genes	F$ASG1.01	Fungal zinc cluster transcription factor Asg1	0,79	319	335	327	tCCGGgctttttccgaa	Ins
F$ARPU	Regulator of pyrimidine and purine utilization pathway	F$PPR1.01	Pyrimidine pathway regulator 1	0,75	320	336	328	ccgggcttttttCCGAac	Ins
F$MGCM	Monomeric Gal4-class motifs	F$RGT1.02	Glucose-responsive transcription factor involved in regulation of glucose transporters	0,87	322	338	330	atgttCGGAaaaagccc	Ins
F$HAL9	Halotolerance 9	F$HAL9.01	Zinc cluster transcription factor HAL9, involved in salt tolerance	0,86	329	347	338	TCTGgagtgatgttcggaa	Ins
F$HAL9	Halotolerance 9	F$HAL9.01	Zinc cluster transcription factor HAL9, involved in salt tolerance	0,86	330	348	339	TCCGaacatcactccagat	Ins

Table F6: Differences in Putative TFBSs Between P*AOX1* and P *ADH2- AOX1-Cat3*

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$WHBS	Winged helix binding sites	F$FHL1.01	Fork head transcription factor Fhl1	0,84	635	647	641	gaaACGCaaatgg	Del
F$MGCM	Monomeric Gal4-class motifs	F$RGT1.02	Glucose-responsive transcription factor involved in regulation of glucose transporters	0,87	636	652	644	acgaaCGGAatatgtt	Ins
F$CSRE	Carbon source-responsive elements	F$CAT8.01	Zinc cluster transcriptional activator binding to carbon source responsive elements (CSRE)	0,86	644	658	651	tccgttcgtCCGAat	Ins
F$MGCM	Monomeric Gal4-class motifs	F$RGT1.02	Glucose-responsive transcription factor involved in regulation of glucose transporters	0,87	644	660	652	agattCGGAcgaacgga	Ins

Table F7: Differences in Putative TFBSs Between P_{AOX1} and $P_{ADH2\text{-}AOX1\text{-}Aca}$

Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	Score	Start position	End position	Anchor position	Sequence	Modification
F$AZF1	Asparagine-rich zinc finger 1	F$AZF1.01	Glucose-dependent zinc finger transcriptional activator	1	506	514	510	aAAAAgaaa	Del
F$YMCM	Yeast cell cycle and metabolic regulator	F$MCM1.02	Yeast factor MCM1 cooperating with MATalpha factors	0,75	515	531	523	cTTCCaaaagtcggcat	Del
F$MOT3	Modifier of transcription 3	F$MOT3.01	Modifier of transcription 3, zinc finger transcription factor involved in repression of a subset of hypoxic genes	0,91	489	509	499	acaatAGGCactggccgttag	Ins
F$MMAT	M-box interacting with Mat1-Mc	F$MAT1MC.01	HMG-BOX protein interacts with M-box site, cooperativity with HMG-Box STE11 protein	0,87	502	512	507	cctATTGtaga	Ins
F$YMAT	Yeast mating factors	F$MATALPHA2.02	Homeodomain transcriptional repressor Matalpha2	0,86	501	513	507	gcctaTTGTagac	Ins
F$BZIP	Fungal basic leucine zipper family	F$CST6.01	Chromosome stability, bZIP transcription factor of the ATF/CREB family (ACA2)	0,85	501	521	511	gcctattgtagACGTcaacc	Ins

CURRICULUM VITAE

PERSONAL INFORMATION

Name: Burcu Gündüz Ergün

Nationality: Turkish (TC)

Date of Birth: November 8, 1987

E-mail: burcugunduz3@gmail.com

EDUCATION

Degree	Institution	Year of Graduation
MS	University of Edinburgh, Biotechnology	2014
MS	METU, Biotechnology	2012
BS (Major)	Hacettepe University, Food Engineering	2010
BS (Minor)	Hacettepe University, Chemical Engineering	2010
High School	Lüleburgaz Atatürk High School	2005

WORK EXPERIENCE

Year	Place	Enrollment
2014-	Ministry of Agriculture and Forestry, Biotechnology Research Center, Ankara	Engineer
2017-2018	Uni. Of Natural Resources and Life Sciences, Austria	Visiting Sci.

AWARDS, DISTINCTIONS AND FELLOWSHIPS

- The Scientific and Technological Research Council of Turkey (TUBITAK), International Research Scholarship for PhD studies, 2017-2018.
- The Scientific and Technological Research Council of Turkey (TUBITAK), National PhD Scholarship, 2014-2016
- Turkish Higher Education System Scholarship for MSc in the UK
- METU Graduate Courses Performance Award, 2012-2013
- The Scientific and Technological Research Council of Turkey (TUBITAK), National MSc Scholarship, 2010-2012
- METU Graduate Courses Performance Award, 2010-2011
- The Scientific and Technological Research Council of Turkey (TUBITAK), National Undergraduate Research Fund, 2009-2010

PUBLICATIONS

- Gündüz Ergün B, Hüccetoğulları D, Öztürk S, Çelik E, Çalık P. in press: *2018*. Established and upcoming yest expression systems. *Recombinant Protein Prodution in Yeast-Methods in Molecular Biology*, Brigitte Gasser and Diethard Mattanovich (ed), Humana Press

- Watkins C, Bartley DJ, Gündüz Ergün B, Yıldızhan B, Ross-Watt T, Morrison AA, Sanmartín MJR, Strathdee F, Andrews L, Free A. in press. Effect of monepantel administration on the ovine intestinal microbiota of sheep infected with sensitive and resistant strains of Teladorsagia circumcincta. *International Journal for Parasitology.*

- Öztürk S, Gündüz Ergün B, Çalık P. 2017. Double promoter expression systems for recombinant protein production by industrial microorganisms. *Applied Microbiology and Biotechnology*. 101(20):7459-7475

- Gündüz Ergün B, Çalık P. 2016. Lignocellulose degrading extremozymes produced by Pichia pastoris: Current state and future perspectives. *Bioprocess and Biosystems Engineering*. 39(1):1-36

PRESENTATIONS AND ABSTRACTS

- Gündüz Ergün B. 2016. Potential applications of transglutaminase enzyme in cereal products. International Cereal and Bread Congress, İstanbul

- Gündüz Ergün B, Çalık P. 2016. Extracellular microbial pro-transglutaminase production by recombinant *Pichia pastoris*. Pichia 2016 Conference, Antalya

- Gündüz B, Çalık P. 2012. "Recombinant transglutaminase production by metabolically engineered *Pichia pastoris*", 15. European Biotechnology Congress, İstanbul.

- Gündüz B, Çalık P. 2012. Transglutaminase Production Using Modified *Pichia pastoris* with Metabolic Engineering, International Food, Agriculture and Gastronomy Congress, Antalya