Institute for Molecular BioTechnology

# Expression of recombinant proteins in *Pichia pastoris* employing synthetic promoters

Dissertation

Claudia Ruth

zur Erlangung des akademischen Grades einer Doktorin der technischen Wissenschaften

an der
Technischen Universität Graz
in Zusammenarbeit mit

VTU technology

Glieder, Anton, Ao.Univ.-Prof. Mag.rer.nat. Dr.rer.nat.
Institut für Molekulare Biotechnologie
Technische Universität Graz
2010

## EIDESSTATTLICHE ERKLÄRUNG

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

Graz, am …… ……..

(Unterschrift)

Englische Fassung:

## STATUTORY DECLARATION

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

…… ……..

date (signature)

## Abstract

The ability to perform posttranslational modifications like S-S bridge formation or glycosylation in combination with simple handling made *Pichia pastoris* a favourable host system for the production of recombinant proteins. Employing the strong methanol induced *AOX1* or constitutive *GAP* promoter many proteins have been successfully expressed using this host system.

The promoter library designed by Hartner et al. (2008) provides a toolbox to fine-tune *Pichia pastoris* gene expression. Based on this previous study different promoter variants were applied for the expression of selected model targets (e.g. porcine trypsinogen, plant epoxide hydrolases). Using different promoters for the expression of porcine trypsinogen resulted in variable production windows depending on the promoter used. For example, using a small synthetic promoter consisting of a basal promoter fragment with one attached transcription factor binding site motif, which is derepressed upon glucose depletion, resulted in 10-fold increased activity after batch growth on glucose if compared to the wild type *AOX1* promoter at this time point. Differently, application of various promoter variants for the expression of potato epoxide hydrolase helped to define the limits reachable through transcriptional tuning. Further, a highly competitive low copy strain expressing potato epoxide hydrolase was generated.

Moreover, exploiting the knowledge about possible regulatory sites, the development of new synthetic promoters was targeted. In GFP screening studies a novel promoter variant based on deletion of a specific region and triplication of the putative Mat1-Mc transcription factor binding site, increased activity in absence of glucose as well as presence of methanol in comparison to the wild type *AOX1* promoter. In addition and as a first step in the generation of a fully artificial promoter, novel synthetic core promoter elements able to initiate transcription were generated.

## Zusammenfassung

Die Fähigkeit zu posttranslationellen Modifikationen wie zum Beispiel Disulfidbrückenbildung und Glykosylierung in Kombination mit einfacher Handhabung machten die Hefe *Pichia pastoris* zu einem bedeutenden Wirtssystem für rekombinante Proteinproduktion. Unter Anwendung des durch Methanol stark induzierbaren *AOX1* und konstitutiven *GAP* Promotors konnten zahlreiche heterologe Proteine erfolgreich exprimiert werden. Zusätzlich bietet die von Hartner et al. (2008) generierte Promotor-Bibliothek eine Möglichkeit zur Feinabstimmung der Genexpression.

Aufbauend auf der Arbeit von Hartner wurden hier unter Anwendung verschiedener Promotorvarianten ausgewählte Modellproteine exprimiert. Die Anwendung verschiedener Promotorvarianten zur Herstellung von Schweine-Trypsinogen resultierte in verschiedenen Produktionsmöglichkeiten, je nachdem welcher Promotor verwendet wurde. Zum Beispiel, verglichen mit dem Wildtyp *AOX1* Promotor, konnte mit einem kurzen, synthetischen Promotor, welcher auf einer Fusion eines Kern-Promotorelements mit einer Transkriptionsfaktorbindestelle basiert und über die Glukosekonzentration reguliert ist, 10-mal mehr Trypsinogen bis zum Ende der Wachstums Phase produziert werden. Weiters konnte durch die Anwendung verschieden starker Promotorvarianten zur Herstellung von Kartoffel-Epoxidhydrolase das durch Transkriptionsoptimierung erreichbare Limit festgelegt und ein wettbewerbsfähiger Expressionsstamm mit niedriger Kopienzahl generiert werden.

Basierend auf dem Wissen über positiv und negativ regulierende Elemente im *AOX1* Promotor wurden neue synthetische Promotorvarianten generiert. In Screeningstudien mit GFP als Reporter Protein steigerte eine neue Promotorvariante, welche auf der Deletion einer bestimmten Region und Verdreifachung der putativen Mat1-Mc Transkriptionsfaktorbindestelle basiert, die Aktivität sowohl mit als auch ohne Methanol Induktion im Vergleich zum Wildtyp *AOX1* Promotor. Als erster Schritt in Richtung eines vollsynthetischen Promotors wurden zusätzlich erstmals synthetische *AOX1* Kern-Promotorelemente hergestellt.

## Acknowledgements

Special thanks go to my boyfriend Mario, my Mum, my Dad and all other family members for their trust, support and understanding. Many thanks also to all the people of the Glieder-Group for a nice working atmosphere. Especially I want to thank the people who contributed to the scientific work: Thomas Zuellig (Diploma thesis), Stefan Ertl (Diploma thesis), Viktorija Vidimce (student), Katrin Weinhandl, Piyush Labhsetwar (student), Astrid Schönbacher (student) and Ariane Klatzer (student). I also want to thank Toni Glieder for this challenging task and his guidance. In addition, I thank VTU-Technology, Roland Weis and Thomas Purkarthofer, for financial and scientific support.

# Table of Contents


Abstract ..... 3
Zusammenfassung ..... 4
Acknowledgements ..... 5
Table of Contents ..... 6
Aim of this thesis ..... 9
Abbreviations ..... 10
1 Synthetic promoter design ..... 12
1.1 Introduction ..... 12
1.1.1 The transcription ..... 12
1.1.2 Transcription factors ..... 13
1.1.3 Promoter design - an in silico approach? ..... 13
1.1.4 Perspectives on Synthetic Promoters for Biocatalysis and Biotransformation ..... 14
1.2 Materials and methods ..... 23
1.2.1 Plasmids ..... 23
1.2.2 Strains ..... 24
1.2.3 Site directed mutagenesis and overlap extension PCR ..... 25
1.2.4 Promoter variants ..... 25
1.2.5 Transformation, screening and fluorescence detection ..... 27
1.3 Experiments and results ..... 27
1.3.1 New generation promoter variants I ..... 27
1.3.2 New generation promoter variants II ..... 33
1.3.3 New generation promoter variants III ..... 37
1.4 References ..... 44
1.5 Annex ..... 47
2 Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in *Pichia pastoris* ..... 52
2.1 Introduction ..... 53
2.2 Materials and methods ..... 54
2.2.1 Strains and plasmids ..... 54
2.2.2 Chemicals and media ..... 55
2.2.3 AOX1 promoter variants ..... 56
2.2.4 Design of the AOX1 promoter variants ..... 56
2.2.5 Generation of the AOX1 promoter variants ..... 57
2.2.6 Transformation, screening and rescreening ..... 59
2.2.7 Fed batch cultivations in bioreactors ..... 59
2.2.8 Trypsin activity assay ..... 60
2.2.9 Specific productivity ..... 60
2.2.10 Protein analysis ..... 60

2.2.11 Real Time PCR ..... 60
2.3 Results ..... 61
2.3.1 Rescreening ..... 61
2.3.2 Copy number estimation ..... 62
2.3.3 Fedbatch cultivations in bioreactors ..... 63
2.4 Discussion ..... 68
2.4.1 Conclusions ..... 69
2.5 Acknowledgments ..... 69
2.6 References ..... 70
2.7 Annex ..... 71
*3* Heterologous expression of soluble plant epoxide hydrolases in *Pichia pastoris* ..... 72
3.1 Introduction ..... 72
3.2 Materials and methods ..... 73
3.2.1 Strains and expression cassettes ..... 73
3.2.2 AOX1 promoter variants ..... 74
3.2.3 Chemicals and media ..... 74
3.2.4 Transformation ..... 75
3.2.5 Cultivation ..... 75
3.2.6 Conversions ..... 75
3.2.7 Thin layer chromatography (TLC) ..... 76
3.2.8 HPLC ..... 76
3.2.9 Red assay ..... 76
3.2.10 Protein concentration and purification ..... 77
3.2.11 Cell lysis ..... 77
3.2.12 Real Time PCR ..... 77
3.3 Results ..... 77
3.3.1 Recombinant expression of 3 plant soluble epoxide hydrolases ..... 77
3.3.2 Euphorbia lagascae epoxide hydrolase ..... 78
3.3.3 Up-scaling StEH1 expression ..... 79
3.3.4 Employing promoter technology for the expression of StEH1 ..... 80
3.4 Discussion ..... 82
3.5 References ..... 83
3.6 Supplementary ..... 85
3.7 Annex ..... 85
4 High level expression of *Bungarus fasciatus* acetylcholinesterase ..... 87
4.1 Introduction ..... 87
4.2 Materials and methods ..... 90
4.2.1 Expression cassettes ..... 90
4.2.2 Strains ..... 90
4.2.3 AOX1 promoter variants ..... 90

4.2.4 Chemicals and media ..... 91
4.2.5 Transformation and cultivation ..... 91
4.2.6 Activity assay ..... 91
4.2.7 Activity staining ..... 91
4.2.8 Protein analysis ..... 92
4.3 Results and discussion ..... 92
4.3.1 Bioreactor cultivations ..... 95
4.4 Conclusions ..... 97
4.5 References ..... 98
4.6 Annex ..... 100
List of Figures ..... 102
List of Tables ..... 106

## Aim of this thesis

Based on the studies of Hartner et al. (2008), in which *AOX1* promoter variants of different strength as well as short synthetic promoters with altered regulatory abilities were generated, the goal of this thesis was to show the applicability of synthetic promoters with different properties for recombinant protein expression in *Pichia pastoris.* Following model proteins, relevant for industry and difficult to express, were chosen for expression studies employing promoter technology: porcine trypsinogen, potato, soy bean and spurge epoxide hydrolase and snake acetylcholinesterase.

Besides the applicability of the synthetic promoters, also the generation of new *AOX1* promoter variants as well as short synthetic promoters with superior properties, in regard to methanol and derepression induced protein production, was targeted. As a final step in the design of a synthetic promoter the generation of an artificial core promoter fragment was in addition aim of this study.

## Abbreviations

| | |
|---|---|
| AA..... | amino acid |
| AChE..... | acetylcholinesterase |
| AOX1..... | alcohol oxidase 1 |
| bp..... | base pairs |
| BSA..... | bovine serum albumin |
| C..... | carbon |
| DNA..... | desoxyribonucleic acid |
| EU..... | Ellman units |
| fw..... | forward |
| g..... | gram |
| GFP..... | green fluorescent protein |
| h..... | hour |
| HPLC..... | high performance liquid chromatography |
| Kan..... | kanamycin |
| kDa..... | kilodalton |
| kV..... | kilovolt |
| l..... | liter |
| LB..... | Luria-Bertani Broth |
| MeOH..... | methanol |
| µF..... | microfarad |
| min..... | minute |
| nm..... | nanometer |
| mM..... | millimolar |
| no..... | number |
| OD..... | optical density |
| OE-PCR..... | overlap-extension polymerase chain reaction |
| PTS1..... | peroxisomal targeting sequence 1 |
| RFU..... | relative fluorescence units |
| rpm..... | rounds per minute |
| RT-PCR..... | Real Time PCR |
| rv..... | reverse |
| s..... | second |
| sc..... | single copy |
| SDS-PAGE..... | sodium dodecylsulfate polyacrylamide gel electrophoresis |
| sEH..... | soluble epoxide hydrolase |
| *..... | star |

| | |
|---|---|
| TF..... | transcription factor |
| TLC..... | thin layer chromatography |
| TSO..... | trans-stilbene oxide |
| U..... | unit |
| w/w..... | weight per weight |
| WT..... | wild type |
| YNB..... | yeast nitrogen base |
| Zeo..... | zeocin |

# 1 Synthetic promoter design

## 1.1 Introduction

### 1.1.1 The transcription

Knowledge of certain promoter features and understanding of a promoters principle function are indispensable for the generation of artificial promoters (Heintzman and Ren, 2007; Juven-Gershon and Kadonaga, 2009). Briefly, transcription is regulated by chromatin modifying proteins, transcription factors (TF) and co-factors (Hannenhalli, 2008; Wasserman and Sandelin, 2004). The transcriptional promoter can be seen as the control point of all levels of regulation integrating transcription, epigenic features and signal transduction events (Heintzman and Ren, 2007). The term promoter comprises a core promoter including the transcription start site (TSS). One or multiple repeats of cis-regulatory elements can be found in the upstream promoter binding to enhancers or complexes of such. For transcription initiation the preinitiation complex (PIC) is formed at the core promoter by binding the general transcription factors and RNA polymerase (RNAP) II. In prokaryotes only the core and σ subunit of the RNAP holoenzyme are necessary to initiate transcription (Novina and Roy, 1996). However, also upstream RNAPII α subunit recognition elements with a strong influence on transcriptional activity were found (Gourse et al., 2000; Ross et al., 1993; Zhou and Yang, 2006). In eukaryotes several conserved sequence motifs can be found in a core promoter binding the general TFs (Heintzman and Ren, 2007; Juven-Gershon et al., 2009). The TATA box, located 25-30 nucleotides upstream of TSS, binds the TATA binding protein subunit of the general transcription factor TFIID (Struhl, 1995). Both sites of the TATA box can be flanked by BRE regions (TFIIB recognition elements) (Lagrange et al., 1998). It was found that BRE elements can stimulate or repress transcription. Another common core promoter feature is the initiator element (Inr), which is located directly at TSS and found in promoters containing or lacking a TATA box (Smale and Baltimore, 1989). If TATA and Inr are present, they act synergistically. Two downstream promoter elements have been identified: DPE (downstream promoter element) and MTE (motif ten element) (Burke and Kadonaga, 1997; Burke et al., 1998; Ohler et al., 2002). DPE and MTE were found to be conserved between *Drosophila* and humans (Juven-Gershon et al., 2009).
The eukaryotic promoter database (EPD) provides a yearly updated collection of annotated non-redundant RNAPII promoters (Perier et al., 2000).

Two different modes of transcriptional initiation are reported (Juven-Gershon et al., 2006a,b; Juven-Gershon and Kadonaga, 2009; Smale and Kadonaga, 2003). Focused initiation occurs only at one or a small set of nucleotides and is predominantly found in simpler organisms. Dispersed initiation can include several weak start sites over a region of 100 nucleotides (mainly within CpG island) and is commonly found through vertebrate genes (Juven-Gershon and Kadonaga, 2009). Promoters enabling dispersed initiation generally lack TATA, DPE and MTE motifs. There is indication that dispersed initiation is connected to constitutive expression. In

contrary, focused initiation seems to be associated with regulated genes, suggesting one TSS easier to regulate than several (Carninci et al., 2006).
Regulating eukaryotic chromatin structure histone modification is an important determinant in gene expression. Recent studies even map promoters active or inactive due to their chromatin signatures (Liu et al., 2005). Modifications connected to promoter activity are acetylation of histone H3 and H4 and methylation of histone variant H3K4 (Pokholok et al., 2005). Applying high resolution nucleosome mapping in yeast a nucleosome free region (NFR) was found located 200 base pairs upstream of the start codon (Yuan et al., 2005). The found NFR was flanked by histone variant H2A.Z, which is assumed to play a role in NFR formation or maintenance (Raisner et al., 2005). Generally, NFRs are assumed to play a role in TF and RNAPII positioning by allowing higher chromatin accessibility at promoters or regulatory elements (Felsenfeld, 1996). A detailed description of these epigenic effects can be found in Heintzman and Ren (2007).

### 1.1.2 Transcription factors

Recently available genomic DNA sequences enabled to study gene regulatory networks and TFs as whole network or specific parts of it (Westholm et al., 2008). Experimental and computational methods to elucidate gene regulatory networks and cis-regulatory elements are reviewed in detail by Elnitski et al. (2006), Hannenhalli (2008), Tompa et al. (2005) and van Nimwegen (2007). Briefly, experimental methods for the identification of TF binding sites include the identification of DNAse I hypersensitive regions indicating NFRs (DNA footprinting) and deletion/mutation assays. Common also ChIP-chip or ChIP-seq methods, which enable TF identification after DNA-protein linkage and immunoprecipitation with antibodies. Computational methods advanced the identification of transcription factor binding sites in recent years, but often large scale accuracy was low and sensitivity and specificity imbalanced (Hannenhalli, 2008). Many of the computational methods are based on experimental identified data collected in TF databases such as JASPAR (open access, Sandelin et al., 2004), TRANSFAC (partly licensed, Matys et al., 2006) or MatInspector (free trial license, Cartharius et al., 2005). In these databases TF binding sites are converted into position weight matrixes (PMWs), which are probabilistic representations (Stormo, 2000). An extension of PWM algorithms are PMM (phylogenetic motif models) scanning algorithms, which scan multiple alignments of orthologous sequences (Hawkins et al., 2009). Still, sequence based models do not take the dynamic cell state into account leafing the *in vivo-in vitro* gap (Hannenhalli, 2008). Recent models also incorporate nucleosome occupancy, nucleosome positioning as well as unmethylated CpG islands (Fang et al., 2006; Noble et al., 2005; Segal et al., 2006). A quantitative model of transcription factor activated gene expression was published by Kim and O'Shea (2008).

### 1.1.3 Promoter design - an in silico approach?

Similar to protein engineering a promoter engineer can choose between rational, semi-rational or directed evolution approaches. Taking the rational approach Venter (2007) describes the *in-silico* design of an artificial promoter. The instructions of Venter (2007) provide a simple alternative. The challenge lies in the identification of common motifs associated to different conditions (Pilpel et al., 2001). However, the complexity of transcriptional regulation might lead to differences between design and reality. Library and semi-rational

approaches can be used to overcome such limitations and in addition generate new knowledge for future *in-silico* approaches.

User Manual by Venter (2007):

1. Identification of useful cis-regulatory motifs from clusters of co-expressed genes.
2. Construction of a motif synergy map by a Boolean AND-NOT-OR cis-motif logic* (see also Pilpel et al., 2001). The created map evaluates motif occurrence in regard to specific conditions and in relation to other motifs. Motif spacing relative to TATA is also considered.
3. Evaluation of the selected combinations of cis-motifs in an expression study after fusion to a core promoter.

* According to Kinkhabwala and Guet (2008) promoters can be understood as DNA-based processing units which use TR inputs to convert signals into ON and OFF transcriptional outputs. Such processing units can be described in the computational language of logic functions (e.g. ANDN logic: A and not B).

### 1.1.4 Perspectives on Synthetic Promoters for Biocatalysis and Biotransformation

(published as concept paper in Chembiochem. 2010 Feb 28;11(6):761-765.)

Claudia Ruth and Anton Glieder
Institute of Molecular Biotechnology
Graz University of Technology
Petersgasse 14, A-8010 Graz, AUSTRIA
Fax: (+43) 316 873 4071
E-mail: a.glieder@tugraz.at

**Abstract**

Acting on the transcriptional level, synthetic promoters have been useful tools for controlling gene expression and have applications in many fields. Here, we discuss synthetic promoters and libraries in regard to current and future applications in the field of biocatalysis or biotransformation. We also focus on synthetic promoter design principles and distinguish between prokaryotic and eukaryotic destinations. The natural toolboxes available for tuneable gene expression and the regulation of enzyme function are limited and primarily host specific. Synthetic biology offers generally applicable concepts and quick implementation. Smart alternatives to transcriptional regulation enrich the engineer's tool box for optimizing industrial enzyme production and host-cell physiology for whole-cell processes. Industrially applicable, tuneable enzyme cascades and artificial circuits for iterative up- and down-regulation will soon be achieved.

Synthetic biologists investigate nature and engineer natural systems for tasks[1] involving biocatalysis and biotransformation. These investigations take advantage of two major aspects of synthetic biology:
A) The design and construction of new biological parts and systems;
B) The redesign of existing natural biological systems for useful purposes.
A common first engineering step is the regulation and (re-) design of defined parts of the whole system. In all cases, biocatalyst expression must be balanced with other enzymes; when they act as isolated enzymes or as enzymes in a natural environment (i.e. growing or resting cells), and when they are overexpressed, or take part in multi-component catalytic systems. The generation of improved whole cell biocatalysts or production strains (Designer Bugs) has relied heavily on rational engineering of metabolic pathways and harnessing the catalytic potential of microorganisms with gene knock-out or overexpression strategies. However, during overexpression of recombinant proteins, many stress induced problems occur that cannot be resolved by simple engineering steps. In addition, unbalanced overexpression of single genes generates new bottle necks in production strains; moreover, co-overexpression of several genes raises the complexity and poses a challenge to common improvement strategies. Consequently, synthetic biology for simple, reliable, and controllable gene expression may have an unexpected impact that increases with the complexity of the catalytic system.

On the other hand, genomes and metabolic activities can be minimized and restructured within natural systems to the level of complexity that is necessary and sufficient for technological applications.[2-4] We expect that synergies from both these strategies might shape future developments and success in industrial biotechnology. Here, we present the roots of the most recent developments to provide perspective on the future impact that synthetic biology might have on biocatalysis and biotransformation. We will focus on artificial regulatory elements.

#### 1.1.4.1 Control point transcription

Synthetic promoters have proven to be useful for transcriptional regulation of gene expression, with applications in many fields.[5,6] The rising popularity of synthetic promoters or promoter libraries might be explained by the severity of common genetic methods; for example, gene knockouts or overexpression.[7] In a few cases, these “on or off” approaches led to the desired results, but in many other cases this strategy failed due to the complicated interplay of several factors.[8] In fact, some proteins can be produced at very high levels without causing negative effects on the cell, while others can harm the cell with very little expression or excessive expression above a certain limit. Every protein behaves differently and thus, requires different adjustments in expression level or strategy. Promoter technology is a proven tool for identifying optimal promoter gene combinations and fine-tuning individual expression. Current knowledge of promoter elements and regulation was reviewed previously[9,10] and provides a foundation for generating artificial promoters and circuits.

#### 1.1.4.2 Minimal Synthetic Promoters for Prokaryotes

There are two main methods of generating synthetic promoter libraries.[7] Briefly, the first method uses oligonucleotides with randomized promoter sequences and a 3’ homologous target gene region.[11] With this

method, the promoter variants and target gene can be amplified together, cloned, and used in expression experiments. In a second, alternative method, mutagenic PCR is used for promoter variant generation.[12] Also recent advanced technologies offer simple and efficient access to highly diversified sequences by the chemical synthesis of large oligonucleotides of 100-200 bases. These can be particularly interesting for the generation of small synthetic promoters.

A cornerstone for prokaryotic synthetic promoter design and generation was set by Jensen,[13] who showed the importance of spacer sequences surrounding the two consensus boxes at -10 and -35 bases from the transcription start site. A library of synthetic promoters of variable strengths was created for Lactococcus lactis by randomizing the nucleotide composition of the spacer and its surrounding sequence (Figure 1.1). Improved promoter variants relied on intact consensus boxes and a spacer length of 17 bp; weaker promoters resulted from changes in the consensus regions or reduced spacer length.[13,14] Based on this initial study, many later studies used the same principle of randomized spacing to generate synthetic promoters with various strengths, ranging from several orders of magnitude in reporter protein expression.[11,12,14] It was also shown that the activity of *Escherichia coli* promoters could be increased with two TG repeats at position -17 to -14 from the transcription start.[15]

The successful synthetic approach of Jensen[13] was also applicable to prokaryotic organisms other than *E. coli* and *L. lactis*.[16] Further, the principle of randomized flanking sequences was adapted to generate differently regulated stationary-phase promoters for inducible expression in *E. coli*.[17] These promoters were characterized by binding the σS, but not the σ70, subunit of RNAP, thus slightly different consensus motifs were applied.

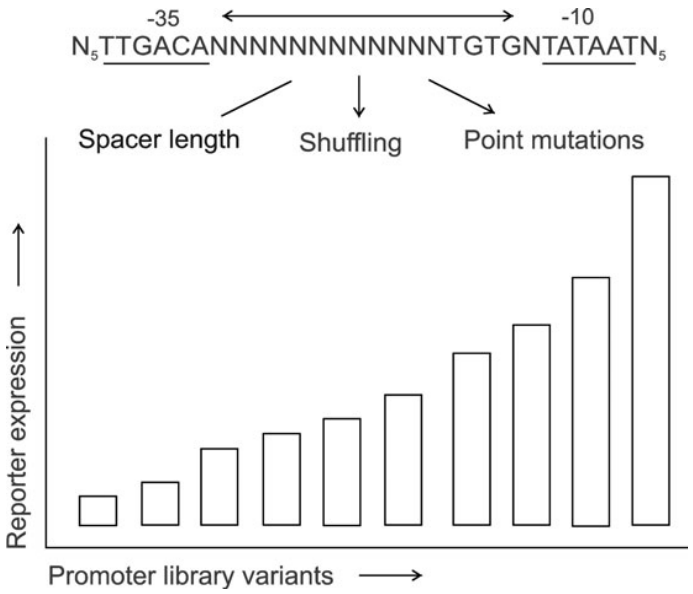



Figure 1.1 Prokaryotic promoter libraries with different promoter strength can be generated by varying the sequence surrounding of the -10 and -35 consensus boxes (shuffling, point mutations, N: A, T, C, G) or the length of the spacer sequence (spacer length), the promoter regulation can be influenced by changes in the consensus boxes

The applicability of synthetic promoters was demonstrated by fine-tuning lycopene production in *E. coli* (isoprenoid pathway).[12] Exchanging the endogenous *dxs* gene promoter with different strong synthetic promoters resulted in an increase in lycopene production, up to a certain level. To link lycopene production almost linearly to *dxs* expression, it was important to fine tune both the single gene and the entire gene pathway. In an other case the metabolic control of the *L. lactis* las operon was studied by applying synthetic promoters to change the expression levels of several individual genes.[6] Furthermore, synthetic promoters were also used to study the *E. coli* glycolytic flux dependence on ATP demand.[18]

#### 1.1.4.3 Model based promoter design for prokaryotes

To generate a generally valid promoter model, the synthetic promoter classification of Jensen,[13] which discriminates between constant core sequences and variable linkers, was systematically reevaluated.[19] Indeed, various promoter strengths were found within all promoter classes, which challenged common wisdom about intact consensus sequences. Therefore, alternative, more complex concepts were investigated. A statistical method that correlated strong and weak expression proved to be a successful model.[19,20] In addition, good results were achieved with a partial least squares statistical method, a potential tool for future *in-silico* promoter design.[19]

#### 1.1.4.4 Artificial regulatory circuits for prokaryotes

To gain a better understanding of cis-regulatory codes for rational promoter design, several studies concentrated on models for bacterial promoter-encoded logic. Experimentally, complex promoter libraries were generated by shuffling *E. coli* transcriptional regulators (TRs), activators, or repressors, to locations within distal (<-35), core (-35/-10), and proximal (>-10) promoter regions. One finding was that complex bacterial promoters can be modelled with simple Boolean logic functions.[21] For example, activator operators had the most influence when positioned distal to the -35 consensus box. In contrast, repressors were functional at all three sites, but had the most effect at the core and the least at distal sites. In addition, one repressor was equivalent to multiple repressors.[22]

#### 1.1.4.5 Minimal Synthetic Promoters for Eukaryotes

In contrast to prokaryotes, transcription initiation in eukaryotes is a more complex process, involving the interplay of several factors at the core promoter. Eukaryotic transcription is also strongly influenced by upstream sequence elements, epigenetic features, and signal transduction.[9] The diversity of core promoters in eukaryotes makes artificial promoter design challenging. Nevertheless, a super core promoter was previously described for metazoan expression that combined 4 core promoter sequence motifs, namely the TATA-box, initiator element (Inr), motif ten element (MTE) and downstream promoter element (DPE).[23]
Most efforts to generate synthetic promoter libraries of different strength for eukaryotes employed upstream activator or repressor binding sites fused to natural core promoters.

#### 1.1.4.6 Synthetic promoters for yeasts

Though there is a lack of uniform consensus sequences an adapted approach of Jensen was also applicable for eukaryotic promoter library generation.[24,25,26] Functional promoters for the yeast *Saccharomyces cerevisiae* can be generated by combining runs of fixed sequences motifs with random bases.[24] Fixed motifs, as for example the TATA box or the transcription start site, here derived from the commonly used *gal1* promoter, ensured promoter function in combination with a TetR (tetracycline repressor) binding site, while random surrounding bases modulated the promoter efficiency.

In the interest of cis-motif discovery for synthetic promoter design and engineering, a library approach was performed for the methanol inducible promoter of the alcohol oxidase 1 gene ($P_{AOX1}$) of the methylotrophic yeast *Pichia pastoris*.[5] $P_{AOX1}$ is tightly repressed in the presence of glucose, glycerol, and other carbon sources. The deletion of *in-silico* predicted TFBS produced differently regulated promoter variants with different potencies (Figure 1.2).

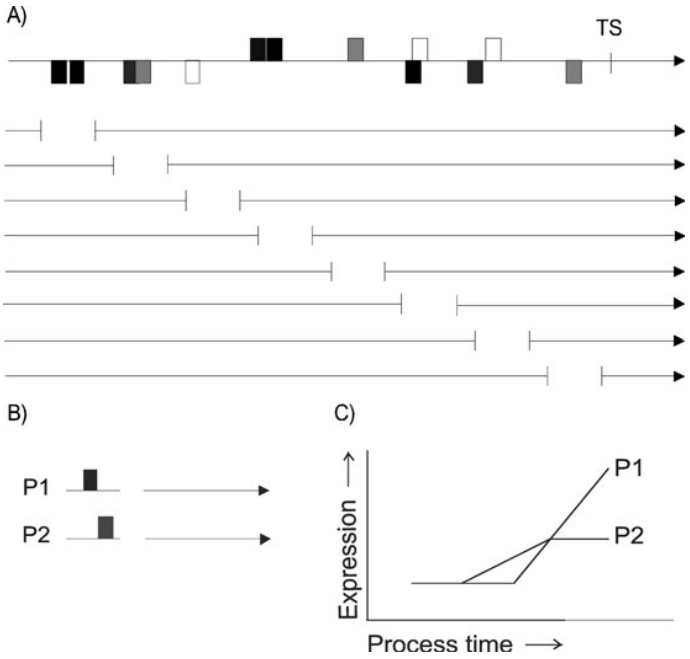



Figure 1.2 Eukaryotic promoter library approaches, A) Promoter library generation and sequence analysis by deletion of putative transcription-factor binding sites (rectangles depict the schematic binding of transcription factors to a promoter sequence, TS: transcription start), B) Synthetic promoter generation by fusion of sequence randomers or identified cis-acting elements to core promoters (synthetic promoters: P1 and P2), C) Different employed regulatory elements (P1 and P2) can alter a promoters mode of regulation and consequently the expression profile

The broad applicability of the generated promoter variants was demonstrated by employing commercial enzymes as reporters. For example, placing the horseradish peroxidase (*hrp*) gene under the control of a strong, derepressed promoter variant caused increased volumetric activity.[5] In contrast, placing the same promoter in control of porcine trypsinogen, an enzyme used for *in vitro* processing of biopharmaceuticals,

caused lower yields, but improved product quality due to a delay, or even prevention, of autoproteolytic product degradation.[27] In addition, employing a novel short synthetic promoter, generated by fusing an identified cis-acting sequence to a core promoter, enabled the production of significant amounts of trypsinogen, even without induction with methanol. Another library variant with increased promoter strength successfully increased the expression of secreted industrial enzymes[28] and improved *P. pastoris* whole cell conversions that required dehydrogenases.[2] In the latter case, this non-natural promoter variant with high reductase expression was combined with an engineered *P. pastoris* strain that had its carbon metabolism redesigned to minimize biomass production and strengthen the existing pathway for NADH regeneration.
This one-for-all promoter strategy uses a well-known, frequently-used, strong promoter sequence to obtain a series of similar sequences with different promoter features. This approach offers new perspectives for innovative expression strategies, including expression cascades or variable transcript production cycles, which can be controlled by variable carbon source feeds. For example, glucose depletion can de-repress the expression of chaperones or foldases prior to target gene expression that is induced by methanol after glucose depletion, and thus improve enzyme titers (Abad and Glieder, manuscript in preparation).

#### 1.1.4.7 Model based promoter design for eukaryotes

There is no simple binary code between the active and inactive states of a promoter in *S. cerevisiae*.[29] However, by employing an operator that binds to a repressor element, it was shown that basal expression (without induction) was increased as the distance between the operator and TATA sequences expanded.[30] Similar to prokaryotic promoter prediction, a thermodynamic model can be used to relate promoter sequences to expression.[31] This model is based on the assumptions that gene expression is regulated by protein-DNA and protein-protein binding, which are associated to changes in free energy. When results from experiments with several promoter libraries were compared to model-generated computational data, 44-59% of the changes in YFP (yellow fluorescent protein) reporter expression were predicted by the model. This model also explained about 60% of the observed changes in expression due to environmental changes (e.g. high glucose concentrations or amino acid starvation).[32] This might become particularly important for future design of reliable fermentation processes or whole cell biotransformations applying high substrate or product concentrations. In addition, this type of model can be used to detect large networks of similar or simultaneously regulated genes,[31] which might provide useful hints for engineering global transcription regulators of industrial production organisms.
Similarly, also a promoter library designed by the adapted approach of Jensen[13] was used to generate a model for gene network prediction.[24] In *S. cerevisiae* these promoters were repressed by TetR and induced by anhydrotetracycline (AtC). Using the generated data a model for an incoherent type II negative feed-forward loop network was generated. The motif, consisting of two repressor genes which exert mutual repression, was then further used to time yeast sedimentation (*flo1* expression).

#### 1.1.4.8 Synthetic promoters for animal cells

A converse and fully synthetic approach is to fuse sequence randomers instead of known TFBS; this generates cis-acting motifs which can be clustered and identified via transcription factor databases like TransFac. Thus, also novel sequence motifs can be discovered.[33]

Synthetic promoters that respond to sonication, triggered state of oxidative stress, were designed for tissue specific expression in cancer cells.[34] With luciferase as a reporter, two promoter library variants from transfected HeLa cells exhibited an increased response to sonication. This promoter library was generated by random ligation of known oxidative stress responsive TFBS to a core promoter fragment. These elements were also used in an earlier approach to design synthetic promoters that responded to X-ray radiation; those promoters were improved with mutagenic PCR.[35]

Cellular eukaryotic promoters like those used for gene therapy often have low transcriptional activity compared to viral promoters. In those cases, the method of two step transcriptional amplification can be used to enhance transcription. This method employs two copies of a cell specific promoter, and the simultaneous expression of a strong transcriptional activator, which stimulates the expression of the target gene. Synthetic bidirectional promoters are used to shorten the size of the expression cassette.[36] Generally, these consist of an activator binding site that contains an effective promoter and, upstream in the opposite direction, a strong viral core promoter fragment, which drives expression of the activator protein (Figure 1.3). This is successful when the upstream elements of the effective promoter can function in both directions. On the other hand, for the expression of a transcription factor, a core promoter with weak transcription might be sufficient or even desired. Similarly, concepts of auto induction, such as the natural prokaryotic nisin inducible (NICE) expression system of *Lactococci*,[37] might be of interest for enhancing expression induced by synthetic promoters.

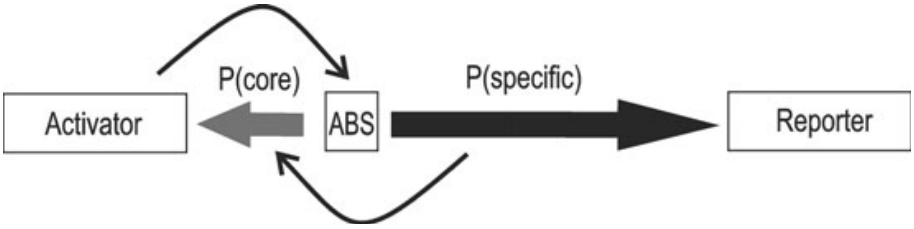



Figure 1.3 Synthetic bidirectional promoter construct, P(specific): full length specific promoter driving the expression of the reporter gene fused to expression boosting activator binding sites (ABS), P(core): short core promoter driving the expression of the activator gene, E(bi): bidirectional activating elements stimulating the activity of P(specific) and P(core)

#### 1.1.4.9 Alternative regulatory systems for fine-tuned gene expression

Regulation mechanisms that influenced gene expression were recently reviewed.[38, 39] In addition to the design of new promoter sequences transcriptional networks were an early and successful focus of synthetic biology.[40, 41] Now also engineering and redesign of more complex networks based on phosphorylation[42], GTPases[43] and RNA interference[44] are included.

Although there are many different approaches we want to highlight a few concepts for regulating transcription, translation and final biological activity in complex pathways by other approaches than (re-) designed promoter sequences.
Nutrient based regulation was utilized when different glucose (repressor) and arabinose (inducer) concentrations were applied to tune the AraC-*ara*BAD promoter-regulated T7 expression system.[45] In another approach, directed evolution of the AraC protein was used to avoid promoter cross-talk between $P_{Bad}$ (a promoter repressed by IPTG) and $P_{lac}$ (a promoter induced by IPTG).[46] Interesting new prospects for the design of new regulatory systems were opened by the discovery that small duplex RNAs could be used as ubiquitous natural tools.[47] Until recently, the common assumption was that these non-coding RNAs (ncRNA) targeted mRNA. However, later findings suggested that they also bind chromosomal DNA and thereby influence gene expression. In addition to the design of cis-acting activator or repressor binding sites and their corresponding (engineered) transcription factors, the design of synthetic ribosome binding sites can be used to control gene expression.[48]
Intriguing natural concepts for the regulation of functional protein complexes are exhibited by cellulosomes. These are fascinating examples that emphasize the importance of fine-tuning structural arrangements.[49] With the use of cohesive, non-catalytic scaffoldins and dockerins, cellulosomes are efficient plant cell wall polysaccharide degrading machines. Adapted from this natural concept, a synthetic protein scaffold complex was applied to improve heterologous mevalonate expression in *E. coli*.[50] Balancing the composition and the orientation of binding domains, co-recruiting the involved enzymes led to a 77-fold increase in expression.

#### 1.1.4.10 Conclusions

Regulating gene expression on the transcriptional level, synthetic promoters and promoter libraries are powerful tools for engineering biocatalysis, biotransformation, and more. Here we summarized synthetic promoter applications in pathway engineering (lycopene production), enzyme production (e.g. *hrp* expression), whole cell conversions (engineered *P. pastoris* strains and promoters), and gene therapy (bidirectional promoters). Synthetic prokaryotic promoters can be designed *de novo* in a very compact manner. Importantly, the composition and length of spacer sequences between consensus blocks, though quite flexible, are strategic tools for tuning promoter strength. The design of fully synthetic promoters for eukaryotes is challenging. Nevertheless, natural promoter elements can be redesigned and engineered to obtain new functionalities. By employing mathematical models, new functions could be predicted with astonishing reliability.
Moreover, initial success stories over the past few years have demonstrated the potential of synthetic promoter technology and alternative artificial regulatory concepts for driving efficient new bioprocesses. Nature has provided many solutions in-between the classic “on or off” regulation. Likewise, future applications in biocatalysis will profit from synthetic biology as smart new concepts that achieve optimal balances in the laboratory become feasible and gain importance. In the near future, innovative expression regulation will deliver higher yields in microbial enzyme production processes enable smart expression cascades, and probably even iterative expression cycles with artificial regulatory circuits. Product- or self-inducible artificial circuits for enzyme expression and stoichiometric, structurally optimized enzyme clusters are feasible. This will impact industrial

biotechnology for single protein production and provide the means to balance enzyme cascades and entire metabolic pathways.

#### 1.1.4.11 References


[1] E. Andrianantoandro, S. Basu, D. K. Karig, R. Weiss, Mol. Syst. Biol. 2006, 2, 2006.0028.

[2] B. Janesch, A. Glieder, F. Hartner, K. Schroer, K. Luef, B. Pscheidt, 2009, WO2009070822.

[3] J. H. Lee, B. H. Sung, M. S. Kim, F. R. Blattner, B. H. Yoon, J. H. Kim, S. C. Kim, Microb. Cell Fact. 2009, 8, 2.

[4] G. Posfai, G. Plunkett 3rd, T. Feher, D. Frisch, G. M. Keil, K. Umenhoffer, V. Kolisnychenko, B. Stahl, S. S. Sharma, M. de Arruda, V. Burland, S. W. Harcum, F. R. Blattner, Science 2006, 312, 1044-1046.

[5] F. S. Hartner, C. Ruth, D. Langenegger, S. N. Johnson, P. Hyka, G. P. Lin- Cereghino, J. Lin-Cereghino, K. Kovar, J. M. Cregg, A. Glieder, Nucleic Acids Res. 2008, 36, e76

[6] B. Koebmann, C. Solem, P. R. Jensen, FEBS J. 2005, 272, 2292-2303.

[7] K. Hammer, I. Mijakovic, P. R. Jensen, Trends Biotechnol. 2006, 24, 53-55.

[8] I. Mijakovic, D. Petranovic, P. R. Jensen, Curr. Opin. Biotechnol. 2005, 16, 329-335.

[9] N. D. Heintzman, B. Ren, Cell. Mol. Life Sci. 2007, 64, 386-400.

[10] T. Juven-Gershon, J. T. Kadonaga, Dev. Biol. 2010; DOI: 10.1016/j.ydbio.2009.08.009.

[11] C. Solem, P. R. Jensen, Appl. Environ. Microbiol. 2002, 68, 2397-2403.

[12] H. Alper, C. Fischer, E. Nevoigt, G. Stephanopoulos, Proc. Natl. Acad. Sci.USA 2005, 102, 12678-12683.

[13] P. R. Jensen, K. Hammer, Appl. Environ. Microbiol. 1998, 64, 82-87.

[14] S. Braatsch, S. Helmark, H. Kranz, B. Koebmann, P. R. Jensen, BioTechniques2008, 45, 335-337.

[15] T. Burr, J. Mitchell, A. Kolb, S. Minchin, S. Busby, Nucleic Acids Res. 2000,28, 1864-1870.

[16] I. Rud, P. R. Jensen, K. Naterstad, L. Axelsson, Microbiology 2006, 152,1011-1019.

[17] G. Miksch, F. Bettenworth, K. Friehs, E. Flaschel, A. Saalbach, T. Twellmann,T. W. Nattkemper, J. Biotechnol. 2005, 120, 25-37.

[18] B. J. Koebmann, H. V. Westerhoff, J. L. Snoep, D. Nilsson, P. R. Jensen, J.Bacteriol. 2002, 184, 3909-3916.

[19] M. De Mey, J. Maertens, G. J. Lequeux, W. K. Soetaert, E. J. Vandamme,BMC Biotechnol. 2007, 7, 34.

[20] K. Jensen, H. Alper, C. Fischer, G. Stephanopoulos, Appl. Environ. Microbiol.2006, 72, 3696-3701.

[21] A. Kinkhabwala, C. C. Guet, PLoS 1 2008, 3, e2030.

[22] R. S. Cox 3rd, M. G. Surette, M. B. Elowitz, Mol. Syst. Biol. 2010; DOI:10.1038/msb4100187

[23] T. Juven-Gershon, S. Cheng, J. T. Kadonaga, Nat. Methods 2006, 3, 917-922.

[24] T. Ellis, X. Wang, J. J. Collins, Nat. Biotechnol. 2009, 27, 465-471.

[25] M. Jeppsson, B. Johansson, P. R. Jensen, B. Hahn-Hagerdal, M. F. Gorwa-Grauslund, Yeast 2003, 20, 1263-1272.

[26] J. Tornøe, P. Kusk, T. E. Johansen, P. R. Jensen, Gene 2002, 297, 21-32.

[27] C. Ruth, T. Zuellig, A. Mellitzer, R. Weis, V. Looser, K. Kovar, A. Glieder, unpublishedresults.

[28] B. Pscheidt, Z. Liu, R. Gaisberger, M. Avi, W. Skranc, K. Gruber, H. Griengl,A. Glieder, Adv. Synth. Catal.

2008, 350, 1943-1948.
[29] M. Ligr, R. Siddharthan, F. R. Cross, E. D. Siggia, Genetics 2006, 172,2113-2122.
[30] K. F. Murphy, G. Balazsi, J. J. Collins, Proc. Natl. Acad. Sci. USA 2007, 104,12726-12731.
[31] J. Gertz, E. D. Siggia, B. A. Cohen, Nature 2009, 457, 215-218.
[32] J. Gertz, B. A. Cohen, Mol. Syst. Biol. 2009, 5, 244.
[33] G. M. Edelman, R. Meech, G. C. Owens, F. S. J 1s, Proc. Natl. Acad. Sci.USA 2000, 97, 3038-3043.
[34] A. Watanabe, S. Kakutani, R. Ogawa, S. I. Lee, T. Yoshida, A. Morii, G.Kagiya, L. B. Feril Jr. , H. Fuse, T. Kondo, J. Med. Ultrasonics 2009, 36, 9-17.
[35] R. Ogawa, S.-i. Lee, G. Kagiya, H. Hirano, S. Fukuda, T. Kondo, T. Kodaki,J. Gene Med. 2008, 10, 316-324.
[36] B. Liu, J. F. Paton, S. Kasparov, BMC Biotechnol. 2008, 8, 49.
[37] O. P. Kuipers, M. M. Beerthuyzen, P. G. de Ruyter, E. J. Luesink, W. M.de Vos, J. Biol. Chem. 1995, 270, 27299-27304.
[38] J. M. Carothers, J. A. Goler, J. D. Keasling, Curr. Opin. Biotechnol. 2009,20, 498-503.
[39] S. Mukherji, A. van Oudenaarden, Nat. Rev. Genet. 2009, 10, 859-871.
[40] C. A. Voigt, Curr. Opin. Biotechnol. 2006, 17, 548-557.
[41] M. Kaern, W. J. Blake, J. J. Collins, Annu. Rev. Biomed. Eng. 2003, 5, 179-206.
[42] S. H. Park, A. Zarrinpar, W. A. Lim, Science 2003, 299, 1061-1064.
[43] B. J. Yeh, R. J. Rutigliano, A. Deb, D. Bar-Sagi, W. A. Lim, Nature 2007,447, 596-600.
[44] K. Rinaudo, L. Bleris, R. Maddamsetti, S. Subramanian, R. Weiss, Y. Benenson,Nat. Biotechnol. 2007, 25, 795-801.
[45] D. R. Wycuff, K. S. Matthews, Anal. Biochem. 2000, 277, 67-73.
[46] S. K. Lee, H. H. Chou, B. F. Pfleger, J. D. Newman, Y. Yoshikuni, J. D. Keasling,Appl. Environ. Microbiol. 2007, 73, 5711-5715.
[47] S. T. Younger, D. R. Corey, ChemBioChem 2009, 10, 1135-1139.
[48] H. M. Salis, E. A. Mirsky, C. A. Voigt, Nat. Biotechnol. 2009, 27, 946-950.
[49] E. A. Bayer, J. P. Belaich, Y. Shoham, R. Lamed, Annu. Rev. Microbiol.2004, 58, 521-554.
[50] J. E. Dueber, G. C. Wu, G. R. Malmirchegini, T. S. Moon, C. J. Petzold, A. V.Ullal, K. L. Prather, J. D. Keasling, Nat. Biotechnol. 2009, 27, 753-759.

## 1.2 Materials and methods

### 1.2.1 Plasmids

The expression cassette of the *Escherichia coli*/*Pichia pastoris* shuttle vector pPpT4 (Figure 1.4) was based on a synthetic *AOX1* promoter and terminator sequence synthesized by GenScript (Piscataway, NJ, USA) based on the sequence information from GenBank (accession no. U96967). To allow selection against the antibiotic zeocin the Ble gene from *S. hindusdanus* (GenBank accession no. A31898), which was optimized for

expression in *E. coli* and *P. pastoris* (Leto Software, Entelechon Corp.), was used. The zeocin resistance gene was under the control of a 3' 34 bp truncated 552 bp long *ILV5* promoter and *AOD* terminator (466 bp), both derived from the *P. pastoris*, strain CBS7435. For bacterial replication the pUC origin (pBR322) was used. For expression of the resistance marker in *E. coli* a synthetic prokaryotic consensus promoter ($P_{EM72}$) was designed and embedded between the 3' truncated eukaryotic promoter and the start of the resistance gene. pPpT2 was constructed similar to pPpT4. However, pPpT4 is characterized by two sequent point mutations in $P_{EM72*}$ ($AC_{-51-52}TT$). These point mutations are assumed to originate from the overlap extension PCR primers, which were used for the template free amplification of $P_{EM72}$ (MPEM71fwOE 5'-CTCTTCCAAATATCGTCTCCACAAATCTAGAGTGTTGACACTTTATACTTCCGG CTCGT ATAATACGACAAGGTG-3', MPEM72rvOE 5'-GGACTGGAACAGCAGAGGTGAGTTTAGCCATGGTTTAGTCCTCCTTACACCTTGTCGT ATTATACGAGCCGGAAG-3').

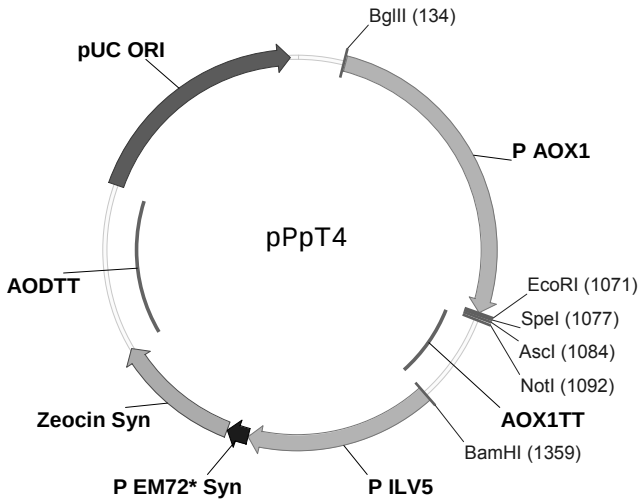



Figure 1.4 *Pichia pastoris* shuttle vector pPpT4 (3546 bp), P AOX1: *AOX1* promoter, AOX1TT: *AOX1* terminator, MCS (multiple cloning site): *Eco*RI, *Spe*I, *Asc*I, *Not*I, P ILV5: promoter of the *P. pastoris ILV5* gene, P EM72*: synthetic *E. coli* promoter, Zeocin Syn: synthetic codon optimized Zeocin gene, AODTT: terminator of the *P. pastoris AOD* gene, pUC ori: origin of replication

### 1.2.2 Strains

Subcloning was done using the *E. coli* strain DH5a-$T1^R$ (Invitrogen, Carlsbad, CA, USA). All *P. pastoris* experiments were performed using strain CBS7435-$Mut^S$ (Δ*AOX1*, see also 3.2.1). The *P. pastoris* strain CBS7435 was obtained from CBS fungal biodiversity center.

### 1.2.3 Site directed mutagenesis and overlap extension PCR

QuickChange 2-step site directed mutagenesis (SDM) was performed according to Wang and Malcolm (1999). Mutations larger then 50 bp were made by two rounds of SDM (e. g. d6*_2xMat_2x201). Mutations were verified by sequencing subsequent to transformation of *E. coli*. For the attachment of the cis-acting elements to core promoters long high quality primers were used. SDM and PCR reactions were performed using PfuUltra™ polymerase (Stratagene Inc.).

### 1.2.4 Promoter variants

Constructs were named according to the deletions, insertions and fusions made. Sequence numeration was done in regard to the natural *AOX1* promoter in upstream direction (GAAACG$_{-1}$ATG). According to Hartner et al. (2008) an *Eco*RI site was introduced at the 3' end of the *AOX1* promoter (GAAAGA$_{-1}$ATTCATG).

A detailed description about already existing *AOX1* promoter variants can be found in Hartner and Glieder (2005) or Hartner et al. (2008). Briefly, while d6 refers to a deletion of 30 bases beginning at position -223, (d6)* refers to an additional deletion of 2 base pairs, located 3' of d6 ($\Delta TA_{-208-209}$).

#### 1.2.4.1 New generation promoter variants I

The multiplied positive acting elements Mat1-Mc and 201-214 are located between -253/-270 and -189/-202. Variant 2x201* was made by deletion of region star (ΔTA) and duplication of region 201-214. The variants 2x201**, d6*_2xMat_2x201* and d6*_2xMat_2x201** are characterized by additional mutations or deletions, $\Delta A_{-201}$ (in 2x201), $\Delta A_{-192}$ (in 2x201) and C-$_{181}$A /A-$_{198}$T (in 1x201), respectively.

#### 1.2.4.2 New generation promoter variants II

To attach cis-acting elements to the core promoters AOX176, AOX201 and AOX201-Z primers binding to the *AOX1* promoter regions -168 to -146 (AOX176 binding), -176 to -202 (AOX201 binding) and -203 to -220 (ZUS binding) were used. Using the long high quality primers shown in Table 1.2 (Annex) following cis-acting elements were fused one time, as repeat or in combination with other elements to the core promoters: d6 (-223/-252), Mat1-Mc (-253/-271), Rapv (-272/-294), InD-d4m (-379/-394), Adr1 (-559/-587), Rap1 (-596/-620) and Stre (-654/-672). In addition, some of the elements were 3' extended with 5-16 bp of the natural sequence to favour binding of the respective transcription factor. An overview of the generated promoter variants and used abbreviations is given in Table 1.1.

Table 1.1 Summary of the new generation promoter variants II, Z: linker region ZUS

| promoter name | core promoter | positive acting elements (abbreviation) |
|---|---|---|
| AOX176-MM | AOX176 | 2xMat1-Mc (MM) |
| AOX176-RR | AOX176 | 2xRap1 (RR) |
| AOX176-R | AOX176 | Rap1 (R) |
| AOX176-MZAA | AOX176 | Mat1-Mc, 2xAdr1 (M, AA) |
| AOX176-MZSR | AOX176 | Mat1-Mc, Stre, Rap1 (M, SR) |
| AOX176-MZMM | AOX176 | 3xMat1-Mc (M, MM) |
| AOX176-RZ | AOX176 | Rap1 (R) |
| AOX176-MZ | AOX176 | Mat1-Mc (M) |
| AOX176-Z | AOX176 | x |
| AOX201-Z | AOX201 | x |
| AOX201-201Z | AOX2x201-Z | 201-214 (201) |
| AOX201-ZRvM | AOX201-Z | Rapv, Mat1-Mc (RvM) |
| AOX201-Zd62x | AOX201-Z | 2xd6 |
| AOX201-ZSR | AOX201-Z | Stre, Rap1 (SR) |
| AOX201-ZS | AOX201-Z | Stre (S) |
| AOX201-ZAA | AOX201-Z | 2xAdr1 (AA) |
| AOX201-Zd6Gcr | AOX201-Z | d6(Gcr1) |
| AOX201-Zd4m | AOX201-Z | d4m |
| AOX201-ZRR | AOX201-Z | 2xRap1 (RR) |
| AOX201-ZR | AOX201-Z | Rap1 (R) |
| AOX201-ZRM | AOX201-Z | Rap1, Mat1-Mc (RM) |
| AOX201-ZMM | AOX201-Z | 2xMat1-Mc (MM) |
| AOX201-ZRMd6 | AOX201-Z | Rap1, Mat1-Mc, d6 (RMd6) |
| AOX201-ZRsp | AOX201-Z | Rsp |

#### 1.2.4.3 New generation promoter variants III

The nucleotide sequences of core promoter 1 and 11 are shown in Figure 1.5 and 1.6. To attach cis-acting elements region ZUS (-203/-220) was used. For the generation of promoter variant core11-ZMM the cis-acting element Mat1-Mc (-253/-271) was attached twice.

>core 1 (182 bp)
agatcttgagcacgaccaacatatctatataaaacaaaagcacctttcctcttttttctccttttttttcatcatcactttcacctcaatttccattacttctggtttcttctca
caagataaatattaatacaacttaagacgcatacaatcttataacaactaaaaacaatgtcaaacttcaaaga$_{(-1)}$attc

Figure 1.5 Synthetic core promoter 1 (consensus based design), underlined: restriction sites *Eco*RI and *Bgl*II

>core 11 (168 bp)
agatcttgagcacgaccaacacatctatataatacaaaagcacctttcctctttttctccttttcttttcgatcataatctaatcagtctcaagaaaccagaagtaatg
gaaattgaggtgaaagtcacaacttaagaccatacaatcttactagatatatcaaacttcaaaga$_{(-1)}$attc

Figure 1.6 Synthetic core promoter 11 (TF based design), underlined: restriction sites *Eco*RI and *Bgl*II

### 1.2.5 Transformation, screening and fluorescence detection

The condensed protocol of Lin-Cereghino et al. (2005) was used for *Pichia pastoris* transformations. *P. pastoris* was transformed with 1 µg of linearized plasmid DNA applying following parameters: 1,5 kV, 25 µF and 200 Ω. For regeneration 1 ml of cold sorbitol was added. Transformed cells were regenerated for 2 hours at 28 °C and plated on selective media.
Screening was done in 96 well plates (microplates). According to Weis et al. (2004) clones were grown 60 hours in 300 µl BMD1% media. Induction was performed by addition of 250 µl BMM2 (0 h induction time) followed by methanol pulses of 50 µl BMM10 after 12, 24 and 48 hours of induction. To follow the production of green fluorescent protein (GFP) samples were taken before each methanol pulse. Intracellular GFP fluorescence was measured using a SPECTRAmax GeminiXS Spectrofluorometer (Molecular Devices, Inc., US). Following parameters were applied: extinction 395 nm, emission 507 nm. Prior measurement samples were 1:4 diluted with $H_2O$. A *P. pastoris* negative control strain, containing linearized pPpT4 without GFP, was in addition evaluated, but similar to the Mut$^S$ wild type strain no GFP fluorescence was detected (background RFU: 0 h: 10, 48 h: 18).

## 1.3 Experiments and results

### 1.3.1 New generation promoter variants I

Based on the studies of Hartner et al. (2008) new deletion and insertion *AOX1* promoter variants were generated incorporating the gained knowledge about possible regulatory sites. In the referred study, performed in 96 well plates and using GFP as reporter, deletion of the regions d6 and d6* resulted in a 4-fold increase in activity before induction compared to the wild type promoter (WT) at single copy level. Differently, deletion of region Mat1-Mc resulted in a decreased methanol induced activity of 42%, indicating this region important for methanol induced regulation in *Pichia pastoris*. In agreement with this Kranthi et al. (2009) identified the key regulator Mxr1p binding to exactly this region. Fusion of region Mat1-Mc to an *AOX1* core promoter fragment (AOX176) resulted in enhanced basal activity in absence of glucose and presence of methanol (Hartner et al., 2008). Similar effects were also found employing the fusion promoter variant AOX176-201-214.
For these reasons the elements d6, d6*, Mat1-Mc and 201-214 were chosen for further studies. Using multiplied activator sites and combined duplication/deletion strategies the following new promoter variants were generated by SDM: d6_2xMat_2x201, d6*_2xMat_2x201, d6_2x201, d6*_2x201, 2x201* and d6_3xMat. By mistake also the variants 2x201**, d6*_2xMat_2x201* and d6*_2xMat_2x201** were generated and included in further

studies. Detailed information on the deletions and insertions made within $P_{AOX1}$ can be found under 1.2.4.1. The generated $P_{AOX1}$ promoter variants are depicted in Figure 1.7.
*Pichia pastoris* was transformed with 9 linearized plasmids coding for GFP expression und the control of the different promoter variants (d6_3xMat, d6_2x201, d6*2x201, d6_2xMat_2x201, d6*_2xMat_2x201, d6*_2xMat_2x201*, d6*_2xMat_2x201**, *2x201 and **2x201). As a control *P. pastoris* was in addition transformed with one $P_{AOX1}$ and one d6* driven reference construct.
To simplify the comparison of the different promoters in screening the average values of all active clones of each construct, including clones with different copy numbers, were calculated and normalized to the corresponding average WT value. This method of analysis is similar to the comparison of putative single copy clones estimated from expression landscapes; however, is more accurate if the copy number steps from the expression landscapes are indistinct. In addition, as the best promoter-gene combination is identified independent from copy numbers, the method is not protein biased (e.g. toxic proteins might give best results with only 1 gene copy integrated).
Here, 46 clones were screened per construct in microplates and the average activity, mean relative fluorescence of all active clones, was evaluated in regard to the WT average after 0 and 48 hours of induction. 0 hours of induction corresponded to 60 hours of microplate batch growth with putative glucose depletion after 24 hours (derepression conditions). At 0 hours of methanol induction 100% WT average was equivalent to 27 relative fluorescence units (RFU), at 48 hours to 1870.


Figure 1.7 $P_{AOX1}$ deletions and insertion variants, WT: wild type *AOX1* promoter and corresponding numeration, blue: Mat1-Mc region, green: region between d6 and element 201-214 (including deletion *: ΔTA), red: element 201-214 and point mutations, brown: transcription start site (TSS/$A_{-113}$) and the putative TATA box (MatInspector prediction)

After 60 hours of microscale batch growth and glucose depletion promoter d6* showed the predicted 4-fold increase in activity. In addition, several other variants with a 10-fold and higher increase in activity were found (Figure 1.8). The constructs d6_2xMat_2x201 and d6*_2xMat_2x201* reached values of 1370% and 1450%, respectively. In addition, also variant d6_3xMat showed a significant enhanced level of 1315%. Seemingly deletion of region d6 in combination with Mat1-Mc duplication triggered putative derepression induction. Also duplication of the region 201-214 facilitated methanol free expression. Comparing the constructs d6_2x201 and d6*_2x201 deletion of region star increased methanol free activity from 450% to 590%. Surprisingly, the found low methanol free activity of construct d6*_2xMat_2x201 (200% in comparison to $P_{AOX1}$ not induced). In contrast, all other constructs based on d6 deletion and Mat1-Mc/201-214 duplication showed at this time point significant enhanced values of 10-fold WT and higher.

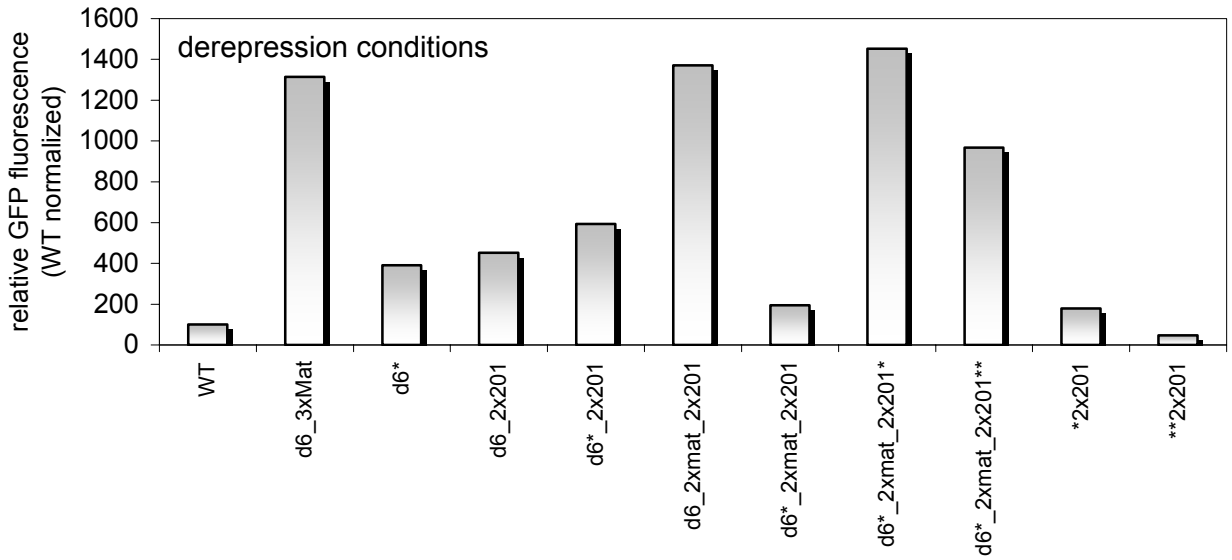



Figure 1.8 GFP expression after 60 h of batch growth and glucose depletion (derepression conditions, 0 h) employing next generation $P_{AOX1}$ deletion and insertion variants (microplate screening), mean value of 46 clones

After 48 hours of methanol induction employment of the reference construct d6* resulted in 36% residual activity. For comparison, Hartner et al. (2008) reported 66% at this time point. However, both studies indicate the region upstream activating. Similar to the methanol free system best results were obtained by duplication of the elements Mat1-Mc and 201-214 in combination with deletion of region d6 reaching a maximal value of 195% (d6_2xMat_2x201, Figure 1.9). Also construct d6_3xMat performed well with 174% relative activity. Comparing the constructs d6_2x201 and d6*_2x201 deletion of region star decreased methanol induced activity from 170% to 59%. This effect was in addition observed comparing the constructs d6_2xMat_2x201 and d6*_2xMat_2x201, for which activity decreased from 195% to 113%. Employment of the variant d6*_2xMat_2x201* resulted in a partly revived activity of 144%, while promoter d6*_2xMat_2x201** showed an even lower residual activity of 97%. Application of the promoter variant 2x201*, characterized by star element

deletion and duplication of region 201-214, resulted in a significant diminished activity of 25%. Similar, employment of variant 2x201** resulted in 23% residual activity only. As according to Hartner et al. (2008) single duplication of region 201-214 resulted in 139% relative wild type activity, region star is seemingly connected to high methanol induced expression and if deleted drastically decreases activity.

.

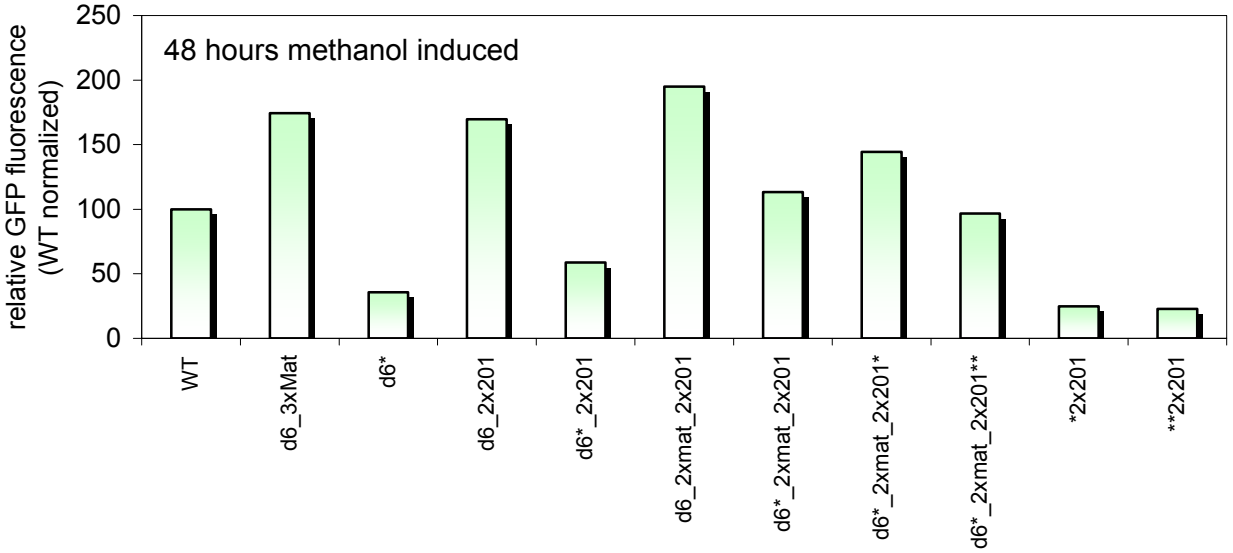



Figure 1.9 Methanol induced GFP expression (48 h) employing next generation $P_{AOX1}$ deletion and insertion variants (microplate screening), mean value of 46 clones *

*Raw data from different methanol induction times of for rescreening selected clones based on the promoters d6_3xMat and d6_2xMat_2x201 can be found in the Annex.

In rescreening 6 clones of each construct representing the whole activity range were inoculated 3 times and the activity was evaluated after 0 and 24 hours of methanol induction (48 hours not determined). Assuming that expression correlates to copy number clones were classified as multi and low copy clones according to their GFP expression levels. Though Real Time PCR is needed for verification, clones with the lowest activities were classified as putative single copy (sc) clones.

The results of the putative single copy clones from rescreening were evaluated in comparison to the average single copy level found in screening. From screening the single copy level was defined as the lowest level of expression with the highest condensation of clones. All results were normalized to the screening WT sc clone average (0 h: 21 RFU, 24 h: 760 RFU).

Results obtained under derepression conditions didn't correlate well with the average single copy screening values. With the exception of construct d6_3xMat, for which the values of screening and rescreening matched, the obtained values were significantly lower. While application of promoter d6_3xMat resulted in a 6-fold increase after batch growth on glucose, employment of construct d6_2xMat_2x201 resulted in a 3,5-fold increase only. For comparison in screening a 10-fold and higher increase was observed for this construct. Construct d6*_2xMat_2x201** performed well and reached 700%. Differently, no response upon derepression

was found for the constructs *2x201 and **2x201. In contrast to screening, the best results were obtained for construct d6*_2xMat_2x201 showing 10-fold enhanced activity before induction. As construct d6*_2xMat_2x201 has proven its ability in methanol free expression of different industrial reporters (VTU, personal communication) one can only speculate that an error of measurement was responsible for reduced values in screening.

Methanol induced rescreening results correlated well with the data obtained from screening (Figure 1.10). Application of promoter construct d6x_3xMat resulted in 168% activity matching the average screening value of 161%. With 169% a similar level was reached by construct d6_2x201, followed by construct d6_2xMat_2x201 reaching 144%. In comparison the corresponding average single copy screening values were a little bit higher with 197% and 207%. Discrepancies might be the result of the average sc estimation possibly also incorporating clones with a copy number higher than one. Two speculations were confirmed by rescreening. First, star region deletion decreases methanol induced activity significantly (30-60%). Second, methanol induced activity can be increased by deletion of region d6 and duplication of Mat1-Mc/201-214 reaching a maximum of 169% in rescreening.

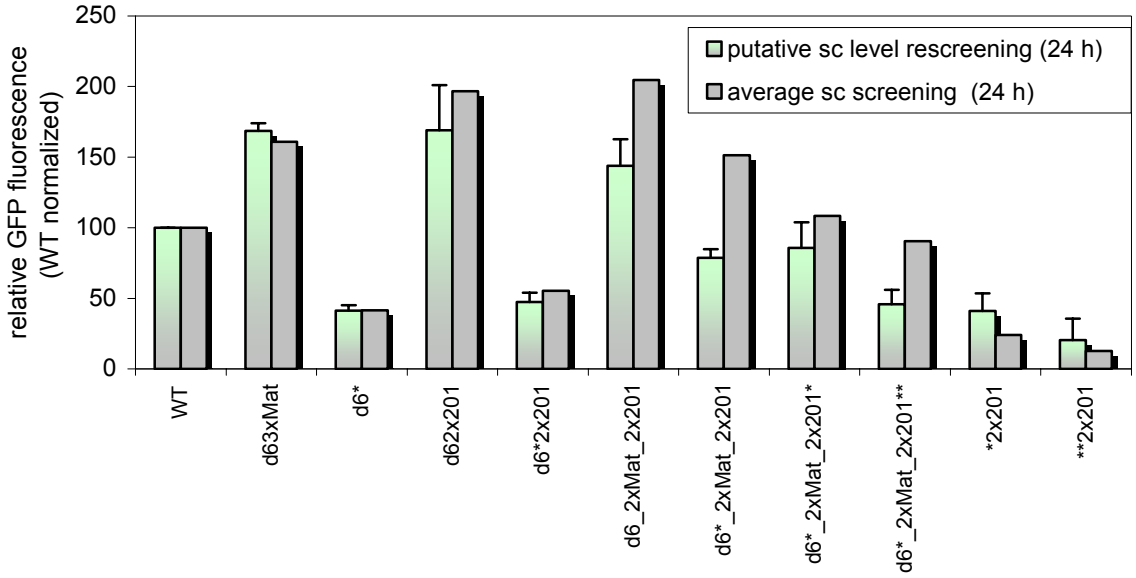



Figure 1.10 Rescreening results of the putative single copy (sc) clones after 24 hours of methanol induction (green) in comparison to the average single copy level obtained in screening after 24 hours of induction (grey)

#### 1.3.1.1 Conclusions

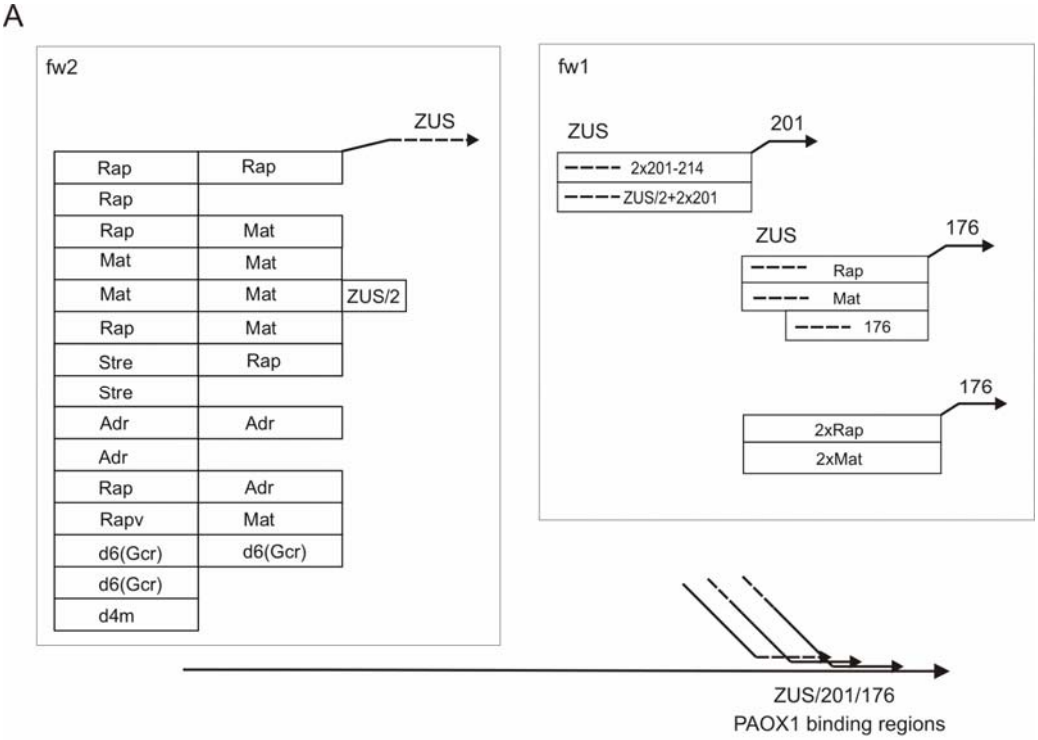
Construct d6_3xMat and d6*_2xMat_2x201 are seemingly interesting for both, high methanol induced and depression based protein expression in *P. pastoris*. Possibly also construct d6_2xMat_2x201 can be used similarly, but the results for this construct have still to be confirmed. Reaching 169% methanol induced WT activity in rescreening, also construct d6_2x201 seems to be interesting for high methanol induced expression. While star element deletion decreased methanol induced activity significantly, a possible positive influence on

derepression induced activity for some of the constructs has still to be elucidated (e.g. d6*_2x201). To create even better promoters, future promoter studies might additionally integrate region d1, identified by Hartner et al. (2008) to be upstream repressing.
On the other hand new studies about relevant transcription factors gave additional information about the mode of regulation of $P_{AOX1}$ and other methanol inducible promoters and thereby offer alternative approaches for the design of efficient new expression systems. For example Kranthi et al. (2009) identified 6 Mxr1p binding sites within $P_{AOX1}$, of which only the one binding to Mat1-Mc was studied here. Yet, in the deletion studies of Hartner et al. (2008) 5 of the 6 Mxr1p binding sites were included. Differently, a recent patent of Tsutsumi et al. (2008) showed naturally methanol inducible promoters independent from methanol by exchanging the promoters of key transcription factors (e.g. Prm1p, Mxr1p). This novel approach offers an alternative access to promoters with different regulatory features.

### 1.3.2 New generation promoter variants II

As a first step in the generation of short synthetic promoters Hartner et al. (2008) fused the positive acting elements Adr1, Stre, Mat1-Mc, Gcr1, 201-214 and 737 to an AOX176 core (basal) promoter fragment. Showing the applicability of the approach, the promoter variants AOX176-201-214, AOX176-Stre and AOX176-Mat1-Mc enhanced expression in absence of glucose and presence of methanol in microscale screening studies using GFP as a reporter and comparing to the basal promoter. Also interesting, while variant AOX176-737 showed even lower activity under derepression conditions, the variants AOX176-Adr1 and AOX176-Gcr1 responded to methanol induction only.
Based on these first results several new short synthetic promoter variants were generated by attaching different cis-acting elements and combinations of said to core promoter fragments. Two core promoter fragments were used in this study (AOX176 and AOX201).
To attach different positive acting elements to core promoter AOX201, region ZUS (Z), located 5' of element 201-214, was used as linker (primer binding region, fw2). In parallel primers (fw1) were designed to attach cis-acting elements to the core promoters AOX176 and AOX201 directly. These primers (fw1) were designed to have a 5' Z binding region. Thus, in a second round of promoter generation each fw2 primer containing a 3' Z region can again be applied, now to attach additional elements to either 5'-Z-element-AOX176-3' or 5'-Z-element-AOX201-3'. An overview of the design strategy and the applied positive acting elements is given in Figure 1.11.

A




B

```
agatctaacatccaaagacgaaaggttgaatgaaacctttttgccatccgacatccacaggtccattc

tcacacataagtgccaaacgcaacaggaggggatacactagcagcagaccgttgcaaacgcaggacct

ccactcctcttctcctcaacacccacttttgccatcgaaaaaccagcccagttattgggcttgattgg

agctcgctcattccaattccttctattaggctactaacaccatgactttattagcctgtctatcctgg

cccCCCTggcgaggttcatgtttgtttatttccgaatgcaacaagctccgcattacACCCgaacatca
   Stre                                                    Rap1
ctccagatgagggctttctgagtgtGGGGtcaaatagtttcatgttccccaaatggcccaaaactgac
                        Adr1
agtttaaacgctgtcttggaacctaatatgacaaaagcgtgatctcatccaagatgaactaagtttgg

ttcgttgaaatgctaacggccagttggtcaaaaagaaacttccaaaagtcggcataccgtttgtcttg

tttggtattgattgacgaatgctcaaaaataatctcattaatgcttagcgcagtctctctatcgcttc
    InD-d4m
tgaaccccggtgcacctgtgccgaaacgcaaatggggaaacACCCgcttttggatgattatgcaTTG
                                        Rapv                Mat1-Mc
TctccacattgtatgCTTCcaagattctggtgggaatactgctgatagcctaacgttcatgatcaaaa
                    d6                  ZUS binding          201
tttaactgttctaaccctacttgacagcaatatataaacagaaggaagctgccctgtcttaaaccttt
  binding                    176 binding
ttttttatcatcattattagcttactttcataattgcgactggttccaattgacaagcttttgattt
        TSS
Taacgacttttaacgacaacttgagaagatcaaaaaacaactaattattgaaaga(-1)attc
```

Figure 1.11 A) Scheme for the generation of the short synthetic promoter variants, squares: cis-acting elements, arrows: primer, applied primer binding regions: ZUS (dotted line), 201 (AOX201), 176 (AOX176), B) $P_{AOX1}$ cis-acting elements (underlined) used for synthetic promoter generation, grey and dark grey: primer binding regions ZUS (fw2), 201 (fw1) and 176 (fw1), bold capital letters: TF binding core similarities (MatInspector, Genomatix, Hartner et al., 2008), red: TSS located at position -113 and the putative TATA box (MatInspector) located between -155/-158 relative to the *AOX1* start codon

Following short promoters were generated by PCR: AOX201-ZRsp, AOX201-ZRMd6, AOX201-ZMM, AOX201-ZRM, AOX201-ZR, AOX201-ZRR, AOX201-Zd4m, AOX201-Zd6Gcr, AOX201-ZAA, AOX201-ZS, AOX201-ZSR, AOX201-Zd62x, AOX201-ZRvM, AOX201-Z, AOX201-201Z, AOX176-Z, AOX176-R, AOX176-RR, AOX176-MZ, AOX176-RZ, AOX176-MM, AOX176-MZMM, AOX176-MZSR and AOX176-MZAA.
A minimum of 50 *P. pastoris* clones was screened per construct and the average activity was evaluated in regard to the WT average after 0 and 48 hours of methanol induction.
After 60 hours of cultivation and glucose depletion several synthetic promoter variants with a putative response on derepression were found (Figure 1.12). Especially clones based on the cis-acting element Rap1 performed well with AOX176-RR reaching 206%. In comparison variant AOX176-R, containing only one Rap1 binding site, reached 148%. This confirms the assumption that repeats of TF binding sites enhance activity in addition to single element fusions. Confirming element Mat1-Mc responsive on derepression, application of variant AOX176-MM resulted in 157% relative activity. Also construct AOX176-RZ reached a value of 152%. Differently, fusion of element Rap1 to core promoter AOX201 (AOX201-ZR) led to a significantly lower residual activity of 66%. Seemingly application of core promoter AOX201 decreased methanol free activity. This effect was confirmed by comparison of the constructs AOX176-RR and AOX201-ZRR, for which application of core promoter AOX201 decreased activity from 206% and 102%. Comparing to Hartner et al. (2008), no significant influence was found under derepression conditions by attaching region Z to AOX176 basal.

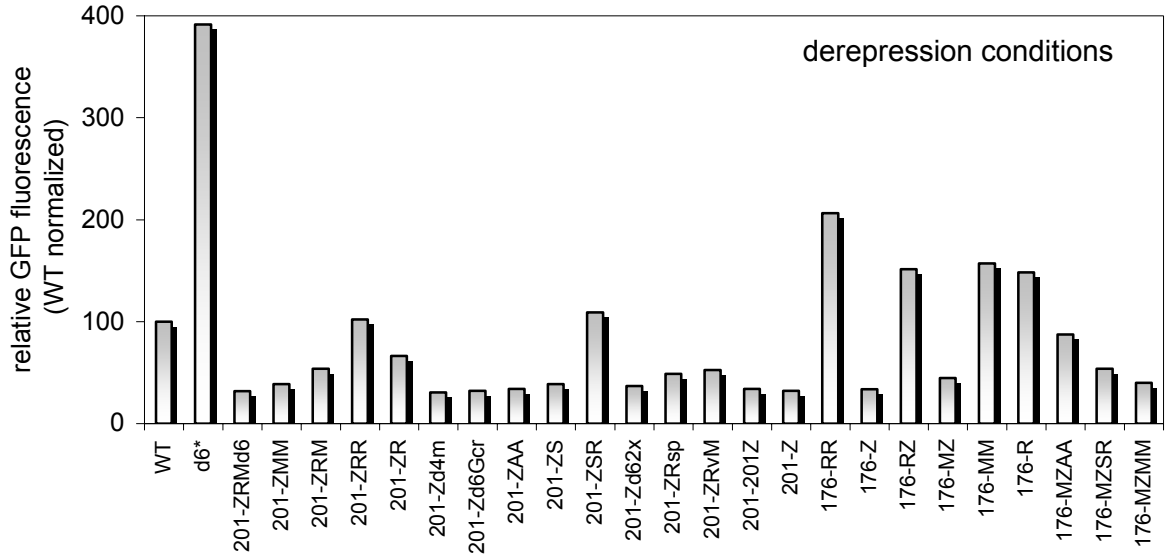



Figure 1.12 GFP expression after 60 h of batch growth and glucose depletion (derepression conditions, 0 h) employing the new short synthetic promoter variants (microplate screening), mean value of minimum 50 clones

After 48 hours of induction the constructs AOX176-MZAA and AOX201-ZMM performed best reaching 42% and 37% induced WT activity (Figure 1.13). Also the promoters AOX201-ZAA and AOX201-ZSR reached methanol induced levels of 27% and 22%. In contrast, no response upon methanol induction (<3%) was found for the constructs AOX201-ZRmd6, AOX201-Zd4m, AOX201-Zd6Gcr and AOX176-Z basal. Interestingly, fusions

containing region d6 (e.g. AOX201-ZRMd6, AOX201-Zd62x and AOX201-Zd6Gcr) with the putative Gcr1 binding site showed very low methanol induced activities (<6%). Yet, this is in agreement with the results from the deletion and insertion promoter variants, which showed high methanol induced expression upon d6 deletion. Little (6%) response on methanol was found for the basal promoter fragment AOX201-Z. In contrast to the methanol free results, core promoter AOX201 did not decrease, but rather enhance methanol induced expression. Also AOX176 basal was suitable for methanol induced expression as demonstrated by the constructs AOX176-MZAA and AOX176-RZ. However, one can speculate that combination of the same positive acting elements with core promoter AOX201 might result in even higher methanol induced activities. Analyzing the influence of region Z on methanol induced expression no consensus regarding enhancement, but also no negative influence was found.

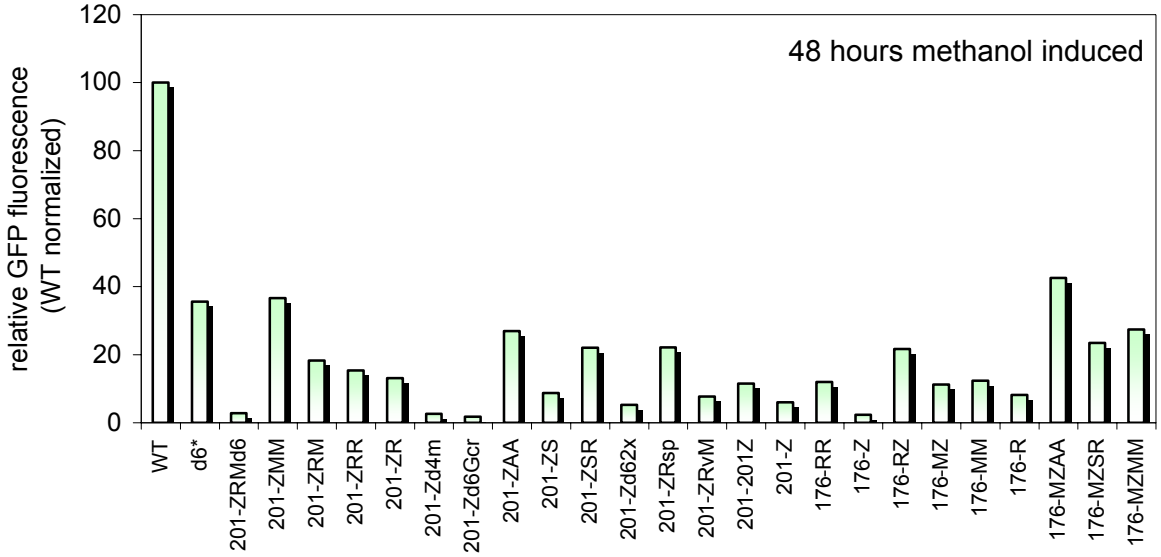



Figure 1.13 Methanol induced (48 h) GFP expression employing the new short synthetic promoter variants (microplate screening), mean value of minimum 50 clones *

* Raw data from different methanol induction times of clones based on the promoters AOX176-RR and AOX176-MM can be found in the Annex.

#### 1.3.2.1 Conclusions

Highest methanol free activity was reached by application of core promoter AOX176 in combination with the cis-acting element Rap1 (AOX176-RR, ~50% of d6*). Also element Mat1-Mc facilitated methanol free expression, which is in agreement to the results obtained from the $P_{AOX1}$ deletion and insertion variants. Comparing the two applied core promoter fragments, AOX176 seems to be advantageous for methanol free expression. While the elements Rap1 and Mat1-Mc stimulated both, expression in absence of glucose and presence of methanol, the elements Adr1 and Stre responded to methanol induction only. Significant methanol induced activity was reached using construct AOX76-MZAA (42% of the WT value). High methanol induced activities were in

addition found for the variants AOX201-ZAA, AOX201-ZSR and AOX176-RZ. In contrast, application of region d6 decreased methanol induced activity significantly.
However, these results are based on microplate experiments only and bioreactor studies need to be done to confirm or disprove the seen effects. One can speculate that the short promoters might perform even better in bioreactor comparisons due to a constant carbon source supply. Glucose depletion after ~24 of 60 hours in microscale batch cultivations might limit the energy available for expression, thus influence promoter comparisons.

### 1.3.3 New generation promoter variants III

Most of the eukaryotic promoter library approaches are based on the fusion of positive acting elements to core promoters. However, additional to TF binding also the core promoter sequence influences promoter strength and regulation (Juven-Gershon and Kadonaga, 2009). The influence of eukaryotic core promoters was demonstrated by the generation of promoter libraries with varying consensus box surroundings (Ellis et al., 2009; Tornoe et al., 2002). In addition, Juven-Gershon et al. (2006a) reported the generation of a super core promoter by the combination of common core promoter sequence elements.
Here, the goal was to design a neutral synthetic *P. pastoris* core promoter able to initiate transcription and depending on the attached positive acting element, to exhibit different regulatory properties. For the design of an artificial core promoter, the core promoter sequences of 4 different regulated promoters were aligned for consensus finding (Figure 1.14A). The core promoter sequence is usually defined between +40/-40 in relation to the transcription start site (TSS) (Juven-Gershon and Kadonaga, 2009), but was in this study extended to approximately +110/-70 according to the $P_{AOX1}$ core promoter sequence. Core promoter sequences of the following promoters were used for alignment: *Pp*$P_{GAP}$, *Pp*$P_{HIS4}$, *Pp*$P_{AOX1}$ and *Sc*$P_{ADH2}$. While $P_{GAP}$ and $P_{HIS4}$ are *P. pastoris* endogenous housekeeping promoters, $P_{AOX1}$ is repressed on glucose and induced by methanol. The *ADH2* promoter, derived from *S. cerevisiae*, is repressed in the presence and induced in the absence of glucose. Alignment resulted in the finding of two roughly defined consensus boxes located between -18/-48 and +33/+42 relative to the $P_{AOX1}$ TSS (-113 from ATG). For the other sequence parts the 50% identity rule was applied. Thus, if 2 out of 4 bases were similar the resulting consensus was used for the design of $P_{core1}$. Sequence gaps without consensus were filled randomly. Finally, $P_{core1}$ was 62% identical to $P_{AOX1}$ (Figure 1.14B). $P_{core1}$ was subsequently analyzed regarding TF binding using the program MatInspector (Genomatix Software Inc., USA). Also the 4 initial template and the following additional promoters were analyzed in regard to TF binding: *Pp*$P_{CAT}$, *Sc*$P_{ADH1}$, *Pp*$P_{FLD1}$, *Pp*$P_{ILV5}$, *Pp*$P_{ARG4}$, *Pp*$P_{DAS1}$ and *Pp*$P_{AOX2}$. TFs of the families yeast TATA binding protein factor, vertebrate TATA binding protein factor, fungal basic leucine flipper, fungal GATA binding factor, yeast heat shock factor, winged helix binding and *Asperg./Neurospora* activation of genes induced by nitrogen were identified as common and used for the TF based optimization of $P_{core1}$, thereby generating $P_{core11}$ (Figure 1.15). For the design of $P_{core11}$ also the natural locations of the applied TFs were taken into account. Comparing $P_{core11}$ to $P_{AOX1}$ an identity of 54% was found. For both, $P_{core1}$ and $P_{core11}$, the

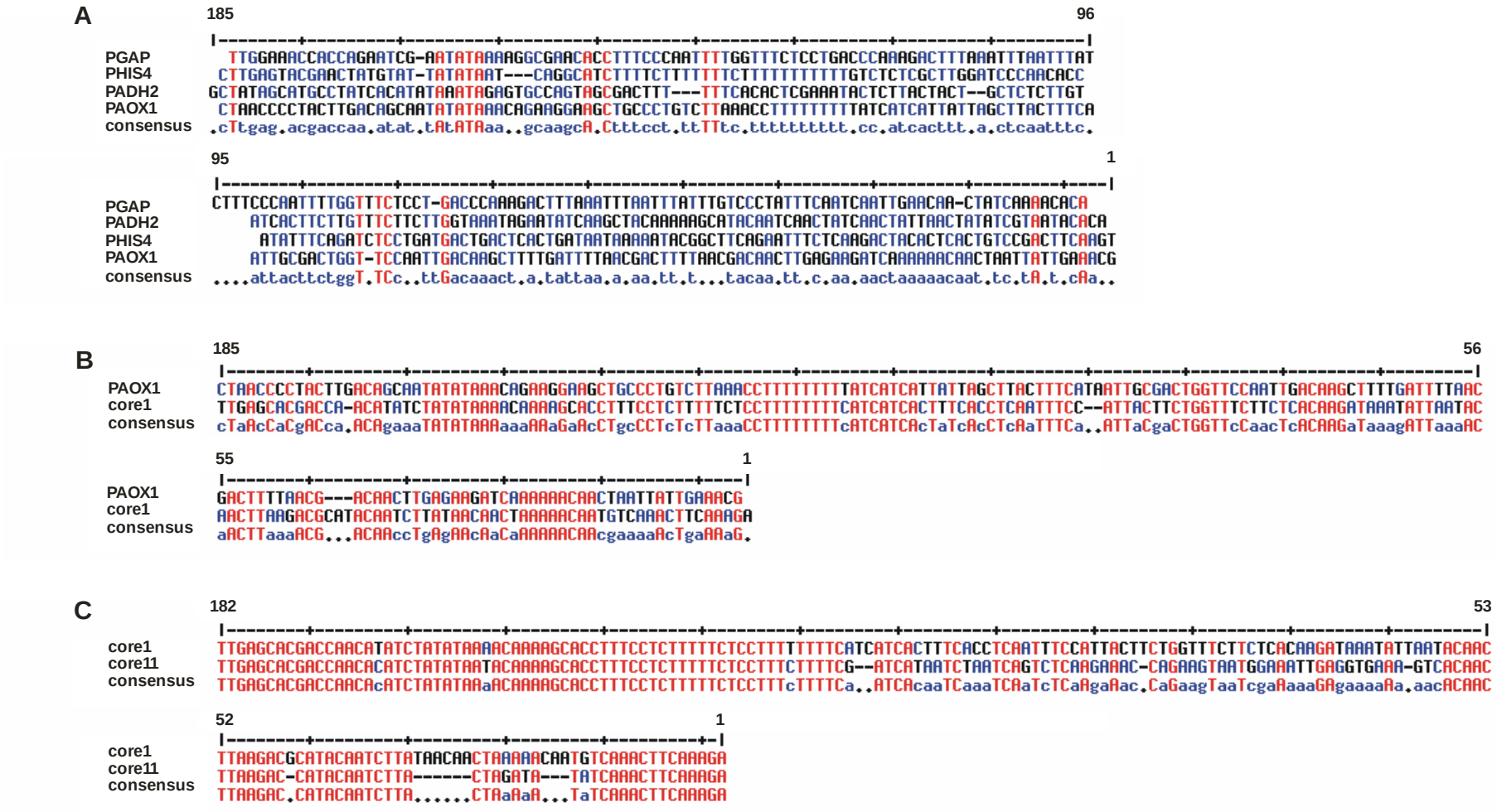
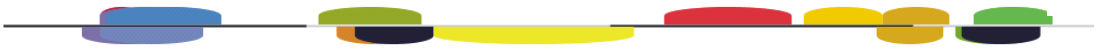



Figure 1.14 A) Sequence alignment of *Pp*$P_{GAP}$, *Pp*$P_{HIS4}$, *Pp*$P_{AOX1}$ and *Sc*$P_{ADH2}$ used for the generation of the consensus promoter core1, B) Sequence comparison of $P_{core1}$ and $P_{AOX1}$, C) Sequence comparison of $P_{core1}$ and $P_{core11}$

A

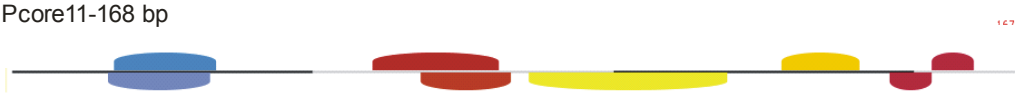
Pcore1-182 bp



| Color code | Family information | Sequence |
|---|---|---|
| | Plant TATA binding protein factor | tttaTATAgatatgt |
| | Asperg./Neurospora activ. of genes induced by nitrogen | TATCtat |
| | Plant TATA binding protein factor | tatcTATAtaaaaca |
| | Yeast TATA binding protein factor | tttgtttTATAtagata |
| | Vertebrate TATA binding protein factor | atctaTATAaaacaaaa |
| | Plant TATA binding protein factor | tctaTATAaaacaaa |
| | Vertebrate TATA binding protein factor | ctataTAAAacaaaagc |
| | Pheromone response elements | atataaAACAaaa |
| | Ribosomal RNA processing element | tccttttttTTTCatcat |
| | Pheromone response elements | tgatgaAAAAaaa |
| | Yeast mating factors | tgatGATGaaaaa |
| | Yeast heat shock factors | agaaaccagaagtaatGGAAattgaggtgaaag |
| | Neurospora crassa QA1 gene activator | caagataaatatTAATacaac |
| | Yeast heat shock factors | aagACGCatacaa |
| | Sterol regulatory element binding | tgtTATAagat |
| | Sterol regulatory element binding | tctTATAacaa |
| | Repressor of hypoxic genes | gacaTTGTtttta |
| | Yeast mating factors | tgacaTTGTtttt |
| | Fungal TALE homeodomain class | aaacaatGTCAaa |

B

Pcore11-168 bp



| Color code | Family information | Sequence |
|---|---|---|
| | Yeast TATA binding protein factor | tttgtatTATAtagatg |
| | Vertebrate TATA binding protein factor | atctaTATAatacaaaa |
| | Fungal basic leucine flipper family | tttcgatcataatcTAATcag |
| | Fungal GATA binding factors | tactGATTagattat |
| | Yeast heat shock factors | agaaaccagaagtaatGGAAattgaggtgaaag |
| | Winged helix binding | aagACGCatacaa |
| | Asperg./Neurospora activ. of genes induced by nitrogen | TATCtag |
| | Asperg./Neurospora activ. of genes induced by nitrogen | TATCaaa |

Figure 1.15 MatInspector transcription factor analyses of the promoters A) $P_{core1}$ (consensus based) and B) $P_{core11}$ (TF optimized)

transcription start site was adopted from $P_{AOX1}$ (ATCA, TSS underlined). A comparison of $P_{core1}$ and $P_{core11}$ is shown in Figure 1.14C.

Using region Z as linker for the attachment of positive acting elements, the following fully synthetic promoters were generated: core1, core11, core1-Z, core11-Z and core11-ZMM. After *P. pastoris* was transformed more than 80 clones were screened per construct in microplates and the average activities were evaluated at 0 and 48 hours of induction.

Before induction by methanol no activity was found for the designed core promoters. Further, also no significant activity was found after 48 hours of methanol induction comparing core1, core 1-Z, core11 and core11-Z to the average WT sc level. In detail, activity increased less then 1%. However, benchmarking against the $Mut^S$ negative control differences were observed. For both, core1 and core11, attachment of region Z increased methanol induced activity, 10% and 20%, respectively (Figure 1.16). However, only core11 based constructs increased activity in regard to the background. Application of core11 and core11-Z resulted in an increased activity of 9% and 33%. The best multi

copy clones based on these promoters enhanced activity 2-fold. In comparison, even the best clones of the promoters core1 and core1-Z did not increase activity in regard to the negative control.

The ability of core promoter 11 for efficient transcription initiation was shown by the attachment of the positive acting element Mat1-Mc (construct: core11-ZMM). After 48 hours of induction application of the best multi copy clones of this construct increased activity 20-fold if compared to the background level (Figure 1.17). Analyzing activity before induction, attachment of two Mat1-Mc repeats resulted in a 2-fold increase in activity, confirming the artificial promoter core11 also suitable for methanol free expression (Figure 1.18).

#### 1.3.3.1 Conclusions

Consensus and putative TF binding site based design was used to generate fully artificial promoters able to initiate transcription. However, being on the beginning of a long road application of promoter core11-ZMM resulted in a 10-times lower activity if compared to the corresponding natural core sequence based construct AOX201-ZMM. To study the applicability of the artificial core promoters different other positive acting elements have to be attached and expression data analyzed.
Though a first secondary structure analyses of the generated core and used template promoters with the RNAfold Webserver resulted in diverse free energy profiles a more detailed study of RNA structures might still reveal interesting new facts. Incorporation of common core promoter motifs similar to Juven-Gershon et al. (2006a) can also be applied for further optimization (see Figure 1.19). In addition, a comparison of such new core promoters in the context of a full *GAP* and *AOX1* promoter in comparison with their native full length promoters would be interesting.

$_{-36}$GTACTTA**TATA**AGGGGGTGGGGGCGCGTTCGTCCTCA$_{+1}$GTCGCGATCGAACA CTC$_{+18}$
GAGC$_{+22}$CGAGCAGA$_{+30}$CGT$_{+33}$GCCTACGGACCG$_{+45}$

Figure 1.19 Super core promoter 1 (SCP1) adopted from Juven-Gershon et al. (2006a), bold: TATA-box (from CMV IE1 core promoter), red: TSS, underlined: Inr (based on sequences from AdML and *D. melongaster* G retrotransposon core promoters), blue: MTE (from *D. melongaster Tollo* core promoter), green: DPE (from *Drosophila* G core promoter)*

*Note, the positions shown for the respective elements were speculated from data regarding mutations in these regions, thus might be an unreliable source

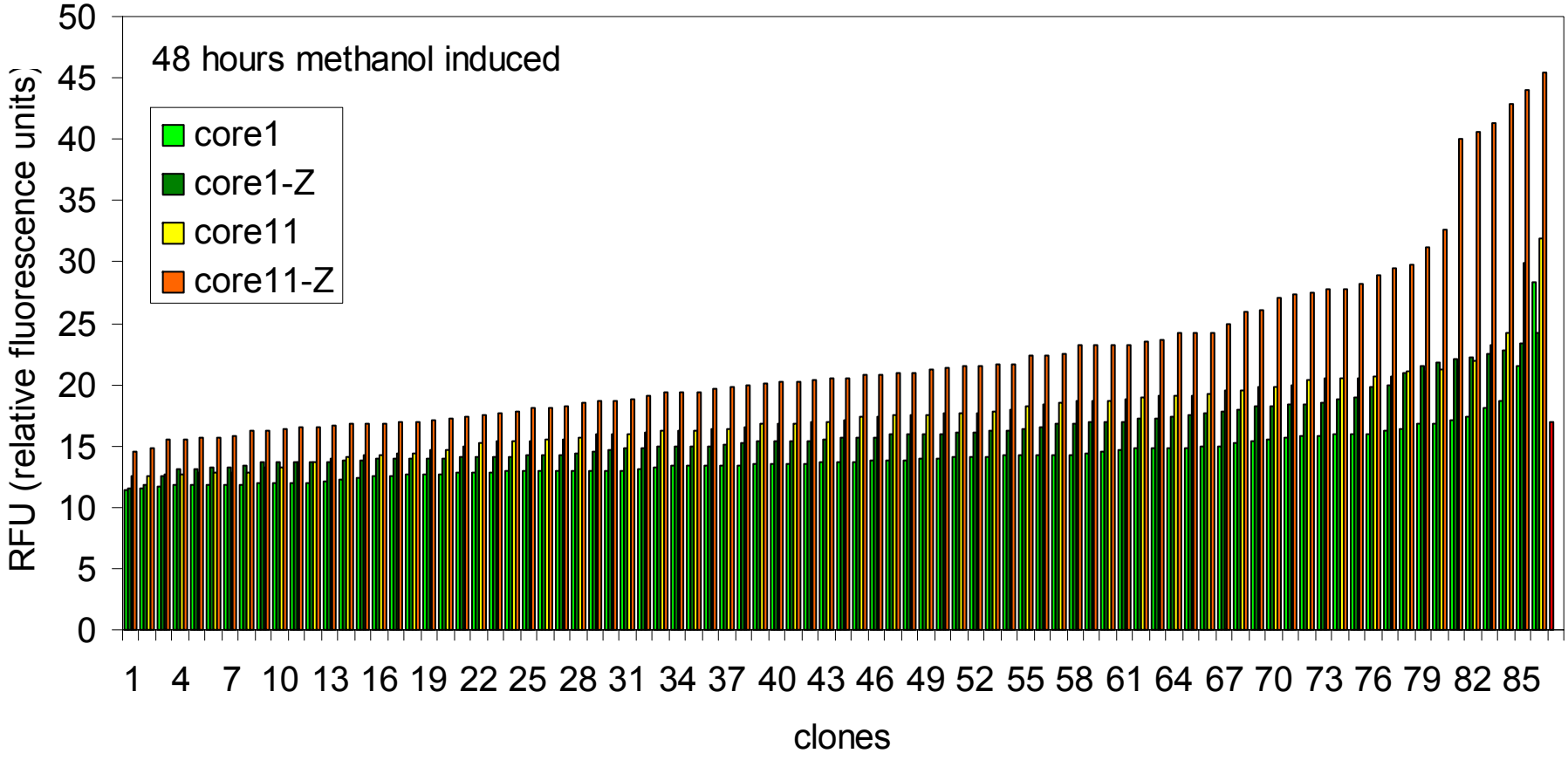



Figure 1.16 Relative GFP fluorescence after 48 hours of methanol induction employing the artificial promoters core1, core1-Z, core11, core11-Z, red: $Mut^S$ background

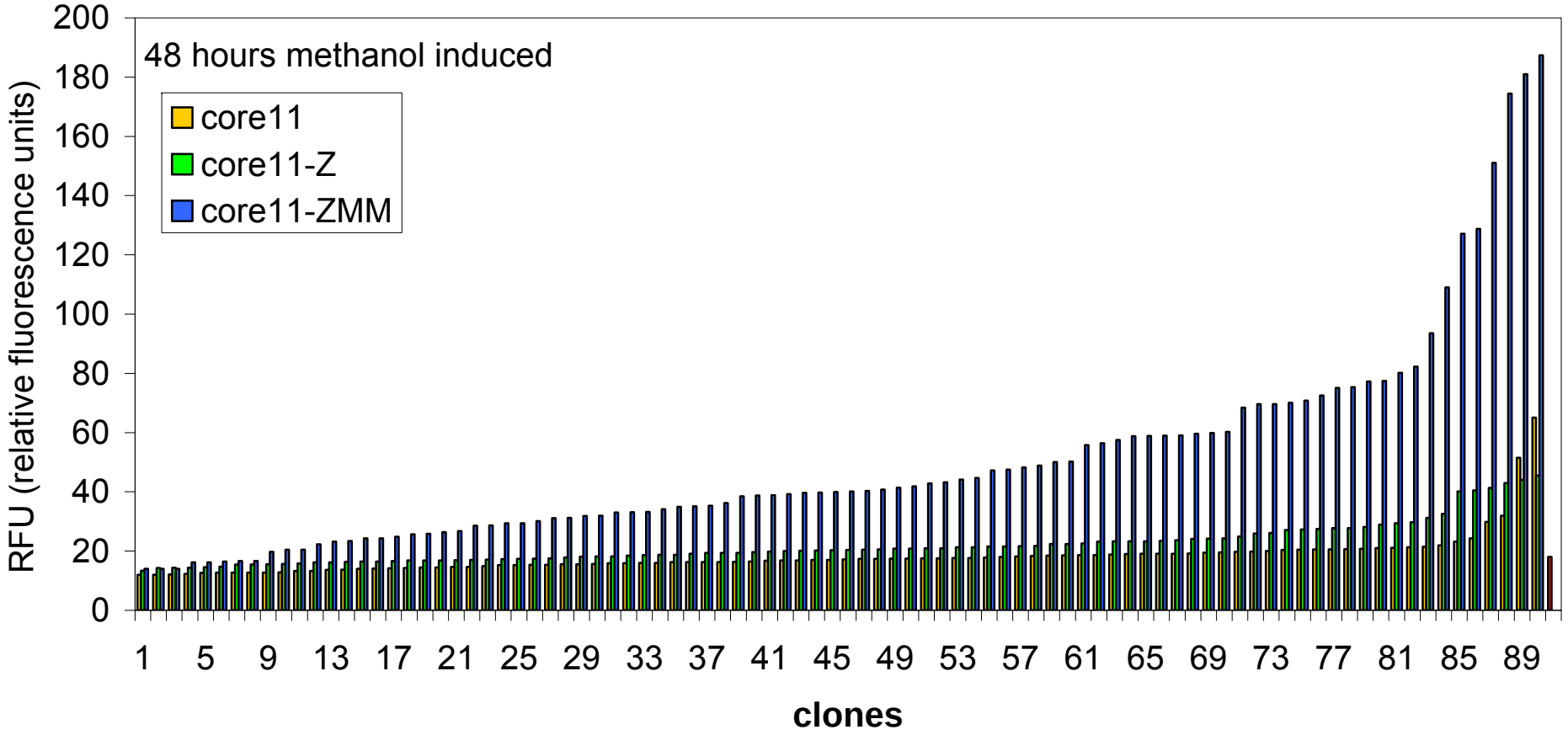



Figure 1.17 Relative GFP fluorescence after 48 hours of methanol induction employing the artificial promoters core11, core11-Z and core11-ZMM, red: MutS background

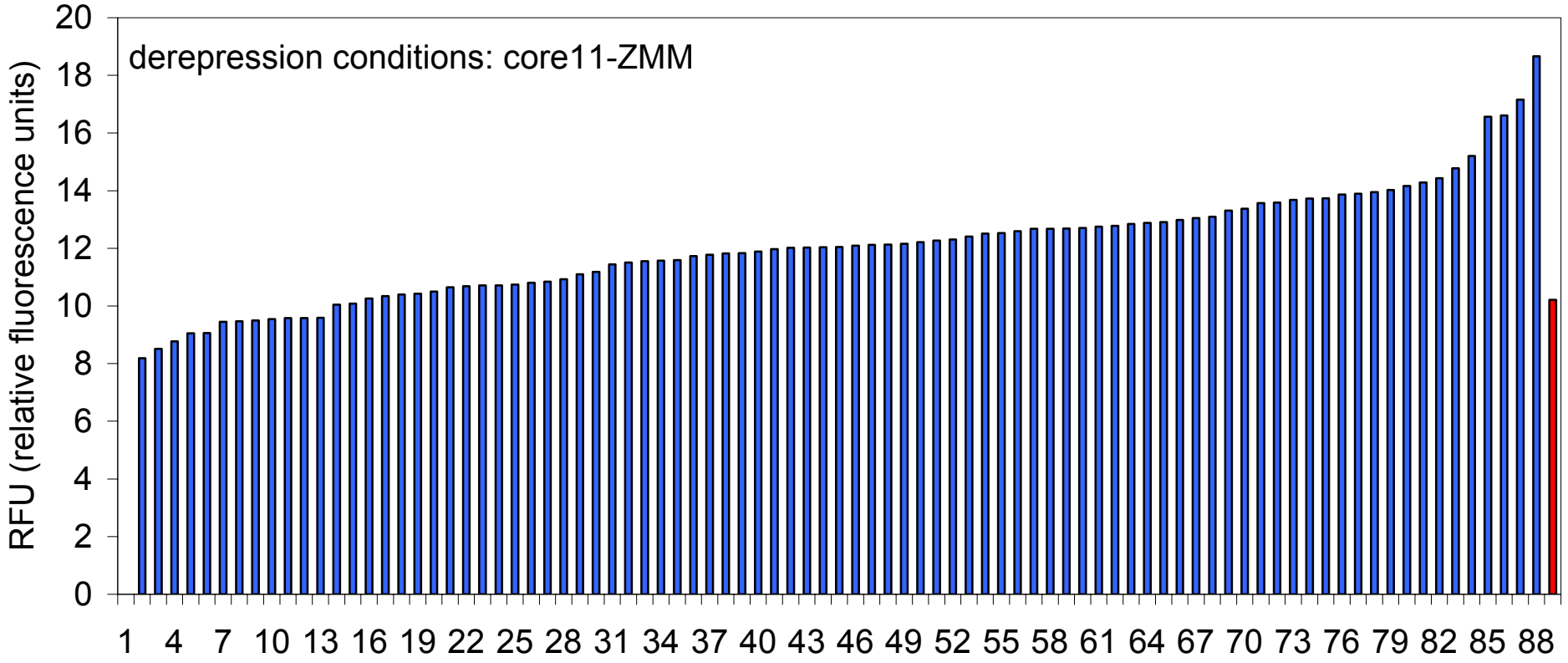



Figure 1.18 Relative GFP fluorescence after 60 hours of batch growth and glucose depletion (derepression conditions) employing the artificial promoter core11-ZMM, red: $Mut^S$ background

## 1.4 References


Burke, T.W., Kadonaga, J.T., (1997) The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. Genes Dev 11, 3020-3031.

Burke, T.W., Willy, P.J., Kutach, A.K., Butler, J.E., Kadonaga, J.T., (1998) The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box. Cold Spring Harb Symp Quant Biol 63, 75-82.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C., Forrest, A.R., Alkema, W.B., Tan, S.L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S.M., Wells, C.A., Orlando, V., Wahlestedt, C., Liu, E.T., Harbers, M., Kawai, J., Bajic, V.B., Hume, D.A., Hayashizaki, Y., (2006) Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet 38, 626-635.

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.

Ellis, T., Wang, X., Collins, J.J., (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. Nat Biotechnol 27, 465-471.

Elnitski, L., Jin, V.X., Farnham, P.J., J 1s, S.J., (2006) Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. Genome Res 16, 1455-1464.

Fang, F., Fan, S., Zhang, X., Zhang, M.Q., (2006) Predicting methylation status of CpG islands in the human brain. Bioinformatics 22, 2204-2209.

Felsenfeld, G., (1996) Chromatin unfolds. Cell 86, 13-19.

Gourse, R.L., Ross, W., Gaal, T., (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. Mol Microbiol 37, 687-695.

Hannenhalli, S., (2008) Eukaryotic transcription factor binding sites--modeling and integrative search methods. Bioinformatics 24, 1325-1331.

Hartner, F.S., Glieder A., (2005) Mutant AOX1 promoters, WO2006089329.

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.

Hawkins, J., Grant, C., Noble, W.S., Bailey, T.L., (2009) Assessing phylogenetic motif models for predicting transcription factor binding sites. Bioinformatics 25, i339-347.

Heintzman, N.D., Ren, B., (2007) The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome. Cell Mol Life Sci 64, 386-400.

Juven-Gershon, T., Cheng, S., Kadonaga, J.T., (2006a) Rational design of a super core promoter that enhances gene expression. Nat Methods 3, 917-922.

Juven-Gershon, T., Hsu, J.Y., Kadonaga, J.T., (2006b) Perspectives on the RNA polymerase II core promoter. Biochem Soc Trans 34, 1047-1050.

Juven-Gershon, T., Kadonaga, J.T., (2009) Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol 339, 225-9

Kim, H.D., O'Shea, E.K., (2008) A quantitative model of transcription factor-activated gene expression. Nat Struct Mol Biol 15, 1192-1198.

Kinkhabwala, A., Guet, C.C., (2008) Uncovering cis regulatory codes using synthetic promoter shuffling PLoS One 3, e2030.

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.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., Ebright, R.H., (1998) New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev 12, 34-44.

Lin-Cereghino, J., Wong, W.W., Xiong, S., Giang, W., Luong, L.T., Vu, J., Johnson, S.D., Lin-Cereghino, G.P., (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. Biotechniques 38, 44-48.

Liu, C.L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S.L., Friedman, N., Rando, O.J., (2005) Single-nucleosome mapping of histone modifications in S. cerevisiae. PLoS Biol 3, e328.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A.E., Wingender, E., (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res 34, D108-110.

van Nimwegen, E., (2007) Finding regulatory elements and regulatory motifs: a general probabilistic framework. BMC Bioinformatics 8 Suppl 6, S4.

Noble, W.S., Kuehn, S., Thurman, R., Yu, M., Stamatoyannopoulos, J., (2005) Predicting the in vivo signature of human gene regulatory sequences. Bioinformatics 21 Suppl 1, i338-343.

Novina, C.D., Roy, A.L., (1996) Core promoters and transcriptional control. Trends Genet 12, 351-355.

Ohler, U., Liao, G.C., Niemann, H., Rubin, G.M., (2002) Computational analysis of core promoters in the Drosophila genome. Genome Biol 3, RESEARCH0087.

Perier, R.C., Praz, V., Junier, T., Bonnard, C., Bucher, P., (2000) The eukaryotic promoter database (EPD). Nucleic Acids Res 28, 302-303.

Pilpel, Y., Sudarsanam, P., Church, G.M., (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. Nat Genet 29, 153-159.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D.K., Young, R.A., (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122, 517-527.

Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J., Madhani, H.D., (2005) Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. Cell 123, 233-248.

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., Gourse, R.L., (1993) A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. Science 262, 1407-1413.

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W., Lenhard, B., (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 32, D91-94.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P., Widom, J., (2006) A genomic code for nucleosome positioning. Nature 442, 772-778.

Smale, S.T., Baltimore, D., (1989) The "initiator" as a transcription control element. Cell 57, 103-113.

Smale, S.T., Kadonaga, J.T., (2003) The RNA polymerase II core promoter. Annu Rev Biochem 72, 449-479.

Stormo, G.D., (2000) DNA binding sites: representation and discovery. Bioinformatics 16, 16-23.

Struhl, K., (1995) Yeast transcriptional regulatory mechanisms. Annu Rev Genet 29, 651-674.

Tompa, M., Li, N., Bailey, T.L., Church, G.M., De Moor, B., Eskin, E., Favorov, A.V., Frith, M.C., Fu, Y., Kent, W.J., Makeev, V.J., Mironov, A.A., Noble, W.S., Pavesi, G., Pesole, G., Regnier, M., Simonis, N., Sinha, S., Thijs, G., van Helden, J., Vandenbogaert, M., Weng, Z., Workman, C., Ye, C., Zhu, Z., (2005) Assessing computational tools for the discovery of transcription factor binding sites. Nat Biotechnol 23, 137-144.

Tornoe, J., Kusk, P., Johansen, T.E., Jensen, P.R., (2002) Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites. Gene 297, 21-32.

Tsutsumi, N., Terui, Y., Takagi, S., Kong, X.Y, (2007) Method for methanol independent induction from methanol inducible promoters in Pichia, WO2008090211.

Venter, M., (2007) Synthetic promoters: genetic control through cis engineering. Trends Plant Sci 12, 118-124.

Wang, W., Malcolm, B.A., (1999) 2-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. Biotechniques 26, 680-682.

Wasserman, W.W., Sandelin, A., (2004) Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet 5, 276-287.

Weis, R., Luiten, R., Skranc, W., Schwab, H., Wubbolts, M., Glieder, A., (2004) Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena. FEMS Yeast Res 5, 179-189.

Westholm, J.O., Xu, F., Ronne, H., Komorowski, J., (2008) Genome-scale study of the importance of binding site context for transcription factor binding and gene regulation. BMC Bioinformatics 9, 484.

Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., Rando, O.J., (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309, 626-630.

Zhou, D., Yang, R., (2006) Global analysis of gene transcription regulation in prokaryotes. Cell Mol Life Sci 63, 2260-2290.

## 1.5 Annex

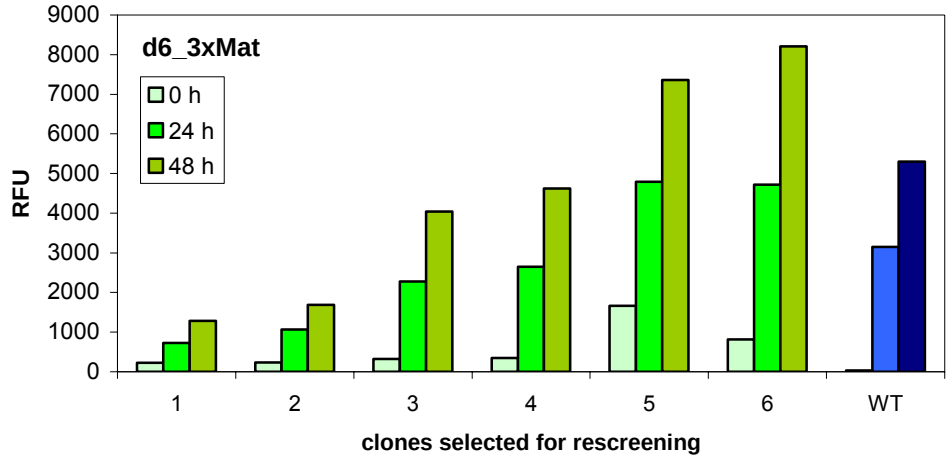



Figure 1.20 Screening results of the 6 for rescreening selected clones of promoter variant d6_3xMat in comparison to the best WT promoter based clone, relative fluorescence units (RFU) at 0, 24 and 48 hours of methanol induction

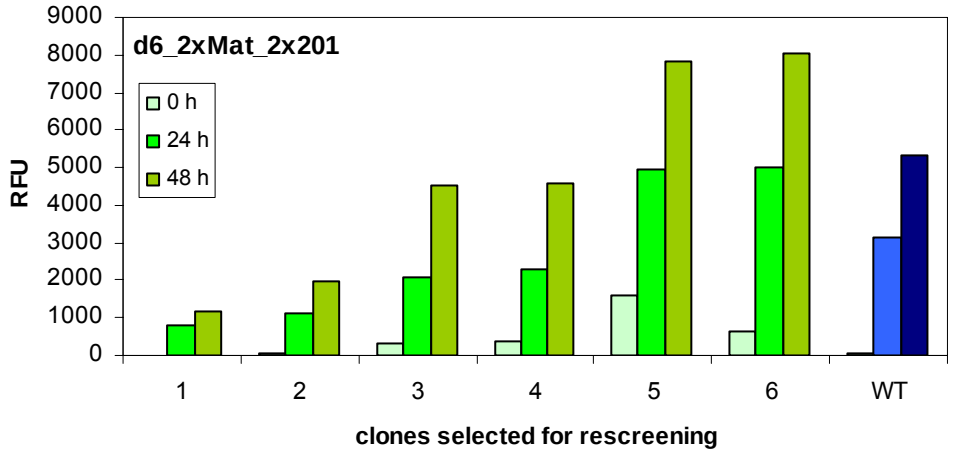



Figure 1.21 Screening results of the 6 for rescreening selected clones of promoter variant d6_2xMat_2x201 in comparison to the best WT promoter based clone, relative fluorescence units (RFU) at 0, 24 and 48 hours of methanol induction

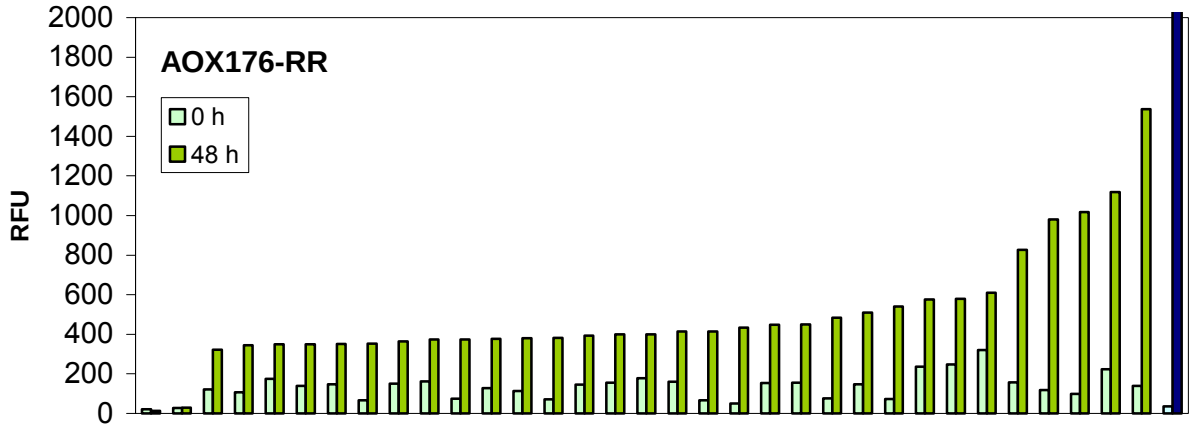



Figure 1.22 Screening results of clones based on promoter variant AOX176-RR in comparison to the best WT promoter based clone, relative fluorescence units (RFU) at 0 and 48 hours of methanol induction

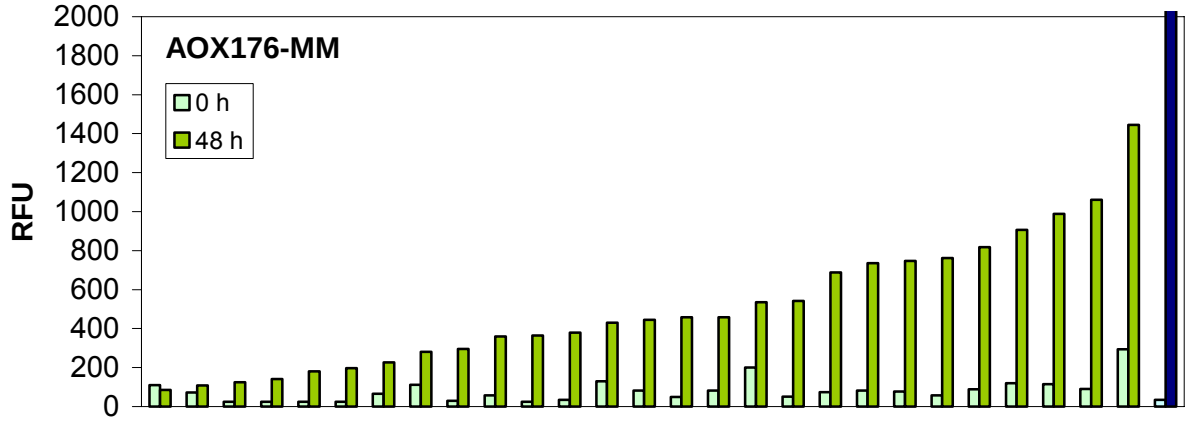



Figure 1.23 Screening results of clones based on promoter variant AOX176-MM in comparison to the best WT promoter based clone, relative fluorescence units (RFU) at 0 and 48 hours of methanol induction

Table 1.2 SDM and PCR primers, bold: positive acting elements, italic: 3' element extensions (see 1.2.4.2), underlined: region ZUS (Z), which links cis-acting elements to core promoters

| name | sequence |
|---|---|
| Matd6fw | GATTATGCATTGTCTCCAC**ATTATGCATTGTCTCCACAC**TGCTGATAGCCTAACGTTC |
| Matd6rv | GAACGTTAGGCTATCAGCAGTGTGGAGACAATGCATAATGTGGAGACAATGCATAATC |
| Matd6*fw | GATTATGCATTGTCTCCAC**ATTATGCATTGTCTCCACAC**TGCTGATAGCCACGTTCATG |
| Matd6*rv | CATGAACGTGGCTATCAGCAGTGTGGAGACAATGCATAATGTGGAGACAATGCATAATC |
| *2x201fw | CTGCTGATAGCCACGTT**CATGATCAAAATTTCATGATCAAAATTT**AACTGTTCT AACCCCTAC |
| *2x201rv | GTAGGGGTTAGAACAGTTAAATTTTGATCATGAAATTTTGATCATGAACGTGGCTATCA GCAG |
| d6fw | GTCTCCACACTGCTGATAGCCTAACGTTC |
| d6rv | GGCTATCAGCAGTGTGGAGACAATGCATAATCATC |
| d6*mut_fw | GATTATGCATTGTCTCCACACTGCTGATAGCCACGTTCATGAT AAAATTTAAC |
| d6*_mut_rv | TTTTGATCATGAACGTGGCTATCAGCAGTGTGGAGACAATGCATAATCATC |
| 2012xfw | GCTGATAGCCTAACGTT**CATGATCAAAATTTCATGATCAAAATTT**AACTGTTC |
| 2012xrv | GTTAAATTTTGATCATGAAATTTTGATCATGAACGTTAGGCTATCAGCAGTATTC |
| fw2_RMd6_Z | TTCAGATCT*CCGAATGCAACAAGCT***CCGCATTACACCCGAACATCACTCCGATTAT GCATTG TCTCCACACATTGTATGCTTCCAAGATTCTGGTGGGAAT**TGCTGATAGCCTAACGTT |
| fw2_2xMat_Z | TTCAGATCT*CTTTTTGGAT***GATTATGCATTGTCTCCACACGATTATGCATTGTCTCCA CAC**TGCTGATAGCCTAACGTT |
| fw2_RapMat_Z | TTCAGATCT*CCGAATGCAACAAGCT***CCGCATTACACCCGAACATCACTCCGATTATG CATTGTCTCCACAC**TGCTGATAGCCTAACGTT |
| fw2_2xRap_Z | TTCAGATCT*CCGAATGCAACAAGCT***CCGCATTACACCCGAACATCACTCCCCGCATTA CACCCGAACATCACTCC**TGCTGATAGCCTAACGTT |
| fw2_Rap_Z | TTCAGATCT*CCGAATGCAACAAGCT***CCGCATTACACCCGAACATCACTCC**TGCTGA TAGCCTAACGTT |
| fw2_2xMat_Z+ | TTCAGATCT*CTTTTTGGATG***ATTATGCATTGTCTCCACACGATTATGCATTGTCTCC ACAC**TGCTGATAGCCTAACGTTTGCTGATAGCCTAACGTT |
| fw1_Z+_2x201 | CCTAACGTTTGCTGATAGCCTAACGTT**CATGATCAAAATTTCATGATCAAAATTT**AACT GTTC |
| fw1_Z_2x201 | TGCTGATAGCCTAACGTT**CATGATCAAAATTTCATGATCAAAATTT**AACTGTTC |
| fw2_d4m_Z | TTCAGATCT*CTTGT***TTGGTATTGATTGACG**TGCTGATAGCCTAACGTT |

| | |
|---|---|
| fw1_2x201_Z | TTCAGATCTTGCTGATAGCCTAACGTT**CATGATCAAAATTTCATGATCAAAATTT** AACTGTTCTAACC |
| fw1_201_Z | TTCAGATCTTGCTGATAGCCTAACGTT**CATGATCAAAATTT**AACTGTTCTAACC |
| fw1_176_Z | TTCAGATCTTGCTGATAGCCTAACGTTGACAGCAATATATAAACAGAAGG |
| fw1_176_RZ | TTCAGATCTTGCTGATAGCCTAACGTT**CCGCATTACACCCGAACATCACTCC**GACA GCAATATATAAACAGAAGG |
| fw1_176_MZ | TTCAGATCTTGCTGATAGCCTAACGTT**GATTATGCATTGTCTCCACAC**GACAGCA ATATATAAACAGAAGG |
| fw1_176_MM | TTCAGATCT**GATTATGCATTGTCTCCACACGATTATGCATTGTCTCCACAC**G ACAGCAATATATAAACAGAAGG |
| fw2_d6_Gcr_Z | TTCAGATCT**ATTGTATGCTTCCAAGATTCTGGTGGGAAT**ACTGCTGATAGCCT AACGTT |
| fw2_Adr_Z | TTCAGATCT**GCTTTCTGAGTGTGGGGTCAAATAGTTTC**TGCTGATAGCCTAACGTT |
| fw2_2xAdr_Z | TTCAGATCT**GCTTTCTGAGTGTGGGGTCAAATAGTTTCGCTTTCTGAGTGTGGG GTCAAATAGTTTC**TGCTGATAGCCTAACGTT |
| fw2_Stre_Z | TTCAGATCT**CCTGGCCCCCCTGGCGAGG**TGCTGATAGCCTAACGTT |
| fw2_SR_Z | TTCAGATCT**CCTGGCCCCCCTGGCGAGGCCGCATTACACCCGAACATCACTCC** TGCTGATAGCCTAACGTT |
| 11fw | AAAGATCTTTGAGCACGACCAACACATC |
| 1fw | AAAGATCTTTGAGCACGACCAACATATC |
| 11rv | TTGAATTCTTTGAAGTTTGATATATCTAG |
| 1rv | TTGAATTCTTTGAAGTTTGACATTGTTTTTAG |
| cr_rspfw* | AAAGATCT**CTTATGATCCGTCTCTCCGGTTACAGC**TGCTGATAGCCTAACGTT |

*rsp: carbon source responsive element, MatInspector *in-silico* analysis ($P_{AOX1}$, $P_{ARG}$, $P_{GAP}$)

Table 1.3 Plasmids coding for GFP expression under the control of different synthetic promoters; *Pichia pastoris* strains are available as glycerol stocks

| strain coll. | plasmid | host |
|---|---|---|
| 5751 | T4_GFP_d6_3xMat | K12 DH5α |
| 5752 | T4_GFP_d6_2x201 | K12 DH5α |
| 5753 | T4_GFP_d6*_2x201 | K12 DH5α |
| 5754 | T4_GFP_d6_2xMat_2x201 | K12 DH5α |
| 5755 | T4_GFP_d6*_2xMat_2x201 | K12 DH5α |
| 5756 | T4_GFP_d6*_2xMat_2x201* | K12 DH5α |
| 5757 | T4_GFP_d6*_2xMat_2x201** | K12 DH5α |
| 5758 | T4_GFP_2x201* | K12 DH5α |

| | | |
|---|---|---|
| 5759 | T4_GFP_2x201 | K12 DH5α |
| 5760 | T4_GFP_2x201** | K12 DH5α |
| 5761 | T4_GFP_d6_2xMat | K12 DH5α |
| 5762 | T4_GFP_d6*_2xMat | K12 DH5α |
| 5763 | T4_GFP_d6** | K12 DH5α |
| 5889 | pPpT4_GFP_Rap_Z_176 | K12 DH5α |
| 5890 | pPpT4_GFP_2xRap_176 | K12 DH5α |
| 5891 | pPpT4_GFP_StreRap_Z_176 | K12 DH5α |
| 5892 | pPpT4_GFP_2xMat_Z | K12 DH5α |
| 5893 | pPpT4_GFP_2xAdr_Z | K12 DH5α |
| 5894 | pPpT4_GFP_2xMat_176 | K12 DH5α |
| 5773 | T4_GFP_d6* | K12 DH5α |
| 5716 | T4_GFP | K12 DH5α |
| 6107 | P176MZMMGFP | K12 DH5α |
| 6108 | P176MZGFP | K12 DH5α |
| 6109 | P214ZRGFP | K12 DH5α |
| 6110 | P214ZRRGFP | K12 DH5α |
| 6111 | pPpT42x201ZGFP | K12 DH5α |
| 6112 | P201ZGFP | K12 DH5α |
| 6113 | P176ZGFP | K12 DH5α |
| 6086 | pPpT4214ZRMd6GFP | K12 DH5α |
| 6092 | PpPT4core1GFP | K12 TOP10F´ |
| 6093 | PpPT4core11GFP | K12 TOP10F´ |
| 6094 | PpPT4214rspGFP | K12 TOP10F´ |
| 6095 | PpPT4176RGFP | K12 TOP10F´ |
| 6096 | PpPT4176MZAAGFP | K12 TOP10F´ |
| 6099 | PpPT4core11ZGFP | K12 TOP10F´ |
| 6100 | PpPT4176MZSRGFP | K12 TOP10F´ |
| 6101 | PpPT4core1ZGFP | K12 TOP10F´ |
| 6102 | PpPT4core11ZMMGFP | K12 TOP10F´ |

# 2 Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in *Pichia pastoris*

This work was based on the Diploma thesis of Thomas Zuellig.
The manuscript was accepted for publication in Systems and Synthetic Biology (Springer) on 02-05-2010.

Ruth, C.[1], Zuellig, T.[3], Mellitzer, A.[1], Weis, R.[4], Looser, V.[3], Kovar, K.[3] and Glieder, A.[1,2]

[1] Institute of Molecular Biotechnology, Graz University of Technology, 8010 Graz, Austria

[2] Research Centre Applied Biocatalysis, 8010 Graz, Austria

[3] Institute of Biotechnology, University of Applied Sciences Zurich (ZHAW), Grüental, 8820 Waedenswil, Switzerland

[4] VTU Technology Technologieentwicklungsgesellschaft m.b.H., 8010 Graz, Austria

**Abstract**

Natural tools for recombinant protein production show technological limitations. Available natural promoters for gene expression in *Pichia pastoris* are either constitutive, weak or require the use of undesirable substances or procedures for induction. Here we show the application of deletion variants based on the well known methanol inducible *AOX1* promoter and small synthetic promoters, where cis-acting elements were fused to core promoter fragments. They enable differently regulated target protein expression and at the same time to replace methanol induction by a glucose or glycerol feeding strategy. Trypsinogen, the precursor of the serine protease trypsin, was expressed using these different promoters. Depending on the applied promoter the production window (i.e. the time of increasing product concentration) changed significantly. In fedbatch processes trypsinogen yields before induction with methanol were up to 10 times higher if variants of the *AOX1* promoter were applied. In addition, the starting point of autoproteolytic product degradation can be predetermined by the promoter choice.

Keywords: *Pichia pastoris*, *AOX1* promoter variants, porcine trypsinogen, production window

## 2.1 Introduction

The applicability of simple genetic techniques, strong expression of foreign genes and the feasibility of high cell density cultures made the methylotrophic yeast *Pichia pastoris* a favourable host system for heterologous protein production (Cregg et al., 1985; Cregg et al., 2000; Cregg et al., 1993). High levels of foreign proteins have been expressed in *P. pastoris* under the control of the strong, methanol inducible *AOX1* promoter (Hasslacher et al., 1997; Werten et al., 1999; Xiong et al., 2006). A benchmark was set in 1997 by the intracellular production of an industrially applied (S)-hydroxynitrile lyase with a titer of 22 g $l^{-1}$ protein (Hasslacher et al., 1997). Generally, the *Pichia pastoris* system is a faster, more simple and less expensive alternative to higher eukaryotic expression systems, such as mammalian cell cultures, still able to perform typical eukaryotic posttranslational modifications like glycosylation or disulfide bridge formation (Cregg et al., 1985; Cregg et al., 1993). Although *Pichia pastoris* has often been used successfully in recombinant protein production still little is known about the regulation of the *AOX1* promoter on a molecular level (Hartner et al., 2008). Briefly, transcription can be regulated by positively and negatively acting sequences in the promoter (Verdier, 1990). Glucose and other repressing sugars can affect the rate of transcription by interfering with activators or repressors directly or by regulating the expression of such transcription factors. Inan et al. (2004) first described two cis-acting elements, A and C, within the *AOX1* promoter ($P_{AOX1}$) binding to yet unknown DNA-binding proteins. In addition, the positive acting transcription factor Mxr1p interacting with the *AOX1* promoter and responsible for the activation of many genes in response to methanol was identified by Lin-Cereghino et al. (2006). Kranthi et al. (2009) identified 6 Mxr1p binding sites within the *AOX1* promoter. In addition, a comprehensive study of Hartner et al. (2008) described the influence of mutations, deletions and multiplications of putative transcription factor binding sites on the regulatory properties of the *AOX1* promoter. Indicating Mxr1p as an important regulator of the methanol utilization in *P. pastoris*, deletion of 5 of the 6 Mxr1p sites resulted in reduced activity on methanol. In agreement also Xuan et al. (2009) showed the *AOX1* promoter region -638 to -510 to be upstream activating. Additionally to mechanistic information, the developed library of Hartner et al. (2008) allows adjusting the level of transcription according to the individual expressed genes. Further, first *AOX1* promoter variants became available, which allow glucose regulated expression based on a simple repression-derepression principle.

In this study we show the effect of different synthetic promoter variants on heterologous gene expression in *Pichia pastoris*. A major interest was put in the question how differently regulated promoters influence production window and product quality under different environmental conditions, here in presence of methanol, glucose or glycerol. For these comparative studies our new *Escherichia coli/Pichia pastoris* shuttle vector (pPpT2) was employed. As model target porcine trypsinogen was selected. Trypsinogen (zymogen) is the inactive precursor of the serine protease trypsin which is secreted in the pancreas (Hanquier et al., 2003; Stroud et al., 1977). In the small intestine it is further activated to trypsin through proteolytic cleavage by enterokinase. Enterokinase cleaves after the lysine of the $(Asp)_4$-Lys sequence in the propeptide and releases active protease (Hanquier et al., 2003). Active trypsin can cleave its own zymogen by cleaving at the carboxyl-terminal ends of lysine and arginine starting an autocatalytic cascade. Trypsin is used in industrial biotechnology for the processing

of enzymes and pharmaceutical proteins (e.g. insulin production). Thus its precursor trypsinogen also became an interesting model target for protein expression studies (Chance et al., 1997; Frank et al., 1995; Hanquier et al., 2003).

## 2.2 Materials and methods

### 2.2.1 Strains and plasmids

Porcine trypsinogen (GenBank accession no. P00761) was codon optimized applying a *Pichia pastoris* high methanol codon usage using the Gene Designer Software (DNA2.0, Menlo Park, CA, USA). For secretion the signal sequence of the yeast mating factor α, also codon optimized, was added by overlap extension PCR (polymerase chain reaction). The obtained construct was ligated into the multiple cloning site (MCS, *Eco*RI/*Not*I) of our *E. coli*/*P. pastoris* shuttle vector pPpT2 (Figure 2.1). The *AOX1* promoter and terminator sequences, which were synthesized by GenScript (Piscataway, NJ, USA) based on the sequence information from GenBank (accession no. U96967), were used for the construction of the plasmid pPpT2. To allow selection against the antibiotic zeocin we used the Ble gene from *S. hindusdanus* (GenBank accession no. A31898), which was optimized for expression in *E. coli* and *P. pastoris* (Leto Software, Entelechon Corp.). In the referred plasmid pPpT2 the zeocin resistance gene was under the control of a 3' 34 bp truncated 552 bp long fragment of the *ILV5* promoter and the *AOD* terminator (466 bp), both derived from *P. pastoris*, strain CBS7435. For bacterial replication the pUC origin (pBR322) was used. For expression of the resistance marker in *E. coli* a synthetic prokaryotic consensus promoter (P EM72) was embedded between the 3' truncated eukaryotic promoter and the start of the resistance gene.
All experiments were performed using the *Pichia pastoris* strain CBS7435, obtained from CBS fungal biodiversity center. Subcloning was done using the *E. coli* strain DH5a-T1$^R$ (Invitrogen, Carlsbad, CA, USA).

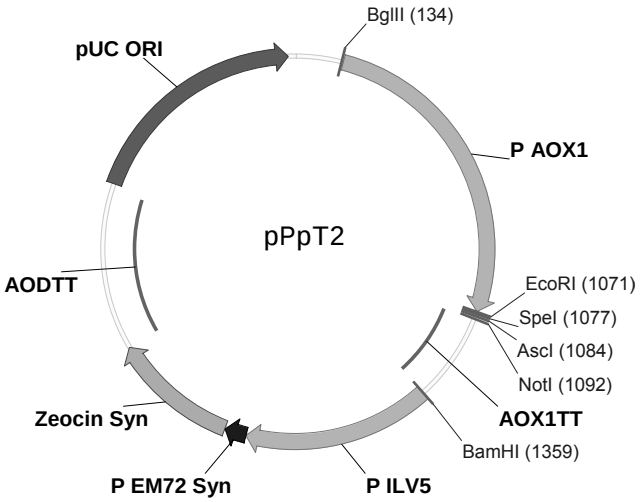



Figure 2.1 *Pichia pastoris* shuttle vector pPpT2 (3546 bp), P AOX1: *AOX1* promoter, AOX1TT: *AOX1* terminator, MCS (multiple cloning site): *Eco*RI, *Spe*I, *Asc*I, *Not*I, P ILV5: promoter of the *P. pastoris ILV5* gene, P EM72: synthetic *E. coli* promoter, Zeocin Syn: synthetic Zeocin gene, AODTT: terminator of the *P. pastoris* AOD gene, pUC ori: origin of replication

### 2.2.2 Chemicals and media

Oligonucleotide primers were obtained from Eurogentec (Seraing, Belgium). For plasmid isolation the GeneJET™ Plasmid Miniprep Kit of Fermentas (Burlington, Ontario, Canada) was used. PCR fragment purification was done with the Wizard® SV Gel and PCR Clean-Up System of Promega (Madison, WI, USA). Chemicals were purchased if not stated otherwise from Becton, Dickinson and Company (Franklin Lakes, NJ, USA), Fresenius Kabi Austria (Graz, Austria), Carl Roth (Karlsruhe, Germany), and Sigma-Aldrich (St Louis, MO, USA).

*E. coli* complex media was composed of 10 g $l^{-1}$ trypton, 5 g $l^{-1}$ yeast extract and 5 g $l^{-1}$ NaCl. For selection in *E. coli* a concentration of 25 µg $ml^{-1}$ zeocin was used. Complex *Pichia pastoris* media (YPD) contained 10 g $l^{-1}$ yeast extract, 20 g $l^{-1}$ peptone and 20 g $l^{-1}$ glucose. For antibiotic selection 100 µg $ml^{-1}$ zeocin were used. 15 g $l^{-1}$ agar was added for plate media. Buffered minimal media BMD (1%), BMM2 and BMM10 consisted per liter of 200 ml 1 M potassium phosphate buffer (pH 6), 13,4 g yeast nitrogen base without amino acids, 0,0004 g $l^{-1}$ biotin and 11 g $l^{-1}$ glucose or 1 or 5% (v/v) methanol, respectively. BMGY media used for inocolum cultures for fedbatch cultivations contained as carbon source 10 g $l^{-1}$ glycerol (Invitrogen, Carlsbad, CA, USA). Batch mineral medium contained per liter 30 g glucose, 0,17 g $CaSO_4\_2H_2O$, 2,86 g $K_2SO_4$, 0,64 g KOH, 2,32 g $MgSO_4\_7H_2O$, 0,22 g NaCl, 0,6 g EDTA disodium salt and 4,25 ml $H_3PO_4$ (85%). Additionally 0.00087 g biotin and 4,35 ml $PTM_1$ mineral salt solution were added to the media after sterilisation. $PTM_1$ mineral salt solution (Invitrogen, Carlsbad, CA, USA) contained per liter 5 ml $H_2SO_4$ (69%), 3,84 g $CuSO_4$, 0,08 g NaI, 3,0 g $MnSO_4\_H_2O$, 0,2 g $Na_2MoO_4\_2H_2O$, 0,02 g $H_3BO_3$, 105 0,92 g $CoCl_2\_6H_2O$, 20,0 g $ZnCl_2$ and 65,0 g $FeSO_4\_7H_2O$. 12 ml $PTM_1$ salt solution and 0,0024 g biotin were added to the feed solutions. To avoid foam antifoam A

(10%) was added in a concentration of 100 µl $l^{-1}$. The glucose substrate feed contained 632 g $kg^{-1}$ α-D(+)-glucose monohydrate, the methanol substrate feed 99% methanol.

### 2.2.3 AOX1 promoter variants

The knowledge about possible regulatory sites on the *AOX1* promoter provided the starting point for this work (Hartner et al., 2008). Representative deletion variants and multiplications of cis-acting elements from this previous study as well as new promoter variants were compared in this study with respect to technological aspects for the expression of porcine trypsinogen.

The promoter variant d1 as well as the improved combination variant $d1^+$ increased expression in screening studies, using green fluorescent protein (GFP) as reporter, to 133 and 166%, if compared to the wild type (WT) *AOX1* promoter on single copy level after 72 hours of methanol induction (Hartner et al., 2008). Similar positive effects were found employing these constructs for other intracellular and secreted model enzymes. In contrast, under the same conditions the promoter variants d6 and d6* resulted in lower expression (44 and 55%, respectively relative to the wild type promoter), but significant expression started already before induction. Chosen short promoter variants were named AOX176-Mat1-Mc and AOX176-201-214. Single copy transformants of these variants showed very low methanol induced GFP expression (5-10%), but the ability to express protein already before induction. From these results it can be speculated that the short synthetic promoters might also be regulated via a glucose depletion-derepression mechanism as it was described for the variants d6 and d6*.

For further studies about the apparently important region d1, a new $P_{AOX1}$ double deletion variant, namely dHap2345-1z1, was generated based on construct dHap2345-1 (Hartner et al., 2008). In addition, a new cis-acting element core promoter fusion called AOX176-Rap1 was synthesized. Although no significant effect on promoter activity was found by Hartner et al. (2008) by doubling the putative Rap1p binding region, the previous deletion studies indicated an importance of the Rap1 region showing only 34% residual activity upon deletion. For this reason the putative Rap1p binding region was also included in further promoter design as described below.

### 2.2.4 Design of the AOX1 promoter variants

Sequence numeration was done in consensus to the natural *AOX1* promoter in upstream direction starting at the start codon of the *AOX1* open reading frame ($CG_{(-1)}A_{(+1)}TG$). Variant d1 consisted of a deletion of 66 bp between position -711 and -776. The variant $d1^+$ was based on d1 and contained an additional multiplication of 14 bases (201-214, located at position -189). Variant d6 was made by a deletion of 30 bases at position -223. Construct d6* contained an additional deletion of 2 base pairs, located 3' of d6 (ΔTA/-208/-209). Construct dHap2345-1z1 was a double deletion variant with deleted regions between -746/-750 and -753/-755. The short synthetic promoters consisted of the putative strong positive acting elements Rap1 (-599/-615), Mat1-Mc (-253/-269) and 201-214 (-189/-202) fused to the core (basal) promoter element AOX176 (168 bp). Figure 2.2 shows these variants in comparison to the wild type *AOX1* promoter.

### 2.2.5 Generation of the AOX1 promoter variants

The employed *AOX1* promoter variants d1, d1$^{+}$, d6 and d6* obtained from Hartner and Glieder (2005) were amplified by PCR using following primers (restriction sites underlined): PAOX1Bglfw 5'-TTCAGATCTAACATCCAAAGACGAAAGG-3', PAOX1Ecorv 5'-TTGAATTCTTTCAATA ATTAGTTGTTTTTTG-3'). For the PCR amplifications Phusion™ High-Fidelity DNA Polymerase distributed by New England Biolabs (Ipswich, MA, USA) was used. Promoter variant dHap2345-1z1 was generated by site directed mutagenesis (SDM) according to Wang and Malcolm (1999). Also, small synthetic promoter variants were amplified by PCR, but using PfuUltra™ polymerase (Stratagene Inc.). To generate the small promoter variants the positive acting elements Mat1-Mc, Rap1, 201-214 were attached to the core promoter fragment AOX176 (168 bp, pPpT2) by long high quality PCR primers. According to Hartner et al. (2008) a *Bsp*TI restriction site was introduced as linker between the positive acting element and the core promoter fragment. Following primers were used for amplification: Mat1-McAOX176 fw 5'-AAAGATCT**TTATGCATTGTCTCCAC**CTTAAG*GACAGCAATAT ATAAACAGAAG*-3', Rap1AOX176fw 5'-AAAGATCT**TTACACCCGAACATCAC**CTTAAG *GACAGCAATATATAAACAGAAG*-3', 201-214AOX176fw 5'-AAAGATCT**CATGAT CAAAATTTCATGATCAAAATTT**CTTAAG*GACAGCAATATATAAACAGAAG*-3' (reverse primer PAOX1Ecorv, restriction sites underlined, positive acting elements bold, core promoter binding site italic). Amplified promoters were digested with *Bgl*II/*Eco*RI and ligated into pPpT2. Fermentas enzymes (Burlington, Ontario, Canada) were used for all cloning steps.

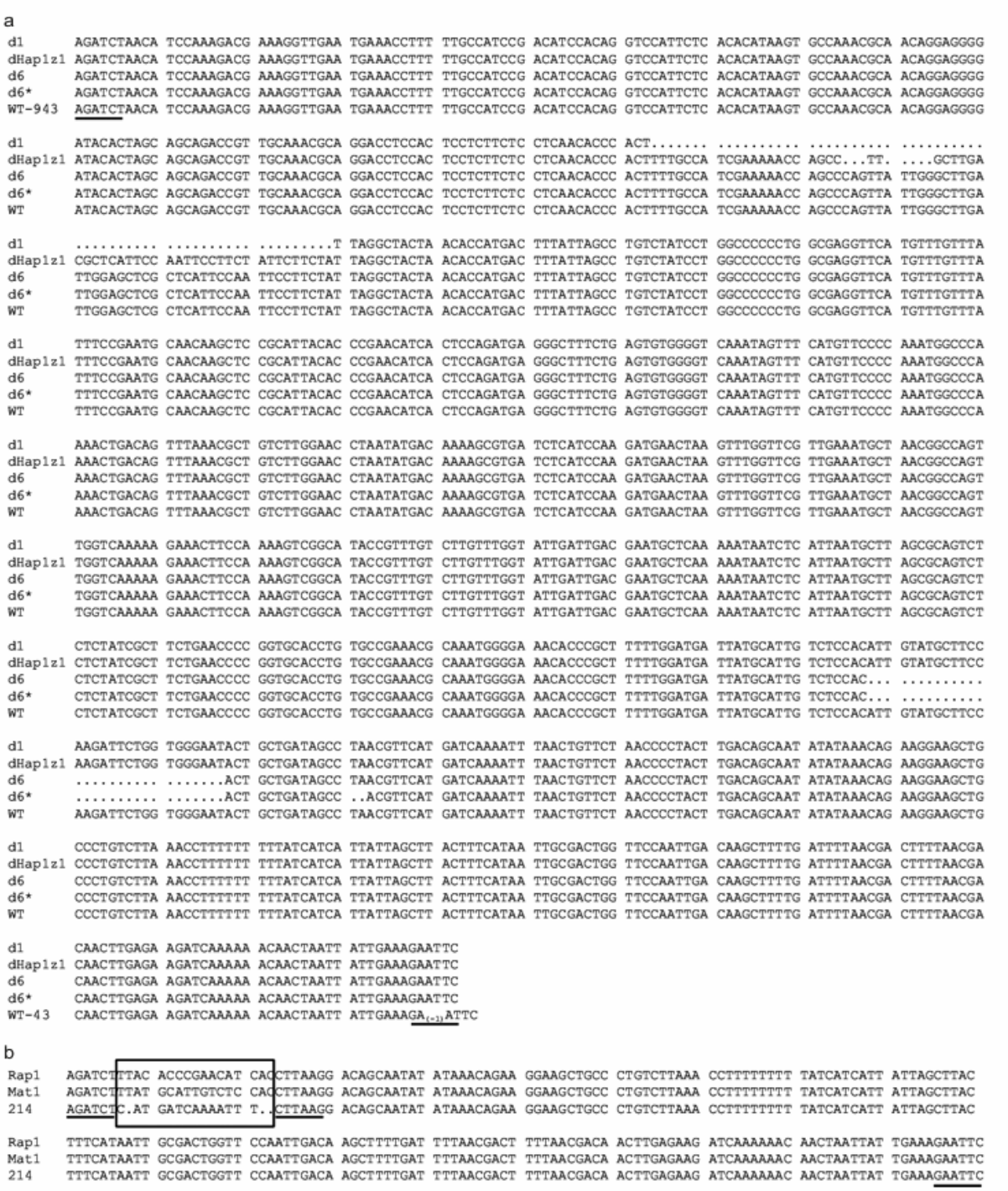



Figure 2.2 Sequence alignment of the *AOX1* wild type (WT) promoter and *AOX1* promoter variants, a) wild type $P_{AOX1}$ (WT), deletion variants (d6, d1), double deletion variants (d6*, dHap2345-1z1), b) synthetic variants AOX176-Rap1 (Rap1), AOX176-Mat1-Mc (Mat1) and AOX176-201-214 (214); underlined: *Eco*RI, *Bgl*II, *Bsp*TI restriction sites, (-1): 3' promoter end, rectangle: cis-acting elements (Rap1, Mat1-Mc, 201-214)

### 2.2.6 Transformation, screening and rescreening

The condensed protocol of Lin-Cereghino et al. (2005) was used for *Pichia pastoris* transformations. Therefore 1 µg of linearized plasmid DNA was prepared for each construct and electroporation was done using following parameters: 1,5 kV, 25 µF and 200 Ω. For regeneration 1 ml of cold sorbitol was added. Transformed cells were regenerated for 2 hours at 28 °C and plated on selective media.
*Pichia pastoris* was transformed with 9 different trypsinogen expression constructs, including 8 different promoter variants (d1, $d1^+$, d6, d6*, dHap2345-1z1, AOX176-Rap1, AOX176-Mat-1MC, AOX176-201-214) and $P_{AOX1}$ wild type. 88 transformants of each construct were picked for screening and expression analysis. Screening was done in 96 deep well plates according to Weis et al. (2004). Clones were grown 60 hours in 300 µl BMD1% media. Induction was performed by addition of 250 µl BMM2 media followed by methanol pulses of 50 µl BMM10 after 12 and 24 hours of methanol induction (in rescreening also at 48, 72 and 96 hours). Samples (50 µl) were taken before induction and subsequently every 24 hours. From samples cells were harvested by centrifugation at 4000 rpm and 4 °C for 10 minutes in an Eppendorf (Hamburg, Deutschland) 5810R centrifuge. Obtained supernatant was used for activity measurements.
Seven clones of each construct representing the whole activity range in the initial primary screen were chosen for rescreening. In the rescreening process, also performed in 96 deep well plates, selected clones were inoculated 4 times each and the standard deviation of the activity values was calculated.

### 2.2.7 Fed batch cultivations in bioreactors

Fedbatch cultivations were performed using the 1,5 l fedbatch-pro® bioreactor system (DASGIP AG, Juelich, Germany). Cultivation parameters were as follows: 28 °C, pH 6. Total batch volume was 650 ml, inoculum volume 50 ml. In the batch phase cells were grown on glucose with a calculated amount of 7,8 $g_C$. The batch phase (phase A) was followed by a fedbatch phase (phase B) in which glucose was fed exponentially with a rate of $1{,}4*e^{(0{,}2*t)}$ $g_C\,h^{-1}$ for 12,25 h. Fedbatch studies using methanol in the production phase (phase C) were characterized by a constant methanol feed of 1,7 $g_C\,h^{-1}$. Fedbatch studies substituting methanol with glucose in phase C were characterized by a constant glucose feed rate of 2,68 $g_C\,h^{-1}$.
Final cultivation experiments were performed using the Infors Multifors system (Infors AG, Bottmingen-Basel, Germany). Using this cultivation system the batch and fedbatch feeds were adjusted according to the $pO_2$ signal of the culture. The total batch volume was 500 ml, inoculum volume 50 ml. To delay product degradation the production temperature was reduced to 20°C. In cultivation studies using methanol for induction the glycerol based batch and fed batch phases were followed by a methanol feed of 0,3 $g_C\,h^{-1}$. Said methanol feed was applied for 20 hours. Subsequently a constant methanol feed of 1,6 $g_C\,h^{-1}$ was set. In fedbatch cultivations using glycerol and glucose in the production phase feed parameters were as follows: 1,6 $g_C\,h^{-1}$ (glycerol) and 1,8 $g_C\,h^{-1}$ (glucose). Note, for cultivations using glucose in the production phase, glucose was also used as substrate during growth phase.

### 2.2.8 Trypsin activity assay

Trypsin activity was measured photometrically in microplates on a Spectramax Plus 384 plate reader (Molecular Devices, Sunnyvale, CA, USA) using N-a-benzoyl-L-arginine p-nitroanilide (BApNA) as a substrate (Hayakawa et al., 1979). Shortly, 5 µl of supernatant were activated by enterokinase for 3 hours at 37°C (0,5 µl enterokinase 0,5 U $ml^{-1}$, 11 µl TEA buffer). 100 µl of BApNA (0,5 mmol $l^{-1}$) were added to start the reaction and kinetics were recorded for 5 minutes at 405 nm. TEA (triethanolamine) buffer for trypsinogen activity assay was used in a concentration of 0,15 mol $l^{-1}$ and a pH of 8,2. For screening and rescreening trypsinogen activity was measured with and without activation by enterokinase. Absolute trypsinogen activity values were calculated by subtracting the activated values from the auto-activated ones.

### 2.2.9 Specific productivity

Specific productivity ($q_P$) (i.e. given in 'units' per gram of biomass and hour) was used to compare the best clones comprising different promoter constructs. It was computed from activity ($dc_p$) and biomass ($x$) data for the time of increasing product concentration ($dt$) according to equation (3.1).

$$q_p = \frac{1}{x} \cdot \frac{dc_p}{dt} \qquad (3.1)$$

### 2.2.10 Protein analysis

Total protein concentration was estimated using the Coomassie (Bradford) Protein Assay Kit of Pierce (Thermo Fisher Scientific Inc., Rockford, IL USA). Electrophoretic separation of protein samples was done using the Agilent (Santa Clara CA, USA) Bioanalyzer 2100. 4 µl of supernatant were denatured for 5 minutes at 95 °C with Agilent sample buffer (2 µl, 3,5% 1 M DTT). Samples were centrifuged and 84 µl of deionized $H_20$ were added. Chip preparation was done according to the Protein 80 Kit manual. Additionally, the Caliper Life Science (Hopkinton, MA 01748 USA) LabChip ® GX II was used to determine protein concentration, following the HT Protein Express LabChip® Assay user instructions.

### 2.2.11 Real Time PCR

For copy number estimation Real Time (RT)-PCR was performed using the ABI PRISM 7300 Real Time PCR System and Power SYBR® Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA), following a procedure described by Abad et al. (2010). The *Pichia pastoris ARG4* gene was used as a reference for quantification. Following primers were used for amplification of the reference gene and inserted construct: ARG4-RTfw (TCCTCCGGTGGCAGTTCTT), ARG4-RTrv (TCCATT GACTCCCGTTTGAG), AOX1-RTfw (GAAGCTGCCCTGTCTTAAACCTT), AOX1-RTrv

(CAAAAGCTTGTCAATTGGAACCA). The PCR program was performed according to Hartner et al. (2008). DNA preparation was adapted from Hoffman and Winston (1987).

## 2.3 Results

### 2.3.1 Rescreening

Trypsinogen activities of the best clones obtained from rescreening after 0, 24, 48, 72 and 96 hours of methanol induction, were normalized by the best obtained wild type $P_{AOX1}$ transformant, which was induced for 96 hours (Table 2.1). 100% wild type activity was equivalent to 16 U $l^{-1}$ $OD^{-1}$. Best results in rescreening were obtained using variants $d1^{+}$ and d6*, showing 218% und 202% relative activity after 72 hours of methanol induction. Interestingly, both constructs showed a decrease in activity after 96 hours of induction, similar to the WT construct. Variant d1 showed in contrast highest activity after 96 hours with 185% relative activity. Following the expression landscape of this construct, maximum values may not have been reached. With regard to expression profiles, variant $d1^{+}$ showed the highest increase in activity over a short period of time. For this construct, the relative trypsinogen activity increased from 7 to 172% within the first 24 h of induction. Compared to the WT strain, expression was increased more than 2-fold at this time point. The analyses for the synthetic construct AOX176-Rap1 showed the possibility to produce trypsinogen up to the end of the glucose batch, with the best clone reaching 77% relative activity. Interestingly, this clone was also methanol inducible, reaching 164% activity after 72 hours of methanol induction.

Due to the strong derepression effect of construct AOX176-Rap1, indicating an altered regulation pattern, four AOX176-Rap1 clones (G3, A7, C8, B3) of different expression strengths were selected for further studies. Excluded from further analyses were the following constructs: d1, d6, dHap2345-1z1, AOX176-Mat1-Mc, AOX176-201-214. Promoter construct d1 was excluded from further analysis due to its lower specific productivity compared to construct $d1^{+}$. Variant d6 was not considered because of its similarity in expression profile to variant d6*. Variant dHap2345-1z1 was excluded due to low activity under both conditions, methanol induction as well as glucose derepression. The short synthetic constructs AOX176-Mat1-Mc and AOX176-201-214 were not selected as both constructs were outranked in expression by the promoter AOX176-Rap1.

Table 2.1 Relative promoter comparison, activities of the best clones from rescreening are given in relation to the value of the WT-strain at 96 hours (i.e. 100%) and the time of methanol induction - resp. time 0 h at the end of glucose batch; standard deviations were calculated using data from 4 different wells

| promoter | 0 h | 24 h | 48 h | 72 h | 96 h |
|---|---|---|---|---|---|
| WT | 8 ± 8 | 86 ± 14 | 127 ± 14 | 136 ± 15 | 100 ± 21 |
| d1 | 4 ± 9 | 83 ± 11 | 86 ± 10 | 152 ± 15 | 185 ± 45 |
| $d1^{+}$ | 7 ± 4 | 172 ± 55 | 189 ± 61 | 218 ± 23 | 167 ± 40 |
| d6* | -6 ± 10 | 124 ± 30 | 141 ± 39 | 202 ± 44 | 123 ± 31 |
| Rap1_G3 | 77 ± 16 | 96 ± 22 | 100 ± 29 | 164 ± 24 | 80 ± 14 |
| Rap1_A7 | 52 ± 10 | 46 ± 35 | 44 ± 17 | -5 ± 51 | 49 ± 6 |
| Rap1_C8 | 39 ± 8 | 36 ± 8 | 49 ± 15 | 172 ± 64 | 49 ± 5 |
| Rap1_B3 | 46 ± 8 | 68 ± 15 | 96 ± 17 | 80 ± 45 | 61 ± 5 |

### 2.3.2 Copy number estimation

To analyze a possible correlation between expression level and the number of integrated expression cassettes (copy number), the copy numbers were determined by RT-PCR for the best clones of each construct. Additionally the clones A7, C8 and B3 of construct AOX176-Rap1 were analyzed. Interestingly, it was shown that all long mutation and deletion $P_{AOX1}$ variants selected through screening and rescreening contained only 1 copy of the integration cassette. In contrast, construct AOX176-Rap1 seemed to perform best with more integrated copies, with clone G3 having over 4 copies of the expression cassette integrated into the *P. pastoris* genome. The other AOX176-Rap1 strains A7, C8 and B3 contained 3-4, 2 and again 4 copies, respectively. Strains and copy numbers are shown in Table 2.2.

Table 2.2 *P. pastoris* trypsinogen expression strains with corresponding copy numbers

| construct | copy number |
|---|---|
| WT | 1 |
| d1 | 1 |
| d1+ | 1 |
| d6* | 1 |
| Rap1-G3 | >4 |
| Rap1-A7 | 3-4 |
| Rap1-C8 | 2 |
| Rap1-B3 | 4 |

### 2.3.3 Fedbatch cultivations in bioreactors

To confirm the data of the 96 deep well screening and rescreening, parallel fedbatch cultivations were performed in bioreactors. Specific productivity, activity increase, production time and biomass growth were calculated for all strains or promoter constructs. However, it has to be considered that specific productivity decreases with longer production times and hence do not give a full picture of the abilities of a certain promoter.

#### 2.3.3.1 Fedbatch with methanol

Trypsinogen expression of the best transformants of the promoter variants $d1^+$, d6* and AOX176-Rap1 was analyzed employing methanol induced fedbatch cultivations. In the cultivations a maximal activity of 150 U $l^{-1}$ was reached for the constructs $d1^+$ and AOX176-Rap1, 55 hours after process start (Figure 2.3). For both constructs a maximal protein titer of approximately 700 mg $l^{-1}$ protein was determined by protein chip analysis. Subsequently the expected degradation of trypsinogen started, which led to reduced absolute trypsinogen activity values after approximately 60 hours process time. After 100 hours of cultivation there was almost 100% activity without prior activation by enterokinase.
Variant d6* showed, in contrast to the tightly repressed promoter $d1^+$, a derepression effect and produced 15 U $l^{-1}$ active enzyme until the end of growth in fedbatch with glucose. Compared to variant $d1^+$ expression was improved 3-fold at this point of time. Regulated over glucose depletion, variant d6* was still methanol inducible, though weaker. This resulted in a calculated specific productivity of 0,03 U $g^{-1}$ $h^{-1}$. The effect of induction via derepression was enhanced several fold by variant AOX176-Rap1 expressing 45 U $l^{-1}$ of active enzyme until the end of fedbatch phase B. In addition, variant AOX176-Rap1 was also strongly methanol inducible reaching values similar to the $d1^+$ based transformant. Variant $d1^+$ and AOX176-Rap1 reached comparative specific productivity levels of 0,07 and 0,06 U $g^{-1}$ $h^{-1}$, but variant $d1^+$ was highest inducible by methanol in a short period of time (28 h) confirming the results obtained in rescreening.

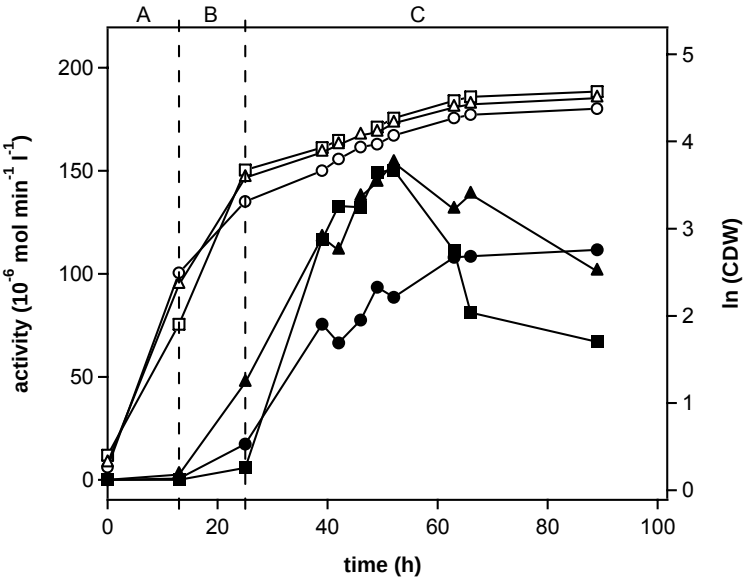



Figure 2.3 Expression of porcine trypsinogen using *AOX1* promoter variants, A – growth in batch with glucose, B – growth in fedbatch with glucose, C – production phase in fedbatch with methanol, open symbols – logarithm of cell dry weight, closed symbols – trypsinogen activity, promoters: $d1^+$ (squares), d6* (circles), AOX176-Rap1 (triangles)

Protein chip analysis of the variants d6* and AOX176-Rap1 revealed a clearly visible trypsinogen/trypsin band at a size of approximately 24,5 kDa, but only up to a certain point of time (Figure 2.4). In correlation with the results from activity measurements, product degradation started after a certain process time. Once trypsinogen was cleaved an autocatalytic cascade started seemingly. According to protein chip analysis we assume that produced trypsin subsequently cleaved all its zymogen to trypsin (23,6 kDa, β -trypsin), which was then degraded into its 2 α-trypsin chains (~13 and ~10 kDa) or even further fragmented.

The start point of autoproteolytic degradation differed depending on the promoter used. Employment of the low producer d6*, with a corresponding maximal activity of 120 U $l^{-1}$, delayed the start point of degradation significantly. In fact, the production time was extended almost 2-fold for this variant. Starting in fedbatch phase B, a total production time of 57 hours was reached. In contrast, using the construct AOX176-Rap1 protein degradation started already after 60 hours process time, with a total production time of 39 hours. However, though degradation started earlier employing this construct, more active protein (700 mg $l^{-1}$) was produced compared to the construct d6* (400 mg $l^{-1}$).

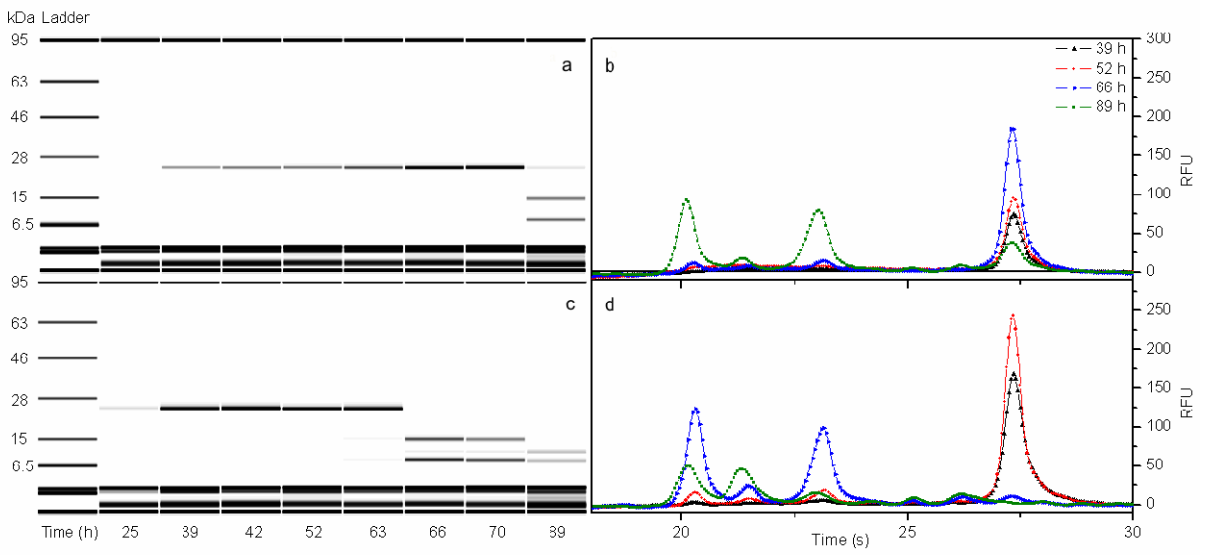



Figure 2.4 Protein chip analyses of the methanol induced fedbatch cultures with the promoters d6* (a, b) and AOX176-Rap1 (c, d); a, c gel images visualizing promoter dependent trypsinogen degradation upon a certain production time; b, d corresponding fluorescence detection based chromatograms used for protein concentration estimation at different cultivation times, RFU: relative fluorescence units

### 2.3.3.2 Fedbatch with glucose

To further analyze the strong derepression effect of the synthetic construct AOX176-Rap1, a second round of fedbatch cultivations was performed employing the strains AOX176-Rap1 G3, A7, C8, B3. Here, the intention was to show the ability of such constructs to express significant amounts of protein via an inducible glucose depletion/derepression based process. Within this series strain AOX176-Rap1 G3 produced up to 170 U $l^{-1}$ active enzyme at maximum level (Figure 2.5). With a production time of 53 hours a specific productivity of 0,03 U $g^{-1}$ $h^{-1}$ was calculated for this strain. For comparison AOX176-Rap1 G3 induced by methanol reached 150 U $l^{-1}$ and 0,06 U $g^{-1}$ $h^{-1}$, respectively. Also the other AOX176-Rap1 constructs produced significant amounts of active enzyme. Strains A7, B3 and C8 showed maximal activities of 145, 160 and 120 U $l^{-1}$ with corresponding productivities of 0,02, 0,03 and 0,02 U $g^{-1}$ $h^{-1}$.

Product degradation started for all strains between 50-70 hours process time in the following order: G3, A7/B3 and C8 (data not shown). In accordance to the previous results the degradation started first for the high production strain G3 (170 U $l^{-1}$) and last in case of the low producing transformant C8 (120 U $l^{-1}$). Again, after about 100 h process time almost all trypsinogen was activated to trypsin.

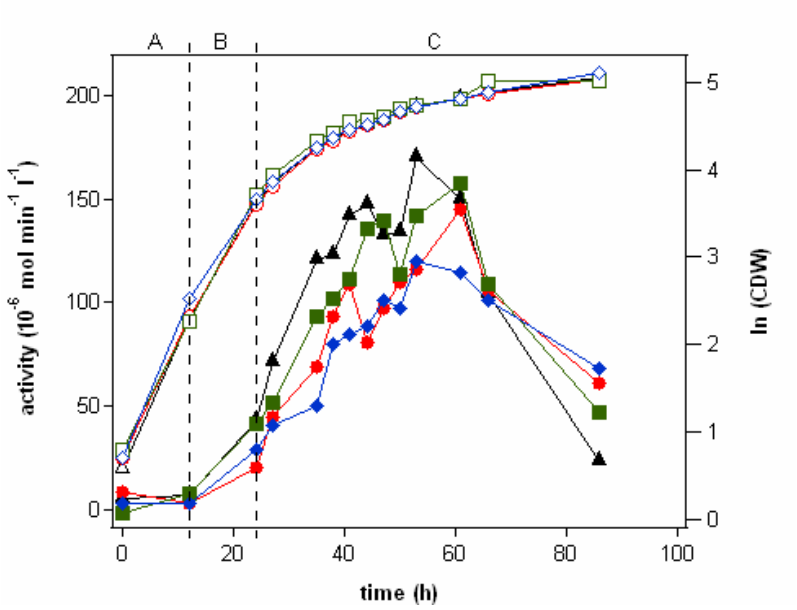



Figure 2.5 Expression of porcine trypsinogen using synthetic promoters, A – growth in batch with glucose, B – growth in fedbatch with glucose, C – production phase in fedbatch with glucose, open symbols – logarithm of cell dry weight, closed symbols – trypsinogen activity, AOX176-Rap1 variants: G3 (triangles), A7 (circles), B3 (squares), C8 (rhomboids)

#### 2.3.3.3 Final fedbatch comparison

To confirm the obtained results and for further analysis of product formation and its degradation a final cultivation series using different substrates for induction was performed. In agreement with the results described above different production windows were found on account of different promoters (Figure 2.6). Interestingly, construct d1$^+$ increased activity per biomass up to 135% reaching 4,9 U $g_X^{-1}$. In addition, promoter d1$^+$ produced already some trypsinogen before induction (1,3 U $g_X^{-1}$). This observation was also confirmed by protein chip analysis showing 25 mg $l^{-1}$ protein at this time point. Surprisingly, employing the promoter variant d6* did not only result in a prolonged production phase, but also no degradation demonstrated by increasing absolute trypsinogen activity over 110 hours of cultivation time (Figure 6). Employing construct d6* a specific activity of 1,6 U $g_X^{-1}$ was reached at maximum level. For comparison the $P_{AOX1}$ wild type transformant reached a value of 3,6 U $g_X^{-1}$. Cultivations substituting methanol with glucose or glycerol employing promoter variant AOX176-Rap1 (Rap1_Gluc and Rap1_Glyc) showed maximum specific activities between 2,5-2,9 U $g_X^{-1}$, which matched with about 75% of the methanol induced WT value. In detail, Rap1_Gluc reached 2,5 and Rap1_Glyc 2,9 U $g_X^{-1}$. Expression profiles of Rap1_Gluc and Rap1_Glyc showed delayed product degradation starting between 60 - 85 hours of process time. In comparison, for the constructs WT and d1$^+$ degradation started already between 40 - 60 hours of cultivation.

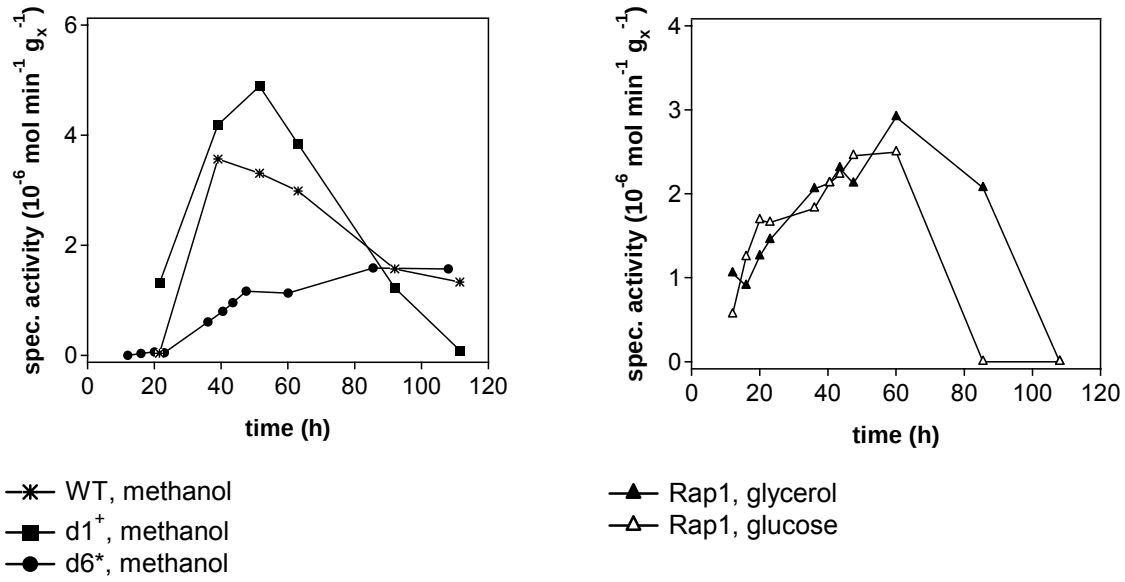



Figure 2.6 Specific trypsinogen activities in fedbatch cultures, left - production phase with methanol (WT, $d1^+$, d6*), right - production phase with glycerol or glucose (AOX176-Rap1)

In addition, protein yields were calculated (mg $g_X^{-1}$). In agreement to already reported results, variant $d1^+$ performed best reaching a value of 3,99 mg $g_X^{-1}$, while the low producer d6* reached only 1,95 mg $g_X^{-1}$. Strategies employing AOX176-Rap1 on glucose or glycerol reached 2,31 and 2,11 mg $g_X^{-1}$, respectively. Protein yields are summarized in Figure 2.7.

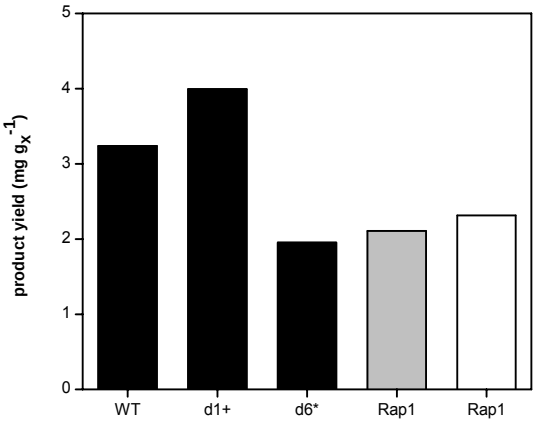



Figure 2.7 Product yields obtained in fedbatch cultures with different promoters (WT, $d1^+$, d6*, AOX176-Rap1) applying different carbon sources in the production phase: methanol (black), glycerol (grey), glucose (white), protein concentration determination: Caliper capillary electrophoresis

## 2.4 Discussion

In this study we have shown the effect of differently regulated, laboratory generated, full length and short synthetic promoter variants on the expression of porcine trypsinogen. Briefly, protein titer could be improved up to 135% by using a stronger *AOX1* promoter variant. In addition, derepressed promoter variants allowed the production of trypsinogen substituting methanol induction by a simple derepression process. Thus, the different regulatory features of construct AOX176-Rap1 now enable to avoid toxic and flammable methanol in a still tightly regulated cultivation process. Moreover, it was possible to shift the time-span of production and proteolytic product degradation.

Evaluating the screening procedure, it can be said that good correlation was found between the results obtained in rescreening and fedbatch cultivations. Product degradation was already indicated in micro scale, showing decreased activity values at 96 hours of induction. Still, it is possible that the observed effect was enhanced by, or just a result of dilution through substrate addition. Interestingly, all expression strains based on full length promoter variants and giving the best results in the screening experiments contained only 1 single integrated copy of the expression cassette. More copies seemed to influence trypsinogen expression negatively, and were therefore not identified as hit during the screening procedure. Low copy numbers might compensate for toxicity effects of trypsinogen. Only the short synthetic construct AOX176-Rap1 showed a positive correlation between copy number and activity with the best clone having over 4 copies. These results might be explained by the fact that 1 copy of construct AOX176-Rap1 expresses a much lower level of trypsinogen compared to the other constructs, hence is far under the level of toxicity influence. This is in agreement with the results of the stronger full length promoter variants.

Rescreening results from deep well plate cultivations were confirmed by fedbatch cultivations. Generally, up-scaling resulted in an approximate 10-fold improvement in trypsinogen production per volume. In methanol induced fedbatch cultivations the expression and degradation profile changed depending on the promoter used. As a result of the altered regulation of construct AOX176-Rap1, finally 10 times more trypsinogen was produced until the end fedbatch phase B (see Figure 3). The previously described d6* promoter variant showed induction by derepression with a 3-fold increase at this point of time. Furthermore, employing this variant allowed to delay or even abolish product degradation. This might be due to the lower specific productivity of the d6* based strain (on methanol), assuming trypsinogen degradation to be at least partly titer connected. Although variant d6* was outranked in final yield by other promoter constructs, employing this variant significantly improved product homogeneity and thereby quality in regard to production time. Surprisingly, with promoter variant $d1^+$ some active protein was produced until the end of fedbatch phase B (see Figure 6), if grown on glycerol. Said effect was not found in cultivations using glucose for growth. However, further experiments are needed to confirm or disprove this particular characteristic of promoter construct $d1^+$. Meanwhile it can only be speculated that different carbon sources, here glucose and glycerol, do effect the regulation of $d1^+$ differently. The strong derepression effect of the strains based on construct AOX176-Rap1 was verified in fedbatch studies using both glucose and glycerol as feed substrates during production phase. It was shown that about 75% of the induced WT level can be reached via a regulated glucose depletion mechanism. In addition, employing AOX176-Rap1 in glucose or glycerol driven cultivations led to

delayed product degradation. Similar to variant d6* lower productivities were found, indicating decelerated specific productivity to be critical for high product quality.
Promoter variant AOX176-Mat1-Mc was not further studied in bioreactor experiments since promoter variant $d1^+$ showed to be better in methanol induced expression and promoter variant AOX176-Rap1 under derepression conditions. Nevertheless the Mat1-Mc sequence fused to a core promoter led to a remarkable response upon methanol induction (~50% of $d1^+$ driven expression). Also multiple Mxr1p binding sites fused to a core promoter might be useful to improve methanol induced expression by small synthetic promoters. However, though Mxr1p seems to be a key factor the full mechanism of methanol induced regulation is not yet understood. DNA footprinting studies involving such small synthetic promoters in future could facilitate the clarification of *AOX1* promoter regulation which might involve several additional factors yet to be identified. Although a variety of different synthetic promoter variants have proven their ability to successfully express recombinant trypsinogen, there is also room for improvement addressing other bottlenecks. Endogenous proteases might play an important role in the activation process. Gel images obtained from capillary electrophoresis showed that once the proteolytic cascade started, trypsin is further fragmented rapidly. Active protease might be deleterious to the cell and cause a production stop. An approach similar to Hanquier et al. (2003) with mutated protease cleavage sites could eliminate the problem of degradation and might enhance protein production. Other bottlenecks, like transport or folding, can also be limiting factors. Co-expression of helper proteins or chaperones like PDI (protein disulfide isomerase) might facilitate expression, as well.

### 2.4.1 Conclusions

Well-studied synthetic promoters that vary in their strength and regulation mechanism offer new opportunities in recombinant protein expression. As for the expression of porcine trypsinogen using the strong WT *AOX1* promoter, the single copy clones performed best, space for expression optimization using the same regulatory system is limited. Said limitation can be overcome by changing the promoter itself, which was demonstrated in this study by employing novel synthetic promoters to optimize the expression of the problematic reporter trypsinogen.

## 2.5 Acknowledgments

We want to express our gratitude to VTU Technology for scientific and financial support. Furthermore, the research was partially supported by the Swiss Academy of Engineering Sciences (SATW) in scope of the "Industrial Biotechnology" program (TK-07-19).

## 2.6 References


Abad, S., Kitz, K., Hörmann, A., Schreiner, U., Hartner, F.S., Glieder, A., Real-time PCR- based determination of gene copy numbers in Pichia pastoris. Biotechnol J 5, 413-420.

Chance, R., Di Marchi, R., Frank, B., Shields, J.E., (1997) Process for preparing insulin analogs. US5700662.

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, 3376-85.

Cregg, J.M., Cereghino, J.L., Shi, J., Higgins, D.R., (2000) Recombinant protein expression in Pichia pastoris. Mol Biotechnol 16, 23-52.

Cregg, J.M., Vedvick, T.S., Raschke, W.C., (1993) Recent advances in the expression of foreign genes in Pichia pastoris. Biotechnology (N Y) 11, 905-10.

Frank, B., Prouty, W., Richard, H., Walden, M., (1995) Process for transforming a human insulin precursor to human insulin. US5457066.

Hanquier, J., Sorlet, Y., Desplancq, D., Baroche, L., Ebtinger, M., Lefevre, J.F., Pattus, F., Hershberger, C.L., Vertes, A.A., (2003) A single mutation in the activation site of bovine trypsinogen enhances its accumulation in the fermentation broth of the yeast Pichia pastoris. Appl Environ Microbiol 69, 1108-13.

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.

Hartner, F.S., Glieder, A., (2005) Mutant AOX1 promoters. WO2006089329.

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.

Hayakawa, T., Kondo, T., Yamazaki, Y., Ito, K., Iinuma, Y., Okumura, N., Sakakibara, A., Naruse, S., Toda, Y., Aoki, I., (1979) Diagnostic significance of serum immunoreactive trypsin in pancreatic diseases. Nippon Shokakibyo Gakkai Zasshi 76, 1513-21.

Hoffman, C.S., Winston, F., (1987) A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli. Gene 57, 267-72.

Inan, M., Meagher, M.M., Benson, A.K., (2004) Alcohol oxidase 1 regulatory nucleotide sequences for heterologous gene expression in yeast. US2004137591.

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.

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, 883-97.

Lin-Cereghino, J., Wong, W.W., Xiong, S., Giang, W., Luong, L.T., Vu, J., Johnson, S.D., Lin-Cereghino, G.P., (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. Biotechniques 38, 44-48.

Stroud, R.M., Kossiakoff, A.A., Chambers, J.L., (1977) Mechanisms of zymogen activation. Annu Rev Biophys Bioeng 6, 177-93.

Verdier, J.M., (1990) Regulatory DNA-binding proteins in yeast: an overview. Yeast 6, 271-97.

Wang, W., Malcolm, B.A., (1999) 2-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. Biotechniques 26, 680-682.

Weis, R., Luiten, R., Skranc, W., Schwab, H., Wubbolts, M., Glieder, A., (2004) Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena. FEMS Yeast Res 5, 179-89.

Werten, M.W., Van Den Bosch, T.J., Wind, R.D., Mooibroek, H., De Wolf, F.A., (1999) High-yield secretion of recombinant gelatins by Pichia pastoris. Yeast 15, 1087-96.

Xiong, A.S., Yao, Q.H., Peng, R.H., Zhang, Z., Xu, F., Liu, J.G., Han, P.L., Chen, J.M., (2006) High level expression of a synthetic gene encoding Peniophora lycii phytase in methylotrophic yeast Pichia pastoris. Appl Microbiol Biotechnol 72, 1039-47.

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, 1271-82.

## 2.7 Annex

Porcine trypsinogen containing *E. coli* and *P. pastoris* strains were used according to the Diploma thesis of Thomas Zuellig.

# *3* Heterologous expression of soluble plant epoxide hydrolases in *Pichia pastoris*

This work was based on the Diploma thesis of Stefan Ertl.
The manuscript is prepared for publication in “Protein Expression and Purification” (Elsevier).

## 3.1 Introduction

Epoxide hydrolases are ubiquitous enzymes found in many kingdoms (Newman et al., 2005). Up to now 7 different classes of epoxide hydrolases are reported including soluble (sEH), microsomal (mEH) and cholesterol (ChEH) epoxide hydrolase. Catalyzing the conversion of epoxides to diols by addition of water, epoxide hydrolases are highly selective biocatalysts with varying stereochemistry depending on the natural host (Morisseau et al., 2000). For example, while human and mouse sEH are both enantioselective for S,S-alkylphenyloxiranes, the plant *Arabidopsis thaliana* sEH is selective for the R,R-enantiomer (Williamson, 2000). In cascade reactions together with monooxygenases epoxide hydrolases enable stereoselective dihydroxylations of unsaturated compounds (Li et al., 2001).
In comparison to mammalian epoxide hydrolases, still little is known about plant sEHs (Morisseau et al., 2000). Plant EHs are found primarily as soluble enzymes in the cytosol, but also in the glyoxysomes. Only the small group of microsomal plant EHs seems to be membrane bound (Blee and Schuber, 1992a). The most striking difference between mammalian and plant sEHs is the 30% shorter N-terminus of the plant proteins (Morisseau et al., 2000). Soluble plant EHs belong to the a/ß-hydrolase fold enzymes and show an approximate molecular weight of 35 kDA. Often multiple isoforms are found (Newman et al., 2005).
sEHs are suggested to play a role in the general defence system of a plant by contributing to the synthesis of cutin monomers and natural antifungal compounds. The biosynthetic pathway of cutin involves sEHs that catalyze the hydration of oxylipins (e.g. vernolic acid) to their corresponding diols (Newman et al., 2005). Generally, plant sEHs seem to prefer trans- over cis-epoxides of sterically hindered substrates like stilbene oxides. Epoxy fatty acids derived from stearic and linoleic acid are the preferred endogenous substrates for the few plant epoxide hydrolases characterized so far (Blee and Schuber, 1992b; Morisseau et al., 2000).
Up to now plant sEHs were mainly isolated from plants or heterologously expressed by *Escherichia coli* or baculovirus systems (Edqvist and Farbos, 2003; Elfström and Widersten, 2005; Morisseau et al., 2000). Known as a simple microbial expression system, for strong overexpression of recombinant proteins and the feasibility of high cell density cultures, *Pichia pastoris* provides an interesting alternative, which has already been reported suitable for plant sEH production (Bellevik et al., 2002a; Bellevik et al., 2002b; Cregg et al., 2000).
In this study we show the expression of soluble epoxide hydrolases of the 3 plants *Solanum tuberosum* (StEH1), *Glycine max* (GmEH1) and *Euphorbia lagascae* (ElEH1) in *Pichia pastoris* to provide enzyme for further biocatalytic characterization and application as well as for a direct comparison of expression and activity of these 3 plant enzymes.

Though homologous StEH1, GmEH1 and ElEH1 show large variations on the amino acid sequence level (identity <60%). GmEH1 possesses an N-terminal extension, which is assumed to play part in the enzymes natural dimeric occurrence. However, there is also indication that the enzyme is monomeric in solution (Arahira et al., 2000). Lacking such an N-terminal extension StEH1 and ElEH1 are monomers. In agreement to the assumed natural sub-cellular localization of the sEHs in the glyoxysomes, the 3 enzymes show a putative C-terminal glyoxysomal targeting sequence (Q/N/E-K-F). However, it is unknown if these putative plant type 1 peroxisomal targeting signal sequences (PTS1) also allow PTS1 mediated entry into *Pichia pastoris* peroxisomes. Targeting to *P. pastoris* peroxisomes might be beneficial (Bellevik et al., 2002a), as peroxisomes do increase significantly upon methanol induction; however was not specifically studied so far.

## 3.2 Materials and methods

### 3.2.1 Strains and expression cassettes

The soluble epoxide hydrolases from *Solanum tuberosum* (accession no. AAA81892), *Glycine max* (accession no. CAA55294) and *Euphorbia lagascae* (accession no. AAO27849) were codon optimized using the Gene Designer software (DNA2.0, Menlo Park, CA, USA) applying *Pichia pastoris* high methanol codon usage (Abad et al., 2010a). In addition, *Solanum tuberosum* sEH1 was codon optimized for expression in *E. coli* applying the class II codon usage of highly expressed genes (www.kazusa.or.jp/codon). The *E. coli* optimized StEH1 gene was cloned into plasmid pMS470Δ8 (*Nde*I/*Sph*I) and the *E. coli* strain BL21 was used for expression. *P. pastoris* was transformed with gel purified, linear PCR cassettes (Liu et al., 2008). The generated linear PCR cassettes were either directly used for *Pichia pastoris* transformation or prior to transformation sub cloned using the Fermentas GeneJET™ PCR Cloning Kit (Burlington, Ontario, Canada). According to Liu et al. (2008) these linear expression cassettes made by PCR contained 3 fragments. Fragment 1 contained the promoter sequence for homologous integration into the genome and to drive the expression of the gene of interest, fragment 2 contained the gene of interest and fragment 3 contained the terminator and selection cassette. Following linker sequences including a 3' Kozak sequence were used for overlap extension (OE)-PCR of the 3 individual fragments: linker 1: ctaggtacttcgaaacgaggacttcacg, linker 2: gtcagatagcgaggtcactcagtc. To facilitate purification from *P. pastoris* cell lysate an N-terminal His-tag (6x His) was attached to the intracellular expression cassettes. In case of the StEH1 containing extracellular expression cassette the C-terminal putative peroxisomal targeting sequence Q-K-F was removed. For protein secretion a codon optimized version of the *S. cerevisiae* mating factor α signal sequence was used (supplementary Figure 3.6).

The *AOX1* promoter ($P_{AOX1}$, 937 bp) and terminator sequences were synthesized by GenScript (Piscataway, NJ, USA) based on the sequence information from GenBank (accession no. U96967). Allowing selection against the antibiotic zeocin, the selection cassette was based on the Ble gene from *Streptoalloteichus hindusdanus* (accession no. A31898), which was optimized for both, expression in *E. coli* and *P. pastoris* (Leto Software, Entelechon Corp.). The zeocin resistance gene was under control of

an 860 bp long *Saccharomyces cerevisiae ADH1* promoter and 300 bp *ADH1* terminator fragment. For application of the selection cassette in *E. coli* a synthetic prokaryotic consensus promoter ($P_{EM72}$) was embedded between the eukaryotic promoter and the start of the resistance gene (Ruth et al., 2010). Gene expression in *P. pastoris* was performed using the strain CBS7435 $Mut^S$, which was constructed in our laboratory by an *AOX1* gene knock-out of strain CBS7435 (Näätsaari et al., article under construction). *Pichia pastoris* strain CBS7435 was obtained from CBS fungal biodiversity center. Strain CBS7435 $Mut^S$ was deposited in the strain collection of IMBT TU-Graz under BT3445. Subcloning was performed using *E. coli* DH5a-$T1^R$ or TOP10F' cells (Invitrogen, Carlsbad, CA, USA).

### 3.2.2 AOX1 promoter variants

The 3' end of the promoter was modified by introducing an *Eco*RI site (similar to accession no. DM049086). However, sequence numeration was done in consensus to the natural *AOX1* promoter in upstream direction, starting at the start codon of the *AOX1* open reading frame (natural $P_{AOX1}$ 3' end: 5'-GAAACG$_{(-1)}$ATG-3' / new $P_{AOX1}$ 3' end: 5'-GAAA*GA$_{(-1)}$ATTC*ATG-3', italic: *Eco*RI site, underlined: *AOX1* start codon).

Using $P_{AOX1}$ as template, the promoter variants $d1^+$ and d6*, described by Hartner et al. (2008), were rebuilt by 2-step site-directed mutagenesis according to Wang and Malcolm (1999). Variant $d1^+$ consisted of a deletion of 66 bp between position -711 and -776 and an additional multiplication of 14 bases called 201-214, located at position -189. Variant d6* was made by a deletion of 30 bases at position -223 (d6) and an additional deletion of 2 base pairs (*), located 3' of d6 ($\Delta TA_{-208-209}$).

### 3.2.3 Chemicals and media

Oligonucleotide primers were obtained from Eurogentec (Seraing, Belgium). For plasmid isolation the GeneJET™ Plasmid Miniprep Kit of Fermentas was used. PCR fragment purification was done with the Wizard® SV Gel and PCR Clean-Up System of Promega (Madison, WI, USA). Chemicals were purchased from Carl Roth (Karlsruhe, Germany) and Sigma-Aldrich (St Louis, MO, USA).

*E. coli* complex media (LB-Agar) was purchased from Roth. Complex *Pichia pastoris* media (YPD Broth/Agar) was purchased from LabM (Bury, UK). For *E. coli* antibiotic selection a concentration of 25 µg $ml^{-1}$ zeocin was used, while for *P. pastoris* a concentration of 100 µg $ml^{-1}$ was applied. Buffered minimal media BMD1%, BMM2 and BMM10 was used according to Weis et al. (2004).

Modified basal salt medium (Invitrogen) contained per liter 40 g glycerol, 0,47 g $CaSO_4.2H_2O$, 9,1 g $K_2SO_4$, 2,07 g KOH, 7,5 g $MgSO_4.7H_2O$, 0,22 g NaCl, 0,6 g EDTA disodium salt and 13,4 ml $H_3PO_4$ (85%). Additionally 0,00087 g biotin and 4,35 ml $PTM_1$ mineral salt solution were added per liter after sterilisation. $PTM_1$ mineral salt solution was adopted from Hartner et al. (2008). 12 ml $PTM_1$ salt solution and 0,0024 g biotin were added to the feed solutions. The methanol feed solution contained 99% methanol, the glycerol feed 60% (w/w) glycerol.

### 3.2.4 Transformation

The condensed protocol of Lin-Cereghino et al. (2005) was used for transformation of *Pichia pastoris*. 1-2 µg of the linear PCR product were used for transformation. Electroporation was done using following parameters: 1,5 kV, 25 µF and 200 Ω. For regeneration 1 ml of ice cold sorbitol/YPD (1:1) was added. Transformed cells were regenerated for 2 hours at 28 °C and plated on selective media.

### 3.2.5 Cultivation

Screening was done in 96 deep well plates according to Weis et al. (2004). Clones were grown 60 hours in 300 µl BMD1% media. Induction was started by addition of 250 µl BMM2 (0 hours of induction time) followed by 50 µl pulses of BMM10 after 12, 24, 48 and 72 hours of induction time. In the rescreening process, also performed in 96 deep well plates, selected clones were inoculated 3 times each. 2 l shake flask experiments were performed using 135 ml BMD1% inoculated to an OD of 0,05 - 0,1. After 60 hours of growth the cultures were induced by addition of 10 ml BMM10, which was followed by pulses of 1,5 ml methanol at 12, 24, 48 and 72 hours of induction time.
Bioreactor cultivations were performed using the Infors Multifors system (Infors AG, Bottmingen-Basel, Germany). Total batch volume was 500 ml, inoculum volume 50 ml. During the whole cultivation pH 5 was applied. In the methanol fed batch phase (production phase) the temperature was lowered from 28°C to 24°C. Fedbatch feeds were adjusted according to the $pO_2$ signal of the culture. The glycerol feeds in fed batch phase 1 were 15,14 and 16,5 g $l^{-1}$ $h^{-1}$ for clone $STEH_{WT}$ and $StEH_{d1+}$, respectively. For clone $STEH_{WT}$ the methanol feed varied between 1,9 - 2,5 ml $h^{-1}$, while for clone $StEH_{d1+}$ a varying methanol feed of 2,8 - 5,7 ml $h^{-1}$ was applied.

### 3.2.6 Conversions

Using trans-stilbene oxide (TSO, stock solution, 100 mM in acetonitrile) as substrate whole cell conversions were done as follows. In deep well screening and rescreening cells were harvested by centrifugation at 4°C and 4000 rpm. The obtained pellet was resupended in 450 µl sodium phosphate buffer (pH 6,8, 100 mM). 50 µl of TSO stock solution were added and the conversions were done for 10 minutes at 28°C and 320 rpm. Employing 9,10-epoxystearic acid as substrate (ESA, stock solution, 10 mM in ethanol) conversions were done for 20 hours. ESA was employed as alternative TLC screening substrate for soy bean and spurge sEHs as higher specific activities were indicated in literature if compared to TSO. After conversion of TSO or ESA 100 µl of cell suspension were transferred into a microplate and 100 µl of acetonitrile were added to stop the reaction. Cells were again centrifuged at 4°C and 4000 rpm and the supernatant was used to measure the conversion rate (TLC, HPLC). Styrene oxide (SO) was in addition used as screening substrate. Being coupled to a photometric assay (red assay) using this substrate allowed rapid screening with high throughput. In this case whole cell conversions were done as follows. In deep well screening and rescreening cells were harvested as above and washed using 800 µl of PBS washing buffer (140 mM NaCl, 10 mM $Na_2HPO_4.2H_2O$, 2,7 mM KCl, 1,8 mM $KH_2PO_4$, pH 7,2). After centrifugation cells were resupended in 180 µl of sodium phosphate buffer (pH 7,2, 100 mM). For conversion 20 µl of SO stock solution (100 mM in acetonitrile) were added

to 20 µl of the cell suspension and 150 µl of PBS buffer. Conversions were performed for 30 minutes at 28°C.
For the activity determination of cells from rescreening and bioreactor experiments TSO was used as substrate. For conversions cells were diluted to an OD of 0,1, 10 and 30 for clones expressing potato, soy bean and spurge sEH, respectively. Conversions were started by addition of 50 µl of TSO stock solution to 500 µl of cell suspension. Conversions were performed in 1,5 ml microtubes at 800 rpm and 28°C and stopped as described above. Conversion times were 10, 90 and 270 minutes for cells expressing potato, soy bean and spurge sEH. In addition, the protocol was adapted for microplates. There, 110 µl of cell suspension with OD 0,5 (for clones expressing StEH1) was used for conversion at 28°C. 12 µl of substrate stock were used to start the reaction and 120 µl acetonitrile added to stop it. Conversions employing cell lysate or purified enzyme were performed using 20 µg of total protein.
To determine the temperature optimum TSO was applied in a final concentration of 1 mM. 50 µl of the purified enzyme were used for conversion in a final volume of 1 ml. The sample volume for product concentration determination was 100 µl. The same applied for the determination of the pH optimum for which sodium phosphate buffer was used between pH 5,3 - 8 and Tris-HCl buffer between pH 8 - 9.

### 3.2.7 Thin layer chromatography (TLC)

Thin layer chromatography was used to detect the conversions of TSO and ESA to their corresponding diols meso-hydrobenzoin (MHB) and 9,10-dihydroxystearic acid (DSA). For the separation of TSO and its diol a mobile phase of petrol and acetic acid ethyl ester in a ratio 2:1 was used. Employing the substrate 9,10-epoxystearic acid a mobile phase of diethyl ether, cyclohexane and acetic acid in a ratio of 70:30:1 was applied. For staining the Hanessian's stain was used consisting of 10 ml $H_2SO_4$ (conc.), 90 ml $H_2O$, 5 g ammonium molybdate and 1,22 g cerium sulphate.

### 3.2.8 HPLC

The conversion of TSO into MHB was in addition analyzed by HPLC. Following parameters were applied. A 5 µm RP- 18e column (Merck, Darmstadt, Germany) was used. An isocratic flow rate of 1 ml $min^{-1}$ with a stop time of 6 minutes was applied. The solvent consisted of 85% acetonitrile and 15% ammonium acetate in $H_20$ (0,77 g $l^{-1}$). The retention time of the diol was 2,9 min, the retention time of the epoxide 4,9 min. Detection was performed photometrically at 220 nm. A linear increase in MHB concentration in regard to the peak area was found between 20 to 1000 µM (calculation factor 6,19).

### 3.2.9 Red assay

The conversion of SO to its corresponding diol (1,2-phenyl ethanediol) was detected using an adapted version of the photometric red assay (Cedrone et al., 2005; Kahakeaw and Reetz, 2008). Based on the oxidation of the formed diol and the titration of the remaining oxidant the assay was performed as follows. After conversion 10 µl NaIO4 stock solution (50 mM in $H_2O$) were added. The suspension was incubated for 5 minutes at room temperature. Cells were centrifuged at 4°C and 4000 rpm and 100 µl

were transferred to a new microplate. 5 µl of adrenaline (50 mM in $H_20$, solubilized by HCl) were added and again incubated for 5 minutes. The decrease of red colour (back titration of the remaining oxidant) was then determined photometrically at 490 nm ($A_{490}$). A linear response of the sodium periodate concentration to $A_{490}$ was found up to a concentration of 1,25 mM. To overcome untransformed *Pichia pastoris* ($Mut^S$) background a higher periodate concentration of 2,5 mM was applied.

### 3.2.10 Protein concentration and purification

Total protein concentration was estimated using the Coomassie (Bradford) Protein Assay Kit (Thermo Fisher Scientific Inc., Rockford, IL USA). Protein purification was done using the His SpinTrap Kit (GE Healthcare, Little Chalfont, UK). Buffers were applied as follows: binding buffer 20 mM imidazole, washing buffer 50 mM imidazole and elution buffer 300 mM imidazole.

### 3.2.11 Cell lysis

Cell disruption was performed using a TS Series Benchtop machine (Constant Systems Ltd., Daventry, UK) with 2,7 kbar pressure difference. Cell lysis buffer (pH 6,5) consisted of 10 mM sodium phosphate buffer, 5 mM EDTA, 0,5 mM NaCl and 0,1% TritonX-100. Alternatively, cells were disrupted by French press (1500 psi, 3 passages).

### 3.2.12 Real Time PCR

Real Time PCR (qPCR) was performed using the ABI PRISM 7300 Real Time PCR System and Power SYBR® Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA), following a procedure described by Abad et al. (2010b). The *Pichia pastoris ARG4* gene was used as a reference for quantification. Following primers were used for amplification of the reference gene and inserted construct: ARG4-RTfw 5'-TCCTCCGGTGGCAGTTCTT-3', ARG4-RTrv 5'-TCCATTGACTCCCGTTTGAG-3', Zeocin-RTfw 5'-GACTCGGTTTCTCCCGTGACT-3', Zeocin-RTrv 5'-CTGCGGAGATGAACAGGGTAA-3'.

## 3.3 Results

### 3.3.1 Recombinant expression of 3 plant soluble epoxide hydrolases

For a first comparative study *Pichia pastoris* was transformed with 3 intracellular expression cassettes containing the epoxide hydrolase genes from *S. tuberosum* (potato, StEH1), *G. max* (soy bean, GmEH1) and *E. lagascae* (spurge, ElEH1) under the control of the wild type *AOX1* promoter (WT). In addition, *Pichia pastoris* was transformed with an alpha factor StEH1 secretion construct ($StEH1_{alpha}$). To take into account that multi copy integrations might happen three 96 well plates were screened per construct by TLC after whole cell conversion of TSO or ESA. For $StEH1_{alpha}$ the activity was determined

for the supernatant. Subsequently the best 10 clones of each construct were cultivated in shake flasks and analyzed regarding activity on TSO.

All 3 intracellular plant epoxide hydrolases were expressed by *Pichia* and enzymatically active. The best obtained clones expressing potato sEH reached specific whole cell activities between 33-415 U $g_{DCW}^{-1}$ towards TSO. Being about 300 times less active on this substrate the best soy bean sEH clones reached values between 0,19-1,25 U $g_{DCW}^{-1}$, while spurge sEH clones showed even less specific whole cell activities reaching only 0,004-0,15 U $g_{DCW}^{-1}$. In contrast, no activity was found in the culture supernatant employing the secretion variant $StEH1_{alpha}$.

To benchmark the obtained results to the commonly used *E. coli* expression system a StEH1 expressing *E. coli* clone was cultured in shake flask and whole cell conversion of TSO studied. In comparison the *E. coli* clone was 3 times less active on whole cell level reaching 150 U $g_{DCW}^{-1}$ only. Figure 3.1 summarizes the results of the best *P. pastoris* clones expressing these 3 plant sEHs in comparison to the StEH1 expressing *E. coli* clone.

A

| *P. pastoris* U $g_{DCW}^{-1}$StEH1 | *P. pastoris* U $g_{DCW}^{-1}$GmEH1 | *P. pastoris* U $g_{DCW}^{-1}$ElEH1 |
|---|---|---|
| 33 ± 1 | 0,187 ± 0,002 | 0,004 ± 0,000 |
| 35 ± 2 | 0,198 ± 0,001 | 0,004 ± 0,000 |
| 52 ± 2 | 0,212 ± 0,170 | 0,004 ± 0,001 |
| 57 ± 2 | 0,219 ± 0,006 | 0,005 ± 0,000 |
| 98 ± 3 | 0,232 ± 0,000 | 0,006 ± 0,000 |
| 113 ± 1 | 0,25 ± 0,001 | 0,012 ± 0,001 |
| 113 ± 4 | 0,269 ± 0,001 | 0,017 ± 0,001 |
| 116 ± 4 | 0,268 ± 0,002 | 0,035 ± 0,001 |
| 143 ± 8 | 0,492 ± 0,002 | 0,088 ± 0,000 |
| 415 ± 18 | 1,248 ± 0,007 | 0,149 ± 0,001 |

B

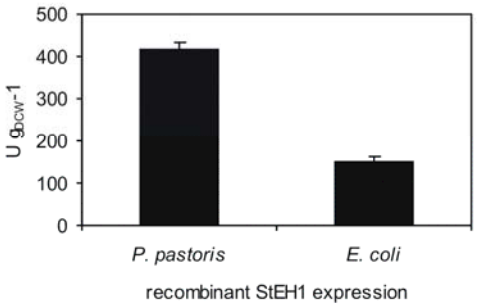



Figure 3.1 A) Specific whole cell epoxide hydrolase activities of the best obtained *P. pastoris* clones (substrate TSO, shake flask cultivations), epoxide hydrolases: StEH1, GmEH1, ElEH1, B) Comparing *P. pastoris* and *E. coli* for recombinant StEH1 expression, whole cell conversion of TSO

### 3.3.2 Euphorbia lagascae epoxide hydrolase

As little is known about spurge sEH this enzyme was further studied. The best obtained ElEH1 expressing *P. pastoris* clone was cultivated in shake flask and the recombinant protein concentration was determined by the method of Bradford after His SpinTrap purification. Accordingly, an expression level of 10 mg $l^{-1}$ and a specific activity of 0,009 U $mg^{-1}$ towards TSO were calculated. In comparison, being expressed at a similar level, a specific activity of 1,1 U $mg^{-1}$ was calculated for the best potato sEH expressing clone. SDS-PAGE of the purified spurge sEH showed a protein band of about 37 kDa, which correlated well with the predicted molecular weight (37,5 kDa including His-Tag, data not shown).

To determine the temperature optimum, but also to study the stability of spurge sEH, the enzyme was incubated at 10°C, 20°C, 30°C, 40°C, 50°C, 55°C and 60°C over a period of 300 minutes. Residual activities were determined after 10, 20, 30, 60, 120, 180 and 300 minutes of incubation. Results showed that the activity of spurge sEH increased with increasing temperature reaching a maximum at 50°C (Figure 3.2A). At 40°C the enzyme was stable over the whole period of measurement ($t_{1/2}$ >300 min). At 55°C not only enzyme activity, but also stability decreased. Half lives at different temperatures are summarized in Figure 3.2A (insert). Non-enzymatic background hydrolysis was observed at temperatures higher than 30°C.

Also the pH optimum of spurge sEH was determined. At a temperature of 28°C the enzyme was incubated at different pH values (5,3 - 9,0) over a period of 400 minutes. Residual activities were determined after 30, 60, 135, 215, 290 and 400 minutes of incubation. The pH optimum of spurge sEH was found at pH 6. Surprisingly, less than 20% residual activity were found at pH values higher than 7,5 (Figure 3.2B). Seemingly the enzyme was inactivated after incubation at higher pH values. However, half life times were not affected (pH 5,3 - 9,0: $t_{1/2}$>400 min, Figure 3.2B insert).

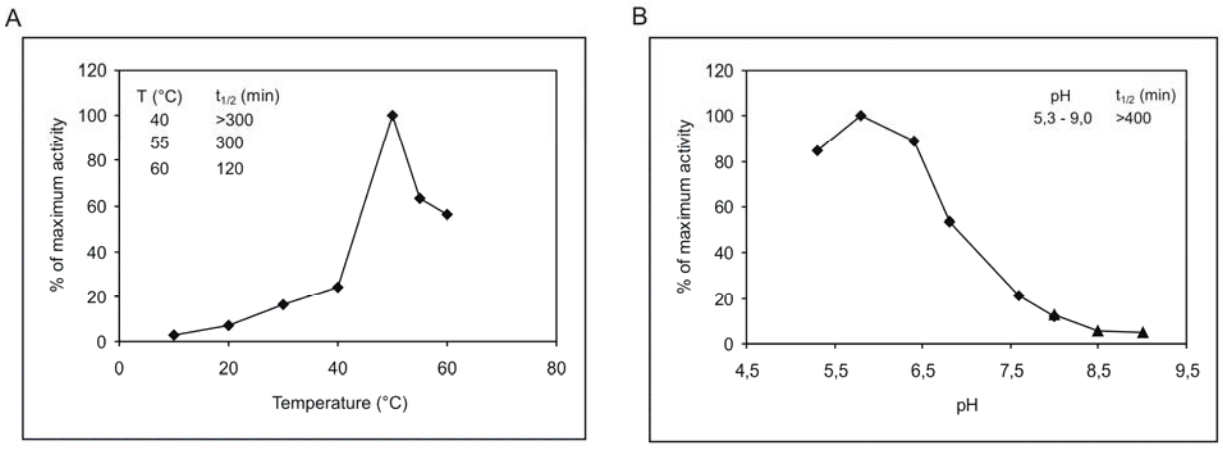



Figure 3.2 A) Temperature and B) pH optimum of spurge soluble epoxide hydrolase, substrate: TSO (final conc. 1 mM), % of maximum activity: values were normalized to the maximal activity (µM $min^{-1}$), inserts: half lives of spurge soluble epoxide hydrolase at A) different temperatures and B) different pH values

### 3.3.3 Up-scaling StEH1 expression

Potato sEH, which showed the highest activity on TSO, was chosen as model enzyme for bioreactor studies.

The best StEH1 expressing *P. pastoris* clone (clone StEH1) was cultivated in a methanol induced fed batch and the time course of expression was determined analysing whole cell conversion of TSO in correlation to induction times. The obtained expression profile showed a steep increase in activity in the beginning reaching 27 U $ml^{-1}$ within 20 hours. Subsequently a reduced increase in potato sEH activity was observed and after 80 hours of production a maximal level of 37 U $ml^{-1}$ was reached (Figure 3.3A).

For further analysis cells were disrupted by a Benchtop TS machine and potato sEH was purified using His SpinTrap columns. SDS-PAGE showed in agreement to the predicted molecular weight a protein band of approximately 37 kDa (Figure 3.3B). Benchmarking protein concentration against BSA standard on SDS-PAGE an expression level of 350 mg $l^{-1}$ was determined and a specific activity of 0,96 U $mg^{-1}$ calculated. Conspicuously, a seemingly low growth rate of 0,004 $h^{-1}$ was found for clone StEH1 on methanol. As a repetition of the bioreactor experiment led to the same results, the physiology of this strain is seemingly distorted.

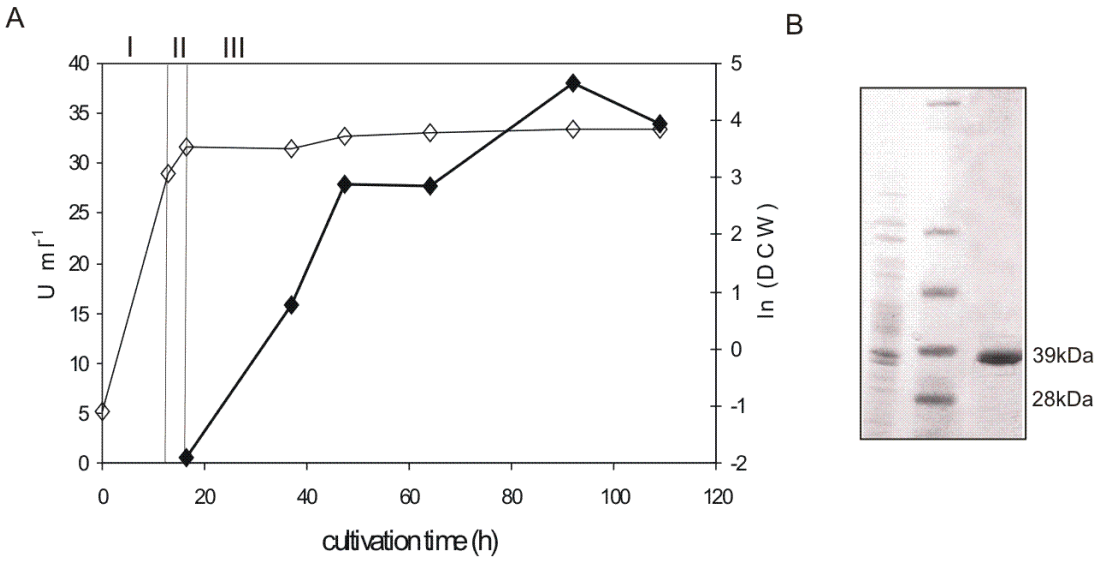



Figure 3.3 A) Time course of $P_{AOX1}$ driven potato sEH expression in a methanol induced bioreactor cultivation, I: glycerol batch phase, II: glycerol fed batch phase, III: methanol fed batch phase (production phase), U $ml^{-1}$: whole cell conversion of TSO (full squares), ln (DCW): natural logarithm of the dry cell weight (empty squares), B) SDS-page of SpinTrap purified potato sEH, lane 1: cell lysate, lane 3: purified potato sEH; 10 µg total protein applied

### 3.3.4 Employing promoter technology for the expression of StEH1

Promoter technology was applied to define the maximum in potato sEH expression, which can be reached by tuning transcription. These promoter technology experiments should also give information if significant amounts of potato sEH can be produced in a methanol free, but regulated way. According to the defined goals the promoter variants $d1^+$ and d6*, showing different induction characteristics, were chosen for the expression of potato sEH. These promoter variants were recently described by Hartner et al. (2008). In the referred study sequence deletions and duplications were introduced into $P_{AOX1}$ according to *in-silico* predicted transcription factor binding sites. Subsequent the promoter library was evaluated expressing green fluorescent protein (GFP) and different industrial reporters. While application of promoter variant $d1^+$ caused an enhanced methanol induced expression of about 160%

compared to $P_{AOX1}$, promoter variant d6* was found to be regulated by glucose depletion (derepression), resulting in a 4-fold increased activity before induction.

*Pichia pastoris* was transformed with two intracellular expression cassettes, in which gene expression was under control of the promoters $d1^+$ and d6* ($StEH1_{d1+}$ and $StEH1_{d6*}$). One 96 well plate was screened per construct using SO and TSO as substrate and the obtained clones were benchmarked against the best WT promoter potato sEH expressing clone ($StEH1_{WT}$). Though less then 100 clones were screened per construct several clones reaching an activity similar to the previously identified best expression clone $StEH1_{WT}$ were found. Confirming the expectations most of the high activity clones were based on promoter variant $d1^+$ (Figure 3.4). For methanol free expression clones based on promoter variant d6* performed best (data not shown). However, comparing methanol induced to methanol free expression, only 20% of the maximal methanol induced WT level was reached via derepression technology.

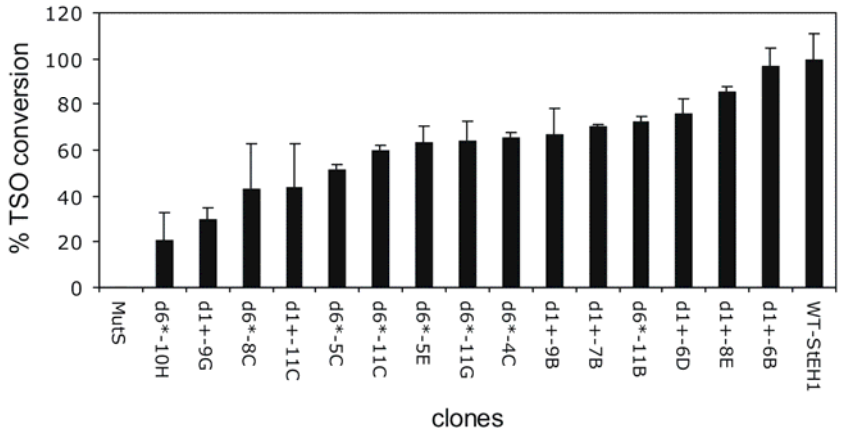



Figure 3.4 Microscale rescreening results of the different promoter variants driving potato sEH expression, % TSO conversion: whole cell conversion of TSO normalized to $StEH1_{WT}$, promoter variants: $d1^+$, d6*

For comparison the best $d1^+$ driven potato sEH expressing clone (clone $StEH1_{d1+}$, clone 6B Figure 3.4) was analyzed in a methanol induced fed batch cultivation. The time course of expression, again determined on whole cell level, showed a steady increase in activity until 80 hours of methanol induction reaching 18 U $ml^{-1}$ (Figure 3.5). Similar as described for $StEH1_{WT}$ cells were disrupted and the recombinant enzyme was purified for further analysis. An expression level of about 200 mg $l^{-1}$ was estimated and in agreement to above a specific activity of 1,2 U $mg^{-1}$ was calculated. Interestingly, applying a similar methanol feed rate a significantly higher growth rate of 0,012 $h^{-1}$ was estimated for the production phase of clone $StEH1_{d1+}$.

As according to Abad et al. (2010b) the gene copy number influences the interpretation of experimental results from promoter studies, the clones $StEH1_{WT}$ and $StEH1_{d1+}$ were analyzed by quantitative PCR (qPCR, real-time PCR). Surprisingly, while clone $StEH1_{WT}$ had multiple expression copies integrated (9-10 copies), clone $StEH1_{d1+}$ was a single copy clone.

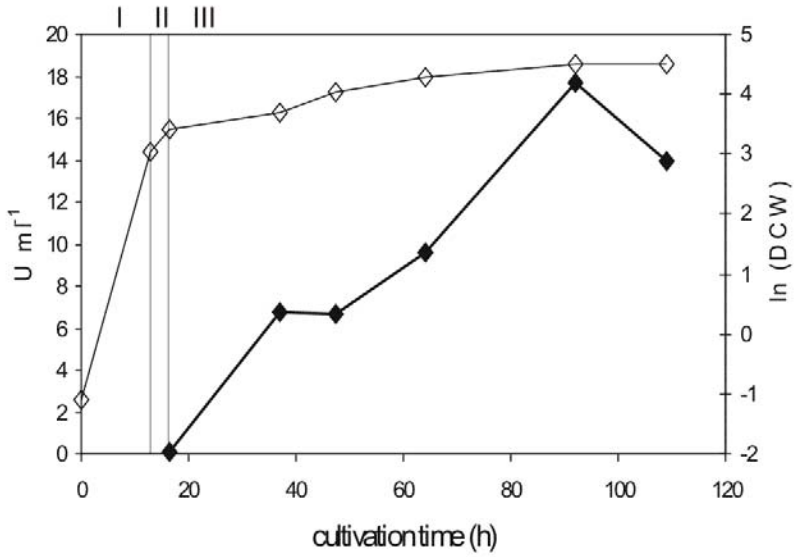



Figure 3.5 Time course of $d1^+$ driven potato sEH expression in a methanol induced bioreactor cultivation, I: glycerol batch phase, II: glycerol fed batch phase, III: methanol fed batch phase (production phase), U $ml^{-1}$: whole cell conversion of TSO (full squares), ln (DCW): natural logarithm of the dry cell weight (empty squares)

## 3.4 Discussion

Confirming *Pichia pastoris* as host system for recombinant plant epoxide hydrolase production, the soluble epoxide hydrolases of *Solanum tuberosum* (StEH1), *Glycine max* (GmEH1) and *Euphorbia lagascae* (ElEH1) were successfully expressed. Similar to Bellevik et al. (2002 a,b), who showed the expression of AtEH1 and BnEH1 in *P. pastoris*, an expression level of 10 mg $l^{-1}$ was reached after intracellular expression of StEH1 in shake flask. However, by using the advantage of *P. pastoris* high cell density cultivations we increased the expression level of StEH1 to 350 mg $l^{-1}$.

Comparing *P. pastoris* and *E. coli* for the expression of StEH1 on whole cell level, the *Pichia pastoris* clone was 3 times more active reaching 415 U $g_{DCW}^{-1}$. Thus, it seems that *Pichia pastoris* is not only an alternative epoxide hydrolase production system, but key to the generation of improved sEH whole cell biocatalysts.

Similar to the other plant sEHs also ElEH1 was expressed at a level of 10 mg $l^{-1}$ in shake flask. Compared to reported inclusion bodies in *E. coli* this result was satisfying and there is still room for further improvement by bioreactor cultivation (Edqvist and Farbos, 2003). Characterizing the spurge sEH a specific activity of 0,009 U $mg^{-1}$ was determined for TSO. In comparison to StEH1 TSO was hydrolyzed 100 times slower. Epoxy fatty acids such as vernolic acid might be the preferred substrate of spurge soluble epoxide hydrolase. Nevertheless, this is the first time specific activity of ElEH1 towards TSO was reported, though the enzyme was known to be active on this substrate (Edqvist and Farbos, 2003).

In this study a specific activity of about 1,1 U $mg^{-1}$ was determined for *Pichia pastoris* recombinant StEH1. Comparing to literature reports differences were found. While for purified recombinant insect cell

StEH1 a value of only 0,375 U $mg^{-1}$ was reported by Morisseau et al. (2000), Elfström and Widersten (2005) determined the high specific activity of 6,1 U $mg^{-1}$ for purified *E. coli* StEH1. Seemingly, the choice of the expression host influenced the specific activity of recombinant StEH1. However, as also not host specific factors as for example purification and measurement methods can influence specific activity values, the reasons for the seen discrepancies remain unclear (Elfström and Widersten, 2005). As a high temperature optimum is commonly reported for plant sEHs the obtained optimum of spurge sEH at 50°C didn't surprise (Bellevik et al., 2002a; Morisseau et al., 2000). However, the low pH optimum of 6 and significantly diminished activities at pH 7,5 and higher surprised. In comparison, *Glycine max* epoxide hydrolase reaches its optimum between pH 7-8, while *Ricinus communis* and *Arabidopsis thaliana* sEH show highest activity even between pH 8-9 (Bellevik et al., 2002a; Blee and Schuber, 1992a; Stark et al., 1995). With a pH optimum between 6-7 *Brassica napus* sEH seems to be most similar to the studied *Euphorbia lagascae* enzyme (Bellevik et al., 2002b).
Employing promoter technology for the expression of recombinant potato sEH gave interesting insights. On the one hand we were able to define the limit of StEH1 expression reachable through transcriptional tuning (promoter, gene copy number) to about 37 U $ml^{-1}$ ($StEH1_{WT}$). Employing a stronger promoter also resulted in a high activity transformant ($StEH1_{d1+}$), which reached a level of 18 U $ml^{-1}$ even with only 1 copy of the expression cassette integrated. On the other hand only low activity, 20% of $StEH1_{WT}$, was found employing the derepressed promoter variant 6*. Seemingly potato sEH is not a suitable target for methanol free production in *P. pastoris* driven by the d6* promoter.
Surprisingly was the low growth rate on methanol of clone $StEH1_{WT}$ in comparison to clone $StEH1_{d1+}$. Possibly extensive overproduction of sEH (37 U $ml^{-1}$, $StEH1_{WT}$) had negative effects on the physiology of the cells. Thus, bottlenecks other then transcription seem to limit further advances in *P. pastoris* recombinant sEH expression. One working point might be the localization of the recombinant enzyme, which might be a limiting factor as previously observed for a D-amino acid oxidase expressed in *Pichia pastoris* (Abad et al., 2010a). Facilitated targeting to the peroxisomes might prevent toxic by-product formation or unwanted enzymatic reactions in the cytosol, thus raise the limit of *P. pastoris* plant sEH production.

## 3.5 References


Abad, S., Nahalka, J., Bergler, G., Arnold, S.A., Speight, R., Fotheringham, I., Nidetzky, B., Glieder, A., (2010a) Stepwise engineering of a Pichia pastoris D-amino acid oxidase whole cell catalyst. Microb Cell Fact 9, 24.

Abad, S., Kitz, K., Hörmann, A., Schreiner, U., Hartner, F.S., Glieder, A., (2010b) Real-time PCR-based determination of gene copy numbers in Pichia pastoris. Biotechnol J 5, 413-420.

Arahira, M., Nong, V.H., Udaka, K., Fukazawa, C., (2000) Purification, molecular cloning and ethylene-inducible expression of a soluble-type epoxide hydrolase from soybean (Glycine max [L.] Merr.). Eur J Biochem 267, 2649-2657.

Bellevik, S., Summerer, S., Meijer, J., (2002a) Overexpression of Arabidopsis thaliana soluble epoxide hydrolase 1 in Pichia pastoris and characterisation of the recombinant enzyme. Protein Expr Purif 26, 65-70.

Bellevik, S., Zhang, J., Meijer, J., (2002b) Brassica napus soluble epoxide hydrolase (BNSEH1). Eur J Biochem 269, 5295-5302.

Blee, E., Schuber, F., (1992a) Occurrence of fatty acid epoxide hydrolases in soybean (Glycine max). Purification and characterization of the soluble form. Biochem J 282 ( Pt 3), 711-714.

Blee, E., Schuber, F., (1992b) Regio- and enantioselectivity of soybean fatty acid epoxide hydrolase. J Biol Chem 267, 11881-11887.

Cedrone, F., Bhatnagar, T., Baratti, J.C., (2005) Colorimetric assays for quantitative analysis and screening of epoxide hydrolase activity. Biotechnol Lett 27, 1921-1927.

Cregg, J.M., Cereghino, J.L., Shi, J., Higgins, D.R., (2000) Recombinant protein expression in Pichia pastoris. Mol Biotechnol 16, 23-52.

Edqvist, J., Farbos, I., (2003) A germination-specific epoxide hydrolase from Euphorbia lagascae. Planta 216, 403-412.

Elfström, L.T., Widersten, M., (2005) Catalysis of potato epoxide hydrolase, StEH1. Biochem J 390, 633-640.

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.

Kahakeaw, D., Reetz, M.T., (2008) A cell-based adrenaline assay for automated high-throughput activity screening of epoxide hydrolases. Chem Asian J 3, 233-238.

Li, Z., Feiten, H.J., Chang, D., Duetz, W.A., van Beilen, J.B., Witholt, B., (2001) Preparation of (R)- and (S)-N-protected 3-hydroxypyrrolidines by hydroxylation with Sphingomonas sp. HXN-200, a highly active, regio- and stereoselective, and easy to handle biocatalyst. J Org Chem 66, 8424-8430.

Lin-Cereghino, J., Wong, W.W., Xiong, S., Giang, W., Luong, L.T., Vu, J., Johnson, S.D., Lin-Cereghino, G.P., (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. Biotechniques 38, 44, 46, 48.

Liu, Z., Pscheidt, B., Avi, M., Gaisberger, R., Hartner, F.S., Schuster, C., Skranc, W., Gruber, K., Glieder, A., (2008) Laboratory evolved biocatalysts for stereoselective syntheses of substituted benzaldehyde cyanohydrins. Chembiochem 9, 58-61.

Morisseau, C., Beetham, J.K., Pinot, F., Debernard, S., Newman, J.W., Hammock, B.D., (2000) Cress and potato soluble epoxide hydrolases: purification, biochemical characterization, and comparison to mammalian enzymes. Arch Biochem Biophys 378, 321-332.

Newman, J.W., Morisseau, C., Hammock, B.D., (2005) Epoxide hydrolases: their roles and interactions with lipid metabolism. Prog Lipid Res 44, 1-51.

Ruth, C., Zuellig, T., Mellitzer, A., Weis, R., Looser, V., Kovar, K., Glieder, A., (2010) Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in Pichia pastoris. Syst Synth Biol, in press.

Stark, A., Houshmand, H., Sandberg, M., Meijer J., (1995) Characterization of the activity of fatty-acid epoxide hydrolase in seeds of castor bean. Planta 197, 84-88.

Wang, W., Malcolm, B.A., (1999) 2-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. Biotechniques 26, 680-682.

Weis, R., Luiten, R., Skranc, W., Schwab, H., Wubbolts, M., Glieder, A., (2004) Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena. FEMS Yeast Res 5, 179-189.

Williamson, K.C., Morisseau, C., Maxwell, J.E. and Hammok, B.D., (2000) Regio and enantioselective hydrolysis of phenyloxiranes catalyzed by soluble epoxide hydrolase. Tetrahedron 11, 4451-4462.

## 3.6 Supplementary

Figure 3.6 Mating factor α signal sequence

>amino acid sequence
M R F P S I F T A V L F A A S S A L A A P V N T T T E D E T A Q I P A E A V I G Y S D L E G D F D V A V L P F S N S T N N G L L F I N T T I A S I A A K E E G V S L E K R E A E A

>nucleotide sequence
atgagattcccatctattttcaccgctgtcttgttcgctgcctcctctgcattggctgcccctgttaacactaccactgaagacgagactgctcaaattcc agctgaagcagttatcggttactctgaccttgagggtgatttcgacgtcgctgttttgcctttctctaactccactaacaacggtttgttgttcattaacacc actatcgcttccattgctgctaaggaagagggtgtctctctcgagaagagaggccgaagct

## 3.7 Annex

Table 3.1 Potato, soy bean and spurge soluble epoxide hydrolase *P. pastoris* expression strains

| strain coll. | host | strain | plasmid | copy number | origin |
|---|---|---|---|---|---|
| 3529 | CBS7435 Mut$^S$ | PotHis linear expression cassette | cas_B1 | 9-10 | Thesis S. Ertl |
| 3530 | CBS7435 Mut$^S$ | EuphHis linear expression cassette | cas_B1 | 10-11 | Thesis S. Ertl |
| 3531 | CBS7435 Mut$^S$ | SoyhHis linear expression cassette | cas_B1 | 4-5 | Thesis S. Ertl |
| 6103 | CBS7435 Mut$^S$ | d1+_PotStEH_1-6B | cas_B1 | 1 | |
| 6106 | CBS7435 Mut$^S$ | d1+_PotStEH_1-5B | cas_B1 | x | |
| 6104 | CBS7435 Mut$^S$ | d6*_PotStEH_2-11B | cas_B1 | x | |
| 6105 | CBS7435 Mut$^S$ | AOX_noEH_N* | cas_B1 | x | |

*negative control without gene of interest

Table 3.2 Plasmids coding for potato, soy bean and spurge soluble epoxide hydrolase expression; Primers were used according to the Diploma thesis of Stefan Ertl

| strain coll. | plasmid | host | origin |
|---|---|---|---|
| 5886 | Potato_EH_opt in pUC57 | K12 Top10 | |
| 5887 | Euphorbia lagascae_EH_opt in pUC57 | K12 Top10 | |
| 5888 | Soybean_EH_opt in pUC57 | K12 Top10 | |
| 5839 | pJet1_Soybean epoxide hydrolase_α factor | K12 DH5α | Thesis S. Ertl |
| 5840 | pJet1_Soybean short epoxide hydrolase_noHis-tag | K12 DH5α | Thesis S. Ertl |
| 5841 | pJet1_Soybean short epoxide hydrolase_α factor | K12 DH5α | Thesis S. Ertl |
| 5842 | pJet1_Soybean short epoxide hydrolase_His-tag | K12 DH5α | Thesis S. Ertl |

| | | | |
|---|---|---|---|
| 5843 | pJet1_Soybean epoxide hydrolase_His-tag | K12 DH5α | Thesis S. Ertl |
| 5850 | pJet1_Soybean epoxide hydrolase_noHis-tag | K12 DH5α | Thesis S. Ertl |
| 5844 | pJet1_Euphorbia epoxide hydrolase_α factor | K12 DH5α | Thesis S. Ertl |
| 5845 | pJet1_Euphorbia epoxide hydrolase_His-tag | K12 DH5α | Thesis S. Ertl |
| 5846 | pJet1_Euphorbia epoxide hydrolase_noHis-tag | K12 DH5α | Thesis S. Ertl |
| 5847 | pJet1_Potato epoxide hydrolase_noHis-tag | K12 DH5α | Thesis S. Ertl |
| 5848 | pJet1_Potato epoxide hydrolase_His-tag | K12 DH5α | Thesis S. Ertl |
| 5849 | pJet1_Potato epoxide hydrolase_α factor | K12 DH5α | Thesis S. Ertl |
| 6048 | pJET_d1+_PotHIS | K12 DH5α | |
| 6049 | pJET_d6*_PotHIS | K12 DH5α | |
| 6050 | pJET_d6.3xMat_PotHIS | K12 DH5α | |
| 6051 | pJET_d6.3xMat_SoyHIS | K12 DH5α | |
| 6052 | pJET_d6.3xMat_EuphHIS | K12 DH5α | |
| 6088 | puc57PotEH E.coli | K12 Top10 F' | |

# 4 High level expression of *Bungarus fasciatus* acetylcholinesterase

## 4.1 Introduction

Acetylcholinesterases (AChEs) are highly interesting enzymes, which terminate acetylcholine-mediated neurotransmission at synapses by hydrolyzing acetylcholine to acetate and choline (Selwood et al., 1993). Organophosphate and carbamate compounds used in insecticides, but also for the development of nerve agents, irreversibly inhibit acetylcholinesterases (Boublik et al., 2002). AChEs, recombinant or naturally isolated, can be used in antidotal therapies against organophosphates, but also as biological elements in biosensors for the detection of such (Taylor et al., 2007). Besides their role in cholinergic transmission, AChEs are also connected to cell growth and death (Jiang and Zhang, 2007). Though the mechanism behind the involvement of AChEs in such physiological processes is not fully understood, it is assumed that AChE expression influences the expression of certain gene groups like for example those involved in apoptosis. Also the involvement of AChEs in Alzheimer diseases (characterized by amyloid fibril deposition), where acetylcholinesterase fragments are suspected to be connected to amyloid b fibril formation, is under investigation (Jiang and Zhang, 2007).
On protein level AChEs are highly polymorphic enzymes, which are in vertebrates encoded by one single gene (Massoulie et al., 1999). Alternative splicing generates different subunits, which contain the same catalytic domain, but different C-terminal peptides (Massoulie et al., 1993). H-peptides contain a free cystein able to form inter-chain disulfide bonds. H-subunits produce amphiphilic, glycophosphatidylinositol anchored dimers. Similar also T-peptides contain a free cystein near the C-terminus. T-subunits produce monomers, homo-oligomers and by incorporation of hydrophobic membrane anchors also hetero-oligomers (Morel and Massoulie, 1997; Cousin et al., 1998). Differently, *Bungarus fasciatus* (banded krait) venom produces a hydrophilic soluble monomeric subform of AChE containing neither T nor H, but a cystein free hydrophilic C-terminal S-peptide (Cousin et al., 1998). Based on the structure of *Torpedo* AChE a 3-dimensional model was proposed for BfAChE by Cousin et al. (1996). In comparison to *Torpedo* AChE two mutations were found in the peripheral-site of BfAChE (methionine 101 and lysine 316, Figure 4.1). These mutations were found to be connected to the much lower sensitivity of BfAChE to peripheral-site inhibitors as for example propidium (Cousin et al., 1996). BfAChE contains 6 cysteines that form intra-molecular disulfide bonds, which are conserved through cholinesterases. Also conserved is the lack of other cysteines in the sequence. Additionally, BfAChE has 4 potential N-glycosylation sites displaying the N-X-T/S motif. The catalytic triad contains the amino acids serine, glutamic acid and histidine at the positions 231, 358, and 471, respectively (Cousin et al., 1996).
Over the last years *P. pastoris* has proven itself as effective host system for the expression of AChEs. Carp, rat, *Drosophila* and *Electrophorus* AChEs have been successfully expressed using this host system (Morel and Massoulie, 1997; Sato et al., 2009; Simon and Massoulie, 1997; Wu et al., 2008). Also BfAChE has been expressed in *P. pastoris* under the control of the methanol inducible *AOX1* promoter (Morel and Massoulie, 1997). Interestingly, in comparison to other *P. pastoris* recombinant

AChEs the yield of BfAChE was 2-3 times higher (Morel and Massoulie, 1997). Solving the puzzle, Morel and Massoulie (2000) found a stimulatory effect of the BfAChE coding sequence on transcription responsible. However, as demonstrated by Weill et al. (2002) the seen effect was not transferable to a homologous human AChE, which indicates the effect context depending.
In this study for the first time we evaluate different promoters for the expression of BfAChE in *P. pastoris* accounting for the fact that the promoter regulates transcript levels and thereby also the amount of translated product per time which is correlated to protein folding.

A

```
T.m. -----MREMNLLVTSSLGVLLHL--VVLCQADDDSELLVNTKSGKVMRTRIPVLSSHISA 53
B.f. MPSCQPGKMPAPWPWWLQLLLCIPSCVAVLPGRAGELKVSTQTGSVRGLSLPVLDGHVSA 60

T.m. FLGIPFAEPPVGNMRFRRPEPKKPWSGVWNASTYPNNCQQYVDEQFPGFPGSEMWNPNRE 113
B.f. FLGIPFAEPPLGRMRFLRPEPVKPWQHVLDATSYKPACYQMVDTSYPGFQGTEMWNPNRG 120

T.m. MSEDCLYLNIWVPSPRPKSATVMLWIYGGGFYSGSSTLDVYNGKYLAYTEEVVLVSLSYR 173
B.f. MSEDCLYLNIWVPSPRPKDAPVLVWIYGGGFYSGAASLDVYDGRFLTYTQNVILVSLSYR 180

T.m. VGAFGFLALHGSQEAPGNMGLLDQRMALQWVHDNIQFFGGDPKTVTLFGESAGRASVGMH 233
B.f. VGAFGFLGLPGSPEAPGNMGLLDQRLALQWIQNNIHPFGGNPRAVTVFGESAGAASVGMH 240

T.m. ILSPGSRDLFRRAILQSGSPNCPWASVSVAEGRRRAVELRRNLNCNLNSDEDLIQCLREK 293
B.f. LLSTQSRTLFQRAILQSGGPNAPWATVTPAESRGRAALLGKQLGCHFNNDSELVSCLRSK 300

T.m. KPQELIDVEWNVLPFDSIFRFSFVPVIDGEFFPTSLESMLNAGNFKKTQILLGVNKDEGS 353
B.f. NPQELIDEEWSVLPYKSIFRFPFVPVIDGDFFPDTPEAMLSSGNFKETQVLLGVVKDEGS 360

T.m. FFLLYGAPGFSKDSESKISREDFMSGVKLSVPHANDLGLDAVTLQYTDWMDDNNGIKNRD 413
B.f. YFLIYGLPGFSKDNESLISRADFLEGVRMSVPHANDIATDAVVLQYTDWQDQDNREKNRE 420

T.m. GLDDIVGDHNVICPLMHFVNKYTKFGNGTYLYFFNHRASNLVWPEWMGVIHGYEIEFVFG 473
B.f. ALDDIVGDHNVICPVVQFANDYAKRNSKVYAYLFDHRASNLLWPPWMGVPHGYEIEFVFG 480

T.m. LPLVKELNYTAEEEALSRRIMHYWATFAKTGNPNEPHSQESKWPLFTTKEQKFIDLNTEP 533
B.f. LPLNDSLNYTPQEKELSRRMMRYWANFARTGNPTDPADKSGAWPTYTASQPQYVQLNTQP 540

T.m. IKVHQRLRVQMCVFWNQFLPKLLNAT
T.m.H-peptide                  ACDGELSSSGTSSSKGIIFYVLFSILYLIFY--- 590
T.m.T-peptide                  ETIDEAERQWKTEFHRWSSYMMHWKNQFDQYSRH 593
B.f. LATQPSLRAQICAFWNHFLPKLLNATVDPPRADRRRRSARA------------------- 581

T.m.H-peptide  ------
T.m.T-peptide  ENCAEL 599
B.f.           ------
```

B

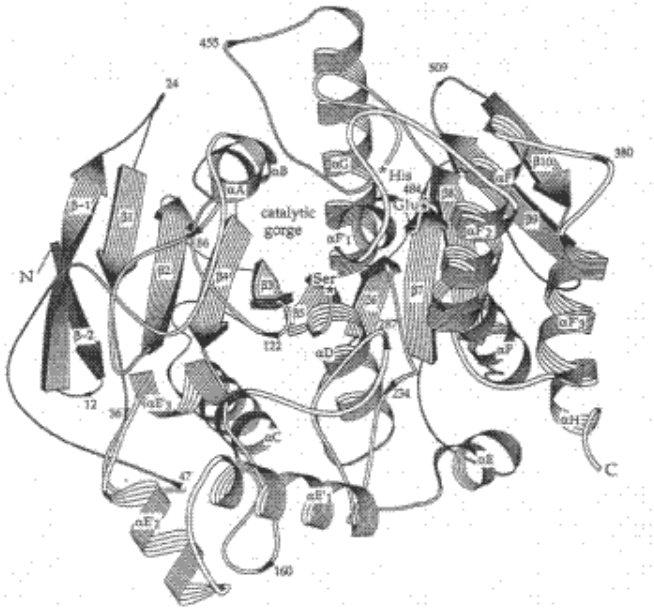



Figure 4.1 A) Alignment of *T. marmorata* (*T.m.*) and *B. fasciatus* (*B.f.*) AChE adopted from (Cousin et al., 1996), circled: catalytic triad S231, E358, and H471, connected boxes: internal disulfide bridges, black dots: positions of the BfAChE residues 101 and 316, which influence the properties of the peripheral site (see text), *B.f.*: S-subunit, *T.m.*: H- and T-subunit B) 3D structure of AChE from *T. californica*, which was used to propose a 3-dimensional model for BfAChE (Cousin et al., 1996, Massoulie et al., 1993; Sussman et al., 1991)

## 4.2 Materials and methods

### 4.2.1 Expression cassettes

*B. fasciatus* acetylcholinesterase (accession no. AAC59905) was codon optimized using the Gene Designer software (DNA2.0, Menlo Park, CA, USA) applying *P. pastoris* high methanol codon usage (Abad et al, 2010a). Expression cassettes were used as described under 3.2.1. According to Liu et al. (2008) the generated linear PCR cassettes were directly transformed or prior transformation sub cloned using the Fermentas GeneJET™ PCR Cloning Kit (Burlington, Ontario, Canada). Following linker sequences were used for the overlap extension (OE)-PCR of the individual fragments of the expression cassette: linker 1: ctaggtacttcgaaacgaggacttcacg, linker 2: gtcagatagcgaggtcactcagtc.

### 4.2.2 Strains

Experiments were performed using a *P. pastoris* CBS7435 *AOX1* knockout strain ($Mut^S$, see 3.2.1). In addition, a mixture of different copy number $Mut^S$-PDI strains (1, 4 and 10 copies; PDI: protein disulfide isomerase, accession no. CAC33587) was employed. PDI containing strains were generated by transformation of strain $Mut^S$ with plasmid pPpKan-PDI coding for PDI expression under the control of the *AOX1* promoter. Plasmid pPpKan was generated from plasmid pPpT2 (see 2.2.1) by exchanging the zeocin against a kanamycin resistance gene (KanMx, described by Oka et al., 1981). Copy number determination was done using Real Time (RT)-PCR as described by Abad et al. (2010b) (see also 3.2.12). A negative control strain N, $Mut^S$ transformed with a cassette without AChE gene, was used to proof activity result of BfAChE expression only. *P. pastoris* strain CBS7435 was obtained from CBS fungal biodiversity centre. Subcloning was performed as described under 3.2.1

### 4.2.3 AOX1 promoter variants

The strong methanol inducible *AOX1* promoter and 4 different *AOX1* promoter variants were employed in this study. The promoter variants $d1^+$ and d6* were used as described by Hartner and Glieder (2005) or Hartner et al. (2008). The short synthetic promoter AOX176-Rap1 was used as described by Ruth et al. (2010) (see also 2.2.4). Promoter variant d6_3xMat1-Mc (referred to as d6_3xMat) was generated by site directed mutagenesis (SDM) according to Wang and Malcolm (1999). Promoter d6_3xMat is characterized by deletion of region d6 (30 bases at position -223) and triplication of the putative Mat1-Mc binding region (-253/-269). Additionally to the applied *AOX1* promoter variants the *P. pastoris* endogenous *GAP* promoter was used in this study. Amplification was done using following primers: Gapfw 5'-gtcttggtgtcctcgtccaatcagg-3', Gaplinkrv: 5'-cgtgaagtcctcgtttcgaagtacctagatagttgttcaattgat tgaaatag-3'.

### 4.2.4 Chemicals and media

Endoglucosidase H was purchased from New England Biolabs (Ipswich, MA, USA). Chemicals and media were used as described under 3.2.3. Buffered minimal media were used according to Weis et al. (2004). Modified basal salt medium (Invitrogen) and bioreactor feed solutions were used similar to 3.2.3.

### 4.2.5 Transformation and cultivation

The condensed protocol of Lin-Cereghino et al. (2005) was used for *P. pastoris* transformations. Screening was done in 96 deep well plates according to Weis et al. (2004). Clones were grown 60 hours in 300 µl BMD1%. Induction was started by addition of 250 µl BMM2 (0 hours) followed by 50 µl pulses of BMM10 after 12, 24 and 48 h induction time. Samples of 50 µl were taken after 0 and 48 h of induction time. Subsequently, cells were harvested at 4°C and 4000 rpm and the supernatant was used for activity measurements. In the rescreening process selected clones were inoculated 4 times each. Bioreactor cultivations of the best obtained clones were performed using the Infors Multifors system (Infors AG, Bottmingen-Basel, Germany). Total batch volume was 500 ml and inoculum volume 50 ml. During all cultivations pH 5 was applied. In fed batch phase 2 the temperature was lowered from 28°C to 24°C. Similar to 3.2.5 fed batch feeds were adjusted according to the $pO_2$ signal of the culture. In the methanol induced cultivations the amount of glycerol used in fed batch phase 1 was 40,6 and 33,3 g for the d6* and *AOX1* based clone. The corresponding feed rates were 23,2 and 4,8 g $l^{-1}$ $h^{-1}$. In fed batch phase 2 the methanol feed varied between 2,1 and 8,2 g $l^{-1}$ $h^{-1}$ for the respective clones. In the methanol free cultivations the amount of glycerol used in fed batch phase 1 was 13,3 (8,9 g $l^{-1}$ $h^{-1}$ ) and 9,4 g (9,4 g $l^{-1}$ $h^{-1}$ ) for the *GAP* and AOX176-Rap1 based clone. In fed batch phase 2 the glycerol feed varied between 9,5 and 6,7 g $l^{-1}$ $h^{-1}$.

### 4.2.6 Activity assay

The Ellman assay was used to determine BfAChE activity (Ellman et al., 1961). Nitrobenzoate absorbance was measured for 5 minutes at 412 nm. Ellman's reagent contained 1 mM acetylthiocholine iodide and 0,5 mM 5,5′-dithiobis(2-nitrobenzoic acid) in 50 mM sodium phosphate buffer, pH 7,4. The reaction was performed in 96 well microplates. To start the reaction 190 µl reagent were added to 10 µl of sample. Kinetics were measured using a SPECTRAmax Plus 384 plate reader (Molecular Devices, Sunnyvale, CA, USA). 1 EU correlated to the amount of enzyme which increased the absorbance by 1 in 1 minute in a 1 ml cuvette and a pathway of 1 cm. On the other hand 1 U was defined as the amount of enzyme which hydrolyzed 1 µmole of substrate in 1 minute. Neither $Mut^S$ nor the negative control strain N showed esterase activity.

### 4.2.7 Activity staining

Novex$^R$ 12% Tris-Glycine native protein gels were used (Invitrogen). Esterase activity staining of the native protein gels was performed using a 1:1 mixture of 1 mM Fast Red Violet LB salt and 2 mM α-

naphthylacetate in 50 mM sodium phosphate buffer, pH 7,4. Brown coloured bands indicated esterase activity. After activity staining gels were in addition Coomassie stained.

### 4.2.8 Protein analysis

Total protein concentration was estimated using the Coomassie (Bradford) Protein Assay Kit of Pierce (Thermo Fisher Scientific Inc., Rockford, IL USA). Electrophoretic separation of the protein samples was done using the Caliper Life Science LabChip ® GX II (Hopkinton, MA 01748 USA), following the HT Protein Express LabChip® Assay user instructions. In addition, NuPAGE$^{R}$ Novex 4-12% Bis-Tris gels were used (Invitrogen).

## 4.3 Results and discussion

Employing the *AOX1* promoter (WT), the *GAP* promoter and the promoter variants d1$^{+}$, d6*, AOX176-Rap1 and d6_3xMat, *P. pastoris* was transformed with 6 different BfAChE expression constructs. In addition, a $P_{AOX1}$-PDI containing strain was co-transformed with 4 different cassettes coding for BfAChE expression under control of the promoters d1$^{+}$, d6*, AOX176-Rap1 and *GAP*.
To facilitate efficient secretion the C-terminal end of the protein was truncated right after the catalytic domain creating the new C-terminal end N-A-T similar to Cousin et al. (1996). For secretion the native and alpha mating factor signal sequence (native/alpha ss) was used (Morel and Massoulie, 1997). As Cousin et al. (1996) reported two potential native ss cleavage sites, alpha ss was fused to both of them (amino acid sequence cleavage positions: 28/29 and 34/35, referred to as a1 and a2).
According to Liu et al. (2008) each expression cassette contained 3 fragments, which were assembled by OE-PCR. Fragment 1 contained the promoter, fragment 2 the gene of interest and fragment 3 the terminator and selection cassette (Figure 4.2A). Due to the short homologous sequence of AOX176-Rap1 to the 5' *AOX1* integration site, the expression cassette was reorganized to allow integration at the *AOX1* terminator. Thus, fragment 1, now containing the selection cassette, was fused to AOX176-Rap1 via linker 2. Using linker 1, AOX176-Rap1 was in turn fused to the gene of interest and *AOX1 t*erminator containing fragment (Figure 4.2B).

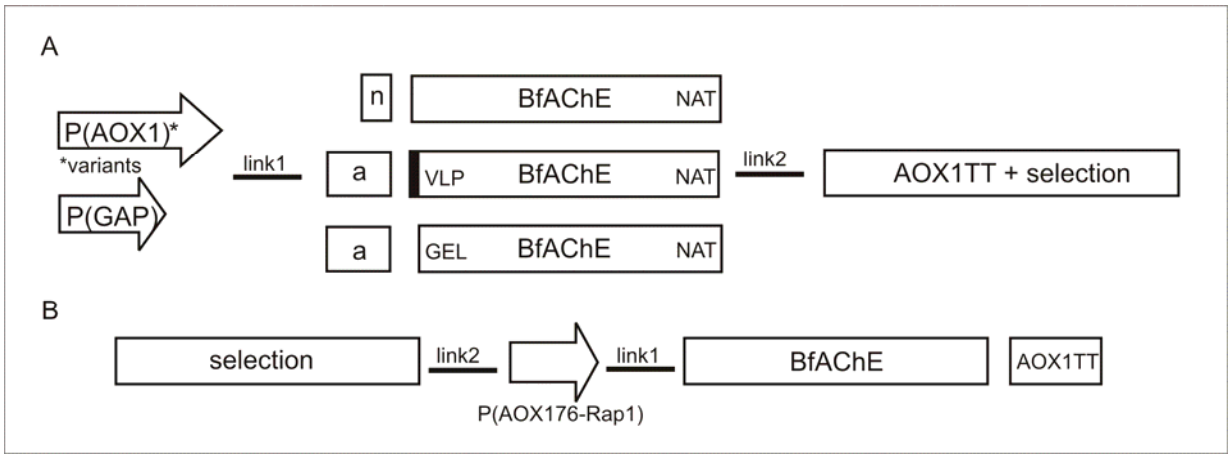



Figure 4.2 BfAChE expression constructs, promoters used: A) *AOX1*, d1$^{+}$, d6*, d6_3xMat, *GAP*, B) AOX176-Rap1, signal sequences: alpha ss (a), native ss (n); VLP, GEL: N-terminal sequences of the mature peptides after cleavage at the first (a1) or alternative (a2) predicted signal sequence cleavage site, NAT: C-terminus (truncation after catalytic domain), AOX1TT: *AOX1* terminator, link 1 and 2: linker 1 and 2 for OE-PCR

A minimum of 80 clones per construct was screened in 96 well plates and evaluated after 0 and 48 hours of induction (data not shown). In rescreening, also performed in 96 well plates, the number of clones was reduced to the best 3-4 clones from screening. A detailed list of all rescreened clones and used abbreviations can be found in Table 4.1.

Table 4.1 AChE expressing *P. pastoris* clones evaluated in rescreening, ss: signal sequence, alpha: alpha mating factor ss, native: native BfAChE ss, a1 and a2 refer to the N-terminal sequence of the mature peptide after cleavage at the first or alternative predicted ss cleavage site

| promoter | ss | strain | abbrev. | rescreen no. | promoter | ss | strain | abbrev. | rescreen no. |
|---|---|---|---|---|---|---|---|---|---|
| d6* | alpha | Mut$^{S}$ | a1-d6* | 1-4 | d6_3xMat | alpha | Mut$^{S}$ | a1-d6_3xMat | 28-29 |
| d6* | alpha | Mut$^{S}$ | a1-d6* | 5-8 | d6_3xMat | alpha | Mut$^{S}$ | a2-d6_3xMat | 30-31 |
| d6* | native | Mut$^{S}$ | n-d6* | 9-11 | d6_3xMat | native | Mut$^{S}$ | n-d6_3xMat | 32-34 |
| *AOX1* | alpha | Mut$^{S}$ | a1-AOX1 | 15-17 | *GAP* | alpha | Mut$^{S}$ | a2-GAP | 35-38 |
| *AOX1* | alpha | Mut$^{S}$ | a2-AOX1 | 22-24 | *GAP* | native | Mut$^{S}$ | n-GAP | 39-41 |
| *AOX1* | native | Mut$^{S}$ | n-AOX1 | 25-27 | AOX176-Rap1 | alpha | Mut$^{S}$ | a2-Rap1 | 42-45 |
| d1$^{+}$ | alpha | Mut$^{S}$ | a1-AOX1 | 12-13 | AOX176-Rap1 | alpha | Mut$^{S}$ PDI | a2-Rap1-PDI | 54-57 |
| d1$^{+}$ | alpha | Mut$^{S}$ | a2-AOX1 | 18-21 | *GAP* | alpha | Mut$^{S}$ PDI | a2-GAP-PDI | 58-59 |
| d1$^{+}$ | native | Mut$^{S}$ | n-AOX1 | 14 | d6* | alpha | Mut$^{S}$ PDI | a2-d6*-PDI | 46-49 |
| | | | | | d1$^{+}$ | alpha | Mut$^{S}$ PDI | a2- d1$^{+}$-PDI | 50-53 |

After 60 hours of batch growth and glucose depletion (derepression conditions) in deep well plates clones based on promoter variant AOX176-Rap1 performed best, with the best clone reaching 33,5 EU ml$^{-1}$. In contrast, but as expected, low to no activity (0,6 EU ml$^{-1}$) was found employing the *AOX1*

promoter in a methanol free system. Differently, but also expected, constructs based on the *GAP* promoter showed good activities. The best *GAP* promoter clone reached 16,2 EU $ml^{-1}$. Clones based on the promoter variants d6* and $d1^+$ reached 12,5 and 9,9 EU $ml^{-1}$, respectively. Being induced by derepression promoter variant d6* confirmed the expectations. In contrast, promoter variant $d1^+$ surprised by its relatively high esterase yield without methanol. In summary, the best methanol free clones were predominantly based on the promoters *GAP* and AOX176-Rap1 in combination with the alpha factor signal sequence fused to cleavage site 2 (Figure 4.3).

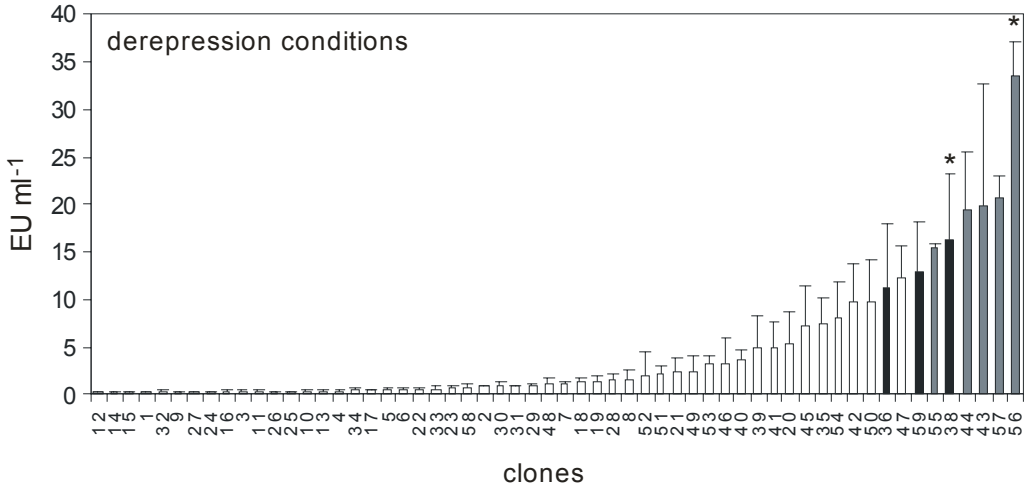



Figure 4.3 BfAChE expression employing different promoter variants – microscale rescreening results after 60 hours of growth and glucose depletion (derepression conditions), emphasized black: best *GAP* promoter based clones, emphasized grey: best AOX176-Rap1 promoter based clones, white: clones based on different promoters variants - see Table 4.1, asterisk: clones selected for bioreactor cultivation, standard deviation: 4 individual cultivations

After 48 hours of methanol induced rescreening clones based on promoter variant d6* performed best. The best clone of this construct reached 23 EU $ml^{-1}$ (Figure 4.4). In comparison, the best *AOX1* promoter based clone reached only 10 EU $ml^{-1}$. Analyzing the employed signal sequences in regard to the level of expression no preferences were found. In fact, good clones were obtained for all 3 signal sequence/esterase fusions, which indicates rather the number of clones screened and with that copy number critical. For example the best obtained n-d6* clone (d6* promoter based, native leader/esterase clone, for abbreviations see Table 4.1) reached 23 EU $ml^{-1}$, but also the best a2-d6* clone reached a level of 21 EU $ml^{-1}$. Interestingly and in agreement to literature the native leader sequence led to the best results in combination with the wild type *AOX1* promoter reaching 10 EU $ml^{-1}$ (Morel and Massoulie, 1997). In contrast, the best a2-AOX1 clone reached 6 EU $ml^{-1}$ only and the best a1-AOX1 clone even less with 3 EU $ml^{-1}$. Thus, it seems that not only copy number, but also the promoter choice influenced the results for the different signal sequences. Similar as for the signal sequences no significant differences were found between clones based on strain $Mut^S$ or the $Mut^S$-PDI strain mixture.

Comparing the methanol induced and the methanol free cultivations similar maximal activities were reached, indicating the methanol free process competitive in this case.

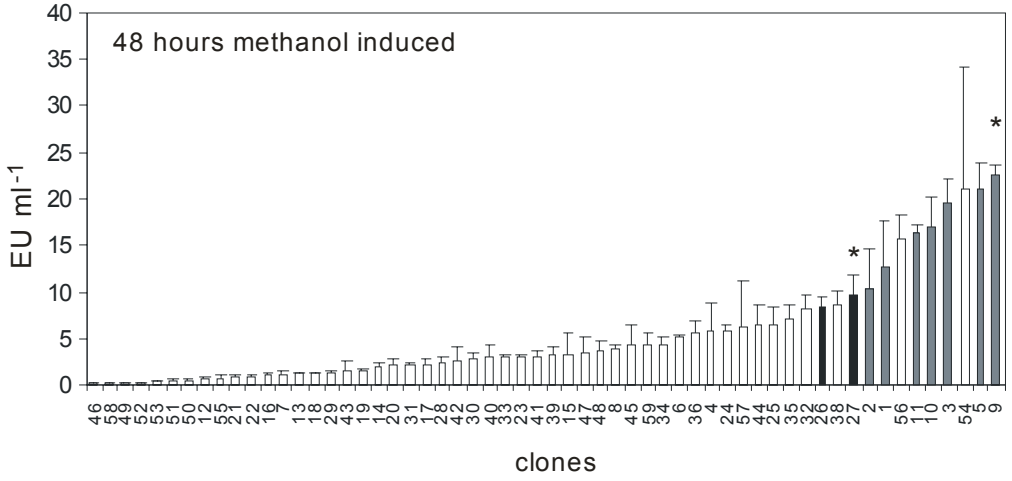



Figure 4.4 BfAChE expression employing different promoter variants – microscale rescreening results after 48 hours of methanol induction, emphasized black: best *AOX1* promoter based clones, emphasized grey: best d6* promoter based clones, white: clones based on different promoter variants - see Table 4.1, asterisk: clones selected for bioreactor cultivation, standard deviation: 4 individual cultivations

### 4.3.1 Bioreactor cultivations

While the best d6* and *AOX1* based clones (clone number 9 and 27) were analyzed in methanol induced bioreactor studies, the best clones based on the promoters *GAP* and AOX176-Rap1 (clone number 38 and 56) were analyzed methanol free applying 6,7 g $l^{-1}$ $h^{-1}$ glycerol in the production phase. Clones based on the promoters $d1^{+}$ and d6_3xMat were excluded from further studies as both were outranked in expression by d6* and AOX176-Rap1 promoter based clones. As the influence of *AOX1* driven PDI co-expression in a methanol free system is non-existent, clone a2-Rap1-PDI is in the following referred to as clone a2-Rap1.

Surprisingly, results were in strong contrast to the data obtained from rescreening (Figure 4.5). After 70 hours of methanol induction clone n-AOX1 outranked clone n-d6* 2-fold reaching 350 EU $ml^{-1}$ (26 U $ml^{-1}$). Further, after 110 hours of methanol free cultivation clone a2-GAP reached 173 EU $ml^{-1}$ (13 U $ml^{-1}$), which was 3 times more in comparison to clone a2-Rap1. In addition, reaching only 50% of the $P_{AOX1}$ level the methanol free processes were not competitive. Analyzing the expression profiles of the clones a2-GAP and a2-Rap1, low activities until 65 hours of cultivation were found. At longer cultivation times a boost in expression was observed, which was surprising, especially for the *GAP* promoter based clone. In fact, the expression profile of clone a2-GAP resembled a MeOH induced profile. Reevaluation of the rescreening did not shed light on the seen discrepancies. In contrast, the methanol induced microscale cultivation protocol has proven its applicability before (expression of porcine

trypsinogen, expression of plant epoxide hydrolases, Hartner et al., 2008). For the microscale glucose batch cultivations inaccuracies might be result of the short cultivation time of 60 hours without subsequent glucose/glycerol feed. Generally, it must be said that in case of BfAChE microscale cultivation was an unreliable tool to screen for the best expression clone, which did not mimic up-scaled and controlled cultivation conditions well enough. One can only speculate that the found trimeric occurrence of the protein (discussed in the next chapter) is connected to the screening inaccuracies and altered expression profiles found in bioreactor cultivations; however evidence remains to be elucidated. It is also possible that the seen effect is a response of the cell to stress caused by expression of this esterase. In this case transcriptomics might provide useful information.

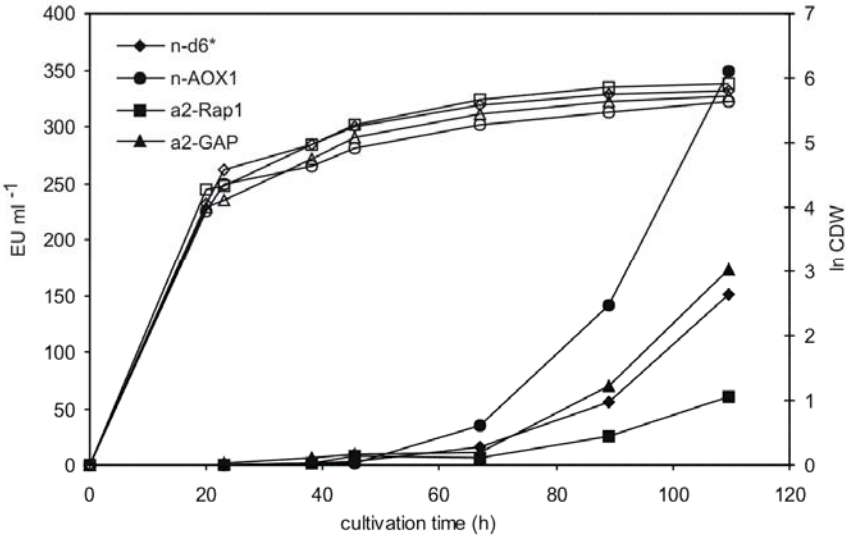



Figure 4.5 Time course of *P. pastoris* BfAChE expression employing the clones n-d6* (rhomboid, MeOH induced), n-AOX1 (circle, MeOH induced), a2-GAP (triangle, MeOH free) and a2-Rap1 (square, MeOH free), EU: Ellman units (full symbols), ln CDW: natural logarithm of the cell dry weight (empty symbols)

Protein SDS-PAGE of the culture supernatant showed a blurry smear at the expected size of 59 kDa (Figure 4.6A). However, according to earlier performed analyses of concentrated $Mut^S$ supernatant we assume *Pichia pastoris* endogenous proteins responsible. Confirming this assumption, endoglucosidase H digest did not result in a protein band of the expected nor any other size. Again these results are in contrast to results obtained from earlier shake flask cultivations, which showed little protein at about 59 kDa and intensive degradation products at lower molecular weights (Figure 4.7, Annex).
Employing Caliper lab-on-chip capillary electrophoresis the BfAChE protein peak was found at a size of approximately 210 kDa indicating the protein multimeric even under denaturing conditions (Figure 4.6B). Using this technology for protein quantification, 3 times more protein was found for the *GAP* promoter based clone, which is in agreement to the results from activity data. In detail, 260 mg $l^{-1}$ protein were estimated for clone a2-GAP after 110 hours of cultivation, while clone a2- Rap1 reached a value of 70 mg $l^{-1}$. A specific activity of 50-60 U $mg^{-1}$ was calculated for the not purified enzyme of both clones.

Native gel electrophoresis and esterase activity staining of a n-d6* BfAChE expressing clone showed esterase activity at a size of approximately 240 kDa after 20 hours of cultivation (Figure 4.6C). This confirmed the results from the capillary electrophoresis, but is in strong contrast to reported monomeric BfAChE after secretion by *P. pastoris* (Morel and Massoulie, 1997). It remains unclear why the monomeric fraction in the current study was little or even non existent. Recently also Sato et al. (2009) reported truncated carp AChE after expression and secretion by *P. pastoris* trimeric under native conditions, however they also proofed the enzyme monomeric and glycosylated after purification from edrophonium-Sepharose chromatography.

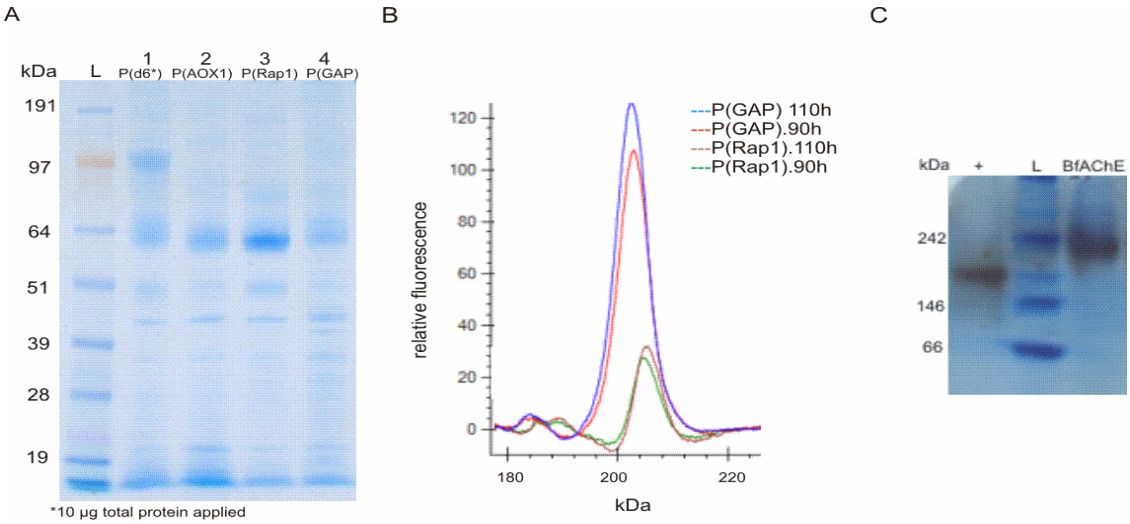



Figure 4.6 A) SDS-PAGE of *P. pastoris* supernatant after 110 hours of BfAChE expression in a bioreactor, clone variants: n-d6* (MeOH induced), n-WT (MeOH induced), a2-Rap1 (MeOH free) and a2-GAP (MeOH free), B) Caliper capillary electrophoresis of the supernatant of the clones a2-GAP and a2-Rap after 90 and 110 hours of methanol free cultivation, C) Native protein gel of a BfAChE expressing strain (concentration factor: 10, clone n-d6*, sample time: end of fed batch phase 1), esterase activity (brown) and Coomassie stain (blue), + (positive control): esterase from porcine liver (3x~60000 kDa), L: NativeMark$^{Tm}$

## 4.4 Conclusions

In this study BfAChE was successfully expressed in *P. pastoris*. Protein analysis showed the protein multimeric independent of the promoter used. Best results were obtained with a combination of the native signal sequence and the *AOX1* wild type promoter reaching a level of 350 EU ml$^{-1}$. In comparison to Morel and Massoulie (1997), who reached 35 EU ml$^{-1}$, the volumetric activity was improved 10-fold; however, Morel and Massoulie used shake flasks only. Analyzing protein titers, employment of clone a2-GAP resulted in a maximum level of 260 mg l$^{-1}$. As clone n-AOX1 gave the best results regarding volumetric activity one can speculate that protein titers might even be higher for this clone. In comparison Morel and Massoulie (1997) reported a level of 2 mg l$^{-1}$ only. Discrepancies might be the

result of multimeric AChE found in this study in comparison to monomeric protein reported by Morel and Massoulie (1997).

Regarding the use of different promoters it is possible that as a result of the inaccurate microscale screening and rescreening good clones were not identified as hit. Thus, the limits in expression reachable through promoter tuning might not have been reached. As finally the *AOX1* promoter based clone performed best, clones based on even stronger promoters, such as $d1^{+}$, are potentially interesting for bioreactor experiments. Further, the influence of the PDI chaperone was possibly also biased on microscale. Thus, clones co-expressing PDI, both, methanol free and methanol induced, might in addition be interesting. Also a comparison between the native and *Pichia pastoris* optimized coding sequence can give new insights as according to literature the native BfAChE coding sequence exerts a context dependent stimulatory effect on protein yields (Morel and Massoulie, 2000).

## 4.5 References


Abad, S., Nahalka, J., Bergler, G., Arnold, S.A., Speight, R., Fotheringham, I., Nidetzky, B., Glieder, A., (2010a) Stepwise engineering of a Pichia pastoris D-amino acid oxidase whole cell catalyst. Microb Cell Fact 9, 24.

Abad, S., Kitz, K., Hörmann, A., Schreiner, U., Hartner, F.S., Glieder, A., (2010b) Real-time PCR-based determination of gene copy numbers in Pichia pastoris. Biotechnol J 5, 413-420.

Boublik, Y., Saint-Aguet, P., Lougarre, A., Arnaud, M., Villatte, F., Estrada-Mondaca, S., Fournier, D., (2002) Acetylcholinesterase engineering for detection of insecticide residues. Protein Eng, 15, 43-50.

Cousin, X., Bon, S., Duval, N., Massoulie, J., Bon, C., (1996) Cloning and expression of acetylcholinesterase from Bungarus fasciatus venom. A new type of cooh-terminal domain; involvement of a positively charged residue in the peripheral site. J Biol Chem 271, 15099-15108.

Cousin, X., Bon, S., Massoulie, J., Bon, C., (1998) Identification of a novel type of alternatively spliced exon from the acetylcholinesterase gene of Bungarus fasciatus. Molecular forms of acetylcholinesterase in the snake liver and muscle. J Biol Chem 273, 9812-9820.

Ellman, G.L., Courtney, K.D., Andres, V.Jr., Feather-Stone, R.M., (1961) A new and rapid colorimetric determination of acetylcholinesterase activity. Biochem. Pharmacol. 7, 88-95.

Hartner, F.S., Glieder A., (2005) Mutant AOX1 promoters, WO2006089329.

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.

Jiang, H., Zhang, X.J., (2008) Acetylcholinesterase and apoptosis. A novel perspective for an old enzyme. Febs J 275, 612-617.

Liu, Z., Pscheidt, B., Avi, M., Gaisberger, R., Hartner, F.S., Schuster, C., Skranc, W., Gruber, K., Glieder, A., (2008) Laboratory evolved biocatalysts for stereoselective syntheses of substituted benzaldehyde cyanohydrins. Chembiochem 9, 58-61.

Massoulie, J., Pezzementi, L., Bon, S., Krejci, E., Vallette, F.M., (1993) Molecular and cellular biology of cholinesterases. Prog Neurobiol 41, 31-91.

Massoulie, J., Anselmet, A., Bon, S., Krejci, E., Legay, C., Morel, N., Simon, S., (1999) The polymorphism of acetylcholinesterase: post-translational processing, quaternary associations and localization. Chem Biol Interact. 119-120, 29-42.

Morel, N., Massoulie, J., (1997) Expression and processing of vertebrate acetylcholinesterase in the yeast Pichia pastoris. Biochem J 328 ( Pt 1), 121-129.

Morel, N., Massoulie, J., (2000) Comparative expression of homologous proteins. A novel mode of transcriptional regulation by the coding sequence folding compatibility of chimeras. J Biol Chem 275, 7304-12.

Oka, A., Sugisaki, H., Takanami, M., (1981) Nucleotide sequence of the kanamycin resistance transposon Tn903. J Mol Biol. 147, 217-26.

Ruth, C., Zuellig, T., Mellitzer, A., Weis, R., Looser, V., Kovar, K., Glieder, A., (2010) Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in *Pichia pastoris.* Syst Synth Biol. in press.

Sato, R., Matsumoto, T., Hidaka, N., Imai, Y., Abe, K., Takahashi, S., Yamada, R.H., Kera, Y., (2009) Cloning and expression of carp acetylcholinesterase gene in Pichia pastoris and characterization of the recombinant enzyme. Protein Expr Purif 64, 205-212.

Selwood, T., Feaster, S.R., States, M.J., Pryor, A.N., Quinn, D.M., (1993) Parallel Mechanisms in Acetylcholinesterase-Catalyzed Hydrolysis of Choline Esters. J. Am. Chem. SOC. 115, 0477-10482.

Simon, S., Massoulie, J., (1997) Cloning and expression of acetylcholinesterase from Electrophorus. Splicing pattern of the 3' exons in vivo and in transfected mammalian cells. J Biol Chem. 272, 33045-55.

Sussman, J.L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L., Silman, I., (1991) Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein. Science 253, 872-879.

Taylor, P., Reiner, E., Kovarik, Z., Radić, Z., (2007) Application of recombinant DNA methods for production of cholinesterases as organophosphate antidotes and detectors. Arh Hig Rada Toksikol. 58, 339-45.

Wang, W., Malcolm, B.A., (1999) 2-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. Biotechniques 26, 680-682.

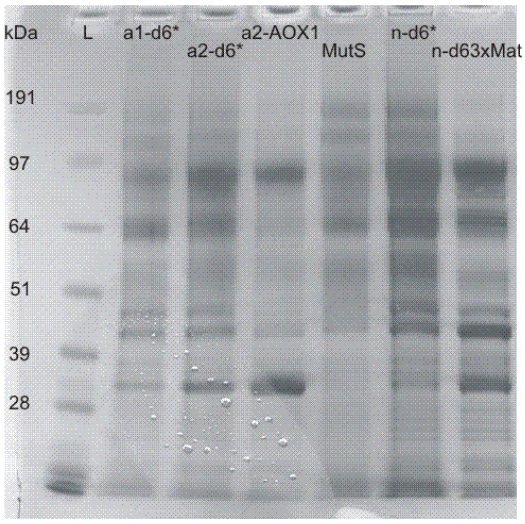
Weill, C.O., Vorlova, S., Berna, N., Ayon, A., Massoulie, J., (2002) Transcriptional regulation of gene expression by the coding sequence: An attempt to enhance expression of human AChE. Biotechnol Bioeng 80, 490-7.

Weis, R., Luiten, R., Skranc, W., Schwab, H., Wubbolts, M., Glieder, A., (2004) Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena. FEMS Yeast Res 5, 179-189.

Wu, A.B., Chen, H.D., Tang, Z.Z., Ye, B.W., Liu, W.J., Jia, H.Y., Zhang, D.B., (2008) Synthesis of Drosophila melanogaster acetylcholinesterase gene using yeast preferred codons and its expression in Pichia pastoris. Chem Biol Interact 175, 403-405.

## 4.6 Annex

A




B

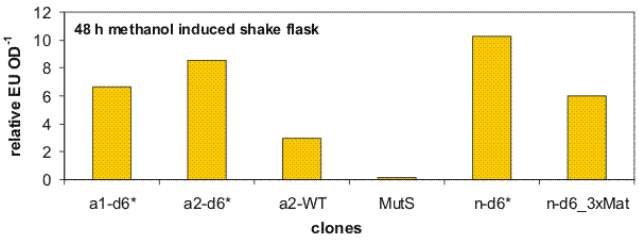



Figure 4.7 A) SDS-PAGE of *P. pastoris* supernatant after 48 hours of methanol induced BfAChE expression in shake flask, clone variants: a1-d6* (rescreening clone no. 1*), a2-d6* (rescreening clone no. 5), a2-AOX1 (no. 47), n-d6* (no. 11) and n-d6_3xMat (no. 34), B) Corresponding relative activity units per OD (relative EU $OD^{-1}$), applied to the gel: ~20-times concentrated supernatant

* numbers according to Table 4.1

Table 4.2 Plasmids coding for AChE expression (Bf: *B. fasciatus*, Ee: *E. electricus*, Dv: *D. viviparus*)

| strain coll. | plasmid | host |
|---|---|---|
| 5860 | pJetB1_pBfAChE a2WT | K12 DH5α |
| 5861 | pJetB1_pBfAChE a1WT | K12 DH5α |
| 5864 | pJetB1_pBfAChE ntWT | K12 DH5α |
| 5862 | pJetB1_pBfAChE a1d63xMat | K12 DH5α |
| 6091 | pJetB1_BfAChEa 2d6* | K12 Top10 F' |
| 5863 | pJetB1_pBfAChE a1d6* | K12 DH5α |
| 6090 | pJetB1_BfAChE $a2d1^{+}$ | K12 Top10 F' |
| 5882 | pJetB1_pBf AChE $ntd1^{+}$ | K12 DH5α |
| 6089 | pJetB1_BfAChE a2Gap | K12 Top10 F' |
| 5744 | puc57_AChE_Ee | K12 DH5α |
| 5745 | puc57_AChE_Bf | K12 DH5α |
| 5746 | puc57_AChE_Dv1 | K12 DH5α |
| 5747 | puc57_AChE_Dv2 | K12 DH5α |
| 6097 | pJet/B1_EeAChE_PAOX1WT | K12 Top10 F' |

Table 4.3 *Bungarus fasciatus* AChE *P. pastoris* expression strains

| strain coll. | host | strain | clone | plasmid |
|---|---|---|---|---|
| 6129 | CBS7435 Mut$^{S}$ | BfAChE_ntd6* | 9 | cas_B1 |
| 6130 | CBS7435 Mut$^{S}$ | BfAChE_ntd6* | 10 | cas_B1 |
| 6131 | CBS7435 Mut$^{S}$ | BfAChE_ntWT | 27 | cas_B1 |
| 6132 | CBS7435 Mut$^{S}$ | BfAChE_a2Gap | 38 | cas_B1 |
| 6133 | CBS7435 Mut$^{S}$PDI | BfAChE_a2RapPDI | 56 | cas_B1 |
| 6134 | CBS7435 Mut$^{S}$ | BfAChE_a2Rap | 44 | cas_B1 |
| 6135 | CBS7435 Mut$^{S}$PDI | BfAChE_a2RapPDI | 57 | cas_B1 |
| 6105 | CBS7435/Mut$^{S}$ | AOX_noEH_N* | x | cas_B1 |

*negative control without gene of interest

Table 4.4 Primers used for OE-PCR of the AChE expression cassettes

| primer coll. | name | sequence |
|---|---|---|
| P08225 | Aoxlongfw | AGATCTAACATCCAAAGACGAAAGG |
| P08207 | pAOXshortrv | CGTGAAGTCCTCGTTTCGAAGTACCTAGTTTCAATAATTAGTTGTTTTTTG |
| P08206 | potAlphafw | CTAGGTACTTCGAAACGAGGACTTCACGATGAGATTCCCATCTATTTTC |
| P08432 | alpha_rv | AGCTTCGGCCTCTCTCTTCTCGAGAGAG |
| P08427 | Bf_n_fw_ov1 | CTAGGTACTTCGAAACGAGGACTTCACGATGCCATCTTGTCAACCTGGTAAG |
| P08429 | Bf_t_rv_ov2 | GACTGAGTGACCTCGCTATCTGACTTAGGTAGCGTTCAGAAGCTTAGG |
| P08430 | a_ol_1_fw | CTCTCGAGAAGAGAGAGGCCGAAGCTGTTTTGCCAGGTAGAGCCGGTG |
| P08431 | a_ol_2_fw | CTCTCGAGAAGAGAGAGGCCGAAGCTGGTGAGTTGAAGGTTTCCACTCAG |
| P08212 | ZeoLinkfw | GTCAGATAGCGAGGTCACTCAGTCTCAAGAGGATGTCAGAATGCC |
| P08226 | Zeorv | TTGTCCTCTGAGGACATAAAATAC |
| P07662 | gap1_r | CGTGAAGTCCTCGTTTCGAAGTACCTAGATAGTTGTTCAATTGATTGAAATAG |
| P08332 | P(GAP)fw | GTCTTGGTGTCCTCGTCCAATCAGG |

# List of Figures

**Figure 1.1** Prokaryotic promoter libraries with different promoter strength can be generated by varying the sequence surrounding of the _10 and _35 consensus boxes (shuffling, point mutations, N: A, T, C, G) or the length of the spacer sequence (spacer length), the promoter regulation can be influenced by changes in the consensus boxes

**Figure 1.2** Eukaryotic promoter library approaches, A) Promoter library generation and sequence analysis by deletion of putative transcription-factor binding sites (rectangles depict the schematic binding of transcription factors to a promoter sequence, TS: transcription start), B) Synthetic promoter generation by fusion of sequence randomers or identified cis-acting elements to core promoters (synthetic promoters: P1 and P2), C) Different employed regulatory elements (P1 and P2) can alter a promoters mode of regulation and consequently the expression profile

**Figure 1.3** Synthetic bidirectional promoter construct, P(specific): full length specific promoter driving the expression of the reporter gene fused to expression boosting activator binding sites (ABS), P(core): short core promoter driving the expression of the activator gene, E(bi): bidirectional activating elements stimulating the activity of P(specific) and P(core)

**Figure 1.4** Pichia pastoris shuttle vector pPpT4 (3546 bp), P AOX1: AOX1 promoter, AOX1TT: AOX1 terminator, MCS (multiple cloning site): EcoRI, SpeI, AscI, NotI, P ILV5: promoter of the P. pastoris ILV5 gene, P EM72*: synthetic E. coli promoter, Zeocin Syn: synthetic codon optimized Zeocin gene, AODTT: terminator of the P. pastoris AOD gene, pUC ori: origin of replication

**Figure 1.5** Synthetic core promoter 1 (consensus based design), underlined: restriction sites EcoRI and BglII

**Figure 1.6** Synthetic core promoter 11 (TF based design), underlined: restriction sites EcoRI and BglII

**Figure 1.7** PAOX1 deletions and insertion variants, WT: wild type AOX1 promoter and corresponding numeration, blue: Mat1-Mc region, green: region between d6 and element 201-214 (including deletion *: ΔTA), red: element 201-214 and point mutations, brown: transcription start site (TSS/A-113) and the putative TATA box (MatInspector prediction)

**Figure 1.8** GFP expression after 60 h of batch growth and glucose depletion (derepression conditions, 0 h) employing next generation PAOX1 deletion and insertion variants (microplate screening), mean value of 46 clones

**Figure 1.9** Methanol induced GFP expression (48 h) employing next generation PAOX1 deletion and insertion variants (microplate screening), mean value of 46 clones

**Figure 1.10** Rescreening results of the putative single copy (sc) clones after 24 hours of methanol induction (green) in comparison to the average single copy level obtained in screening after 24 hours of induction (grey)

**Figure 1.11** A) Scheme for the generation of the short synthetic promoter variants, squares: cis-acting elements, arrows: primer, applied primer binding regions: ZUS (dotted line), 201 (AOX201), 176 (AOX176), B) PAOX1 cis-acting elements (underlined) used for synthetic promoter generation, grey and dark grey: primer binding regions ZUS (fw2), 201 (fw1) and 176 (fw1), bold capital letters: TF binding core similarities (MatInspector, Genomatix, Hartner et al., 2008), red: TSS located at position -113 and the putative TATA box (MatInspector) located between -155/-158 relative to the AOX1 start codon

**Figure 1.12** GFP expression after 60 h of batch growth and glucose depletion (derepression conditions, 0 h) employing the new short synthetic promoter variants (microplate screening), mean value of minimum 50 clones
**Figure 1.13** Methanol induced (48 h) GFP expression employing the new short synthetic promoter variants (microplate screening), mean value of minimum 50 clones
**Figure 1.14** A) Sequence alignment of PpPGAP, PpPHIS4, PpPAOX1 and ScPADH2 used for the generation of the consensus promoter core1, B) Sequence comparison of Pcore1 and PAOX1, C) Sequence comparison of Pcore1 and Pcore11
**Figure 1.15** MatInspector transcription factor analyses of the promoters A) Pcore1 (consensus based) and B) Pcore11 (TF optimized)
**Figure 1.16** Relative GFP fluorescence after 48 hours of methanol induction employing the artificial promoters core1, core1-Z, core11, core11-Z, red: MutS background
**Figure 1.17** Relative GFP fluorescence after 48 hours of methanol induction employing the artificial promoters core11, core11-Z and core11-ZMM, red: MutS background
**Figure 1.18** Relative GFP fluorescence after 60 hours of batch growth and glucose depletion (derepression conditions, 0 h) employing the artificial promoter core11-ZMM, red: MutS background
**Figure 1.19** Super core promoter 1 (SCP1) adopted from Juven-Gershon et al. (2006a), bold: TATA-box (from CMV IE1 core promoter), red: TSS, underlined: Inr (based on sequences from AdML and D. melongaster G retrotransposon core promoters), blue: MTE (from D. melongaster Tollo core promoter), green: DPE (from Drosophila G core promoter)
**Figure 1.20** Screening results of the 6 for rescreening selected clones of promoter variant d6_3xMat in comparison to the best WT promoter based clone, relative fluorescence units (RFU) at 0, 24 and 48 hours of methanol induction
**Figure 1.21** Screening results of the 6 for rescreening selected clones of promoter variant d6_2xMat_2x201 in comparison to the best WT promoter based clone, relative fluorescence units (RFU) at 0, 24 and 48 hours of methanol induction
**Figure 1.22** Screening results of clones based on promoter variant AOX176-RR in comparison to the best WT promoter based clone, relative fluorescence units (RFU) at 0 and 48 hours of methanol induction
**Figure 1.23** Screening results of clones based on promoter variant AOX176-MM in comparison to the best WT promoter based clone, relative fluorescence units (RFU) at 0 and 48 hours of methanol induction
**Figure 2.1** Pichia pastoris shuttle vector pPpT2 (3546 bp), P AOX1: AOX1 promoter, AOX1TT: AOX1 terminator, MCS (multiple cloning site): EcoRI, SpeI, AscI, NotI, P ILV5: promoter of the P. pastoris ILV5 gene, P EM72: synthetic E. coli promoter, Zeocin Syn: synthetic Zeocin gene, AODTT: terminator of the P. pastoris AOD gene, pUC ori: origin of replication
**Figure 2.2** Sequence alignment of the AOX1 wild type (WT) promoter and AOX1 promoter variants, a) wild type PAOX1 (WT), deletion variants (d6, d1), double deletion variants (d6*, dHap2345-1z1), b) synthetic variants AOX176-Rap1 (Rap1), AOX176-Mat1-Mc (Mat1) and AOX176-201-214 (214); underlined: EcoRI, BglII, BspTI restriction sites, (-1): 3' promoter end, rectangle: cis-acting elements

(Rap1, Mat1-Mc, 201-214)

**Figure 2.3** Expression of porcine trypsinogen using AOX1 promoter variants, A – growth in batch with glucose, B – growth in fedbatch with glucose, C – production phase in fedbatch with methanol, open symbols – logarithm of cell dry weight, closed symbols – trypsinogen activity, promoters: d1+ (squares), d6* (circles), AOX176-Rap1 (triangles)

**Figure 2.4** Protein chip analyses of the methanol induced fedbatch cultures with the promoters d6* (a, b) and AOX176-Rap1 (c, d); a, c gel images visualizing promoter dependent trypsinogen degradation upon a certain production time; b, d corresponding fluorescence detection based chromatograms used for protein concentration estimation at different cultivation times, RFU: relative fluorescence units

**Figure 2.5** Expression of porcine trypsinogen using synthetic promoters, A – growth in batch with glucose, B – growth in fedbatch with glucose, C – production phase in fedbatch with glucose, open symbols – logarithm of cell dry weight, closed symbols – trypsinogen activity, AOX176-Rap1 variants: G3 (triangles), A7 (circles), B3 (squares), C8 (rhomboids)

**Figure 2.6** Specific trypsinogen activities in fedbatch cultures, left - production phase with methanol (WT, d1+, d6*), right - production phase with glycerol or glucose (AOX176-Rap1)

**Figure 2.7** Product yields obtained in fedbatch cultures with different promoters (WT, d1+, d6*, AOX176-Rap1) applying different carbon sources in the production phase: methanol (black), glycerol (grey), glucose (white), protein concentration determination: Caliper capillary electrophoresis

**Figure 3.1** A) Specific whole cell epoxide hydrolase activities of the best obtained P. pastoris clones (substrate TSO, shake flask cultivations), epoxide hydrolases: StEH1, GmEH1, ElEH1, B) Comparing P. pastoris and E. coli for recombinant StEH1 expression, whole cell conversion of TSO

**Figure 3.2** A) Temperature and B) pH optimum of spurge soluble epoxide hydrolase, substrate: TSO (final conc. 1 mM), % of maximum activity: values were normalized to the maximal activity (µM min-1), inserts: half lives of spurge soluble epoxide hydrolase at A) different temperatures and B) different pH values

**Figure 3.3** A) Time course of PAOX1 driven potato sEH expression in a methanol induced bioreactor cultivation, I: glycerol batch phase, II: glycerol fed batch phase, III: methanol fed batch phase (production phase), U ml-1: whole cell conversion of TSO (full squares), ln (DCW): natural logarithm of the dry cell weight (empty squares), B) SDS-page of SpinTrap purified potato sEH, lane 1: cell lysate, lane 3: purified potato sEH; 10 µg total protein applied

**Figure 3.4** Microscale rescreening results of the different promoter variants driving potato sEH expression, % TSO conversion: whole cell conversion of TSO normalized to StEH1WT, promoter variants: d1+, d6*

**Figure 3.5** Time course of d1+ driven potato sEH expression in a methanol induced bioreactor cultivation, I: glycerol batch phase, II: glycerol fed batch phase, III: methanol fed batch phase (production phase), U ml-1: whole cell conversion of TSO (full squares), ln (DCW): natural logarithm of the dry cell weight (empty squares)

**Figure 3.6** Mating factor α signal sequence

**Figure 4.1** A) Alignment of T. marmorata (T.m.) and B. fasciatus (B.f.) AChE adopted from (Cousin et al., 1996), circled: catalytic triad S231, E358, and H471, connected boxes: internal disulfide bridges, black dots: positions of the BfAChE residues 101 and 316, which influence the properties of the peripheral site (see text), B.f.: S-subunit, T.m.: H- and T-subunit B) 3D structure of AChE from T. californica, which was used to propose a 3-dimensional model for BfAChE (Cousin et al., 1996, Massoulie et al., 1993; Sussman et al., 1991)

**Figure 4.2** BfAChE expression constructs, promoters used: A) AOX1, d1+, d6*, d6_3xMat, GAP, B) AOX176-Rap1, signal sequences: alpha ss (a), native ss (n); VLP, GEL: N-terminal sequences of the mature peptides after cleavage at the first (a1) or alternative (a2) predicted signal sequence cleavage site, NAT: C-terminus (truncation after catalytic domain), AOX1TT: AOX1 terminator, link 1 and 2: linker 1 and 2 for OE-PCR

**Figure 4.3** BfAChE expression employing different promoter variants – microscale rescreening results after 60 hours of growth and glucose depletion (derepression conditions), emphasized black: best GAP promoter based clones, emphasized grey: best AOX176-Rap1 promoter based clones, white: clones based on different promoters variants - see Table 4.1, asterisk: clones selected for bioreactor cultivation, standard deviation: 4 individual cultivations

**Figure 4.4** BfAChE expression employing different promoter variants – microscale rescreening results after 48 hours of methanol induction, emphasized black: best AOX1 promoter based clones, emphasized grey: best d6* promoter based clones, white: clones based on different promoter variants - see Table 4.1, asterisk: clones selected for bioreactor cultivation, standard deviation: 4 individual cultivations

**Figure 4.5** Time course of P. pastoris BfAChE expression employing the clones n-d6* (rhomboid, MeOH induced), n-AOX1 (circle, MeOH induced), a2-GAP (triangle, MeOH free) and a2-Rap1 (square, MeOH free), EU: Ellman units (full symbols), ln CDW: natural logarithm of the cell dry weight (empty symbols)

**Figure 4.6** SDS-PAGE of P. pastoris supernatant after 110 hours of BfAChE expression in a bioreactor, clone variants: n-d6* (MeOH induced), n-WT (MeOH induced), a2-Rap1 (MeOH free) and a2-GAP (MeOH free), B) Caliper capillary electrophoresis of the supernatant of the clones a2-GAP and a2-Rap after 90 and 110 hours of methanol free cultivation, C) Native protein gel of a BfAChE expressing strain (concentration factor: 10, clone n-d6*, sample time: end of fed batch phase 1), esterase activity (brown) and Coomassie stain (blue), + (positive control): esterase from porcine liver (3x~60000 kDa), L: NativeMarkTm

**Figure 4.7** A) SDS-PAGE of P. pastoris supernatant after 110 hours of BfAChE expression in bioreactor, clone variants: a1-d6* (rescreening clone no. 1*), a2-d6* (rescreening clone no. 5), a2-AOX1 (no. 47), n-d6* (no. 11) and n-d6_3xMat (no. 34), B) Corresponding relative activity units per OD (relative EU OD-1), applied to the gel: ~20-times concentrated supernatant

# List of Tables

**Table 1.1** Summary of the new generation promoter variants II, Z: linker region ZUS

**Table 1.2** SDM and PCR primers, bold: positive acting elements, italic: 3' element extensions (see 1.2.4.2), underlined: region ZUS (Z), which links cis-acting elements to core promoters

**Table 1.3** Plasmids coding for GFP expression under the control of different synthetic promoters; Pichia pastoris strains are available as glycerol stocks only

**Table 2.1** Relative promoter comparison, activities of the best clones from rescreening are given in relation to the value of the WT-strain at 96 hours (i.e. 100%) and the time of methanol induction - resp. time 0 h at the end of glucose batch; standard deviations were calculated using data from 4 different wells

**Table 2.2** P. pastoris trypsinogen expression strains with corresponding copy numbers

**Table 3.1** Potato, soy bean and spurge soluble epoxide hydrolase P. pastoris expression strains

**Table 3.2** Plasmids coding for potato, soy bean and spurge soluble epoxide hydrolase expression

**Table 4.1** AChE expressing P. pastoris clones evaluated in rescreening, ss: signal sequence, alpha: alpha mating factor ss, native: native BfAChE ss, a1 and a2 refer to the N-terminal sequence of the mature peptide after cleavage at the first or alternative predicted ss cleavage site

**Table 4.2** Plasmids coding for AChE expression (Bf: B. fasciatus, Ee: E. electricus, Dv: D. viviparus)

**Table 4.3** Bungarus fasciatus AChE P. pastoris expression strains

**Table 4.4** Primers used for OE-PCR of the AChE expression cassettes