Novel episomal expression systems for *Pichia pastoris*

MASTER'S THESIS

to achieve the university degree of

Diplom-Ingenieur

Master's degree programme: Biotechnology

submitted to

Graz University of Technology

Supervisor

Univ.-Prof. Dipl.-Ing. Dr.techn. Helmut Schwab

Institute of Molecular Biotechnology

Graz, December 2015

AFFIDAVIT

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 indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master‘s thesis dissertation.

December 14, 2015
Date

Signature

Acknowledgements

I would like to thank my supervisor Prof. Helmut Schwab for giving me the opportunity to work on this great project.

Big thanks also to Prof. Anton Glieder and his company bisy e.U., who made the project possible. Thank you for the support during the project and for letting me realize my own ideas.

I would also like to thank my family, especially my parents, for the support during the whole time, not just during the work on the thesis.
Thank you for your personal and also financial support.

Big thanks also to my “old” and “new” friends, who made my time in Graz unforgettable and for all the nice memories we share.

Thomas Vogl also deserves big thanks for his support, especially in the beginning of my work and for critically reading my thesis. But also for his supervision during my project lab at the institute.

I would also like to say thank you to Amneris, who made the life in the lab easier. You were a great support.

Last but not least I would like to thank the whole “Glieder group” for their support during the thesis and of course - even more - for the funny time we had inside and outside of the lab. Even when the lab work was frustrating sometimes you made the time still enjoyable.

Abstract

The methylotrophic yeast *Pichia pastoris* is widely used for heterologous protein production. Due to the rising importance of this expression host the availability of genetic tools is steadily increasing. Commonly used expression vectors are designed for integration into the genome, which leads to stable strains even under the absence of selection pressure. But genomic integration also leads to relatively low transformation efficiencies and heterogeneity between transformants because of multicopy integration and integration in different loci. Also the recovery of the expression cassette is laborious.

An alternative to genomic expression is the use of autonomously replicating sequences (ARSs) that enable episomal expression. ARS vectors show high transformation efficiencies, more uniform and also sometimes higher expression levels than integrative expression vectors and can be easily recovered. For standard protein expression ARS plasmids, mostly based on the *Pichia autonomously replicating sequence 1* (*PARS1*) have seldom been used because of instability issues. However, on a few occasions ARS plasmids were applied for the screening of libraries.

In this thesis an episomal expression kit based on a new heterologous ARS derived from *Candida boidinii* is reported. It comprises two sets of plasmids designed for antibiotic and auxotrophic selection including corresponding knockout strains. Both sets of plasmids offer different promoters for intracellular and secretory expression and are designed for state of the art seamless cloning using type IIS restriction enzymes.

The reported ARS also terminates transcription efficiently giving the possibility to minimize vector size by exchanging the terminator of the selection marker with the ARS. It exhibits even higher transformation efficiency and expression levels than *PARS1* making it the ideal candidate for library screening in protein and promoter engineering approaches and with auxotrophic selection maybe also for larger scale productions.

Kurzfassung

Die methylotrophe Hefe *Pichia pastoris* ist ein wichtiger Organismus für heterologe Proteinproduktion. Durch die zunehmende Bedeutung hat sich auch die Verfügbarkeit von genetischen Werkzeugen ständig erweitert. Typischerweise werden integrative Expressionsvektoren verwendet, die zu stabilen Expressionsstämmen führen, aber auch aufgrund unterschiedlicher Kopienzahl und Integrationsloci relativ niedrige Transformationsraten und heterogene Expressionslevels aufweisen. Auch die Wiedergewinnung der Expressionskassette ist aufwändig.

Eine Alternative zur genomischen Expression stellt die Verwendung autonom replizierender Sequenzen (ARSs) dar, die episomale Expression ermöglichen. ARS-Vektoren zeigen hohe Transformationsraten, sehr einheitliche und manchmal auch höhere Expressionslevels als integrative Vektoren. Zusätzlich können sie leicht wieder isoliert werden für weitere Experimente. Für die Standard-Proteinexpression werden diese Systeme, die meist auf der *Pichia autonom replizierenden Sequenz 1* (*PARS1*) basieren, aufgrund ihrer Instabilität jedoch selten verwendet. Für das Screening von Genbibliotheken haben sie jedoch in einigen Fällen bereits Anwendung gefunden.

In dieser Arbeit wird ein Expressionssystem vorgestellt, das auf einer neuen, heterologen ARS aus *Candida boidinii* basiert. Es besteht aus zwei verschiedenen Sets von Vektoren, die für Selektion mit Antibiotika und auxotrophe Selektion konzipiert sind inklusive der dazugehörigen auxotrophen Stämmen. Beide Vektor-Typen beinhalten verschiedene Promotoren und die Möglichkeit intrazellulärer und extrazellulärer Expression und sind für modernes Klonieren mittels Typ IIS Restriktionsenzymen ausgelegt, um keine unnötigen DNA Elemente nach der Klonierung zu hinterlassen.

Die heterologe ARS kann auch als effizienter Transkriptionsterminator verwendet werden wodurch die Plasmidgröße weiter verringert werden kann, indem der Transkriptionsterminator des Selektionsmarkers ersetzt wird. Plasmide, die auf der neuen ARS basieren, zeigen höhere Transformationseffizienzen und Expressionslevels als jene, die auf *PARS1* basieren. Das macht sie zu den idealen Kandidaten für Library Screenings im Zuge von Protein- und Promotor-Engineering Experimenten und mit auxotropher Selektion eventuell auch für Anwendungen im größeren Maßstab.

Table of Contents

1 Introduction

1.1 *Pichia pastoris* as an expression host

The methylotrophic yeast *Pichia pastoris* (*Komagataella phaffii* (1)) is a powerful protein production host since it combines different features from lower and higher organisms such as the ease of cultivation and genetic manipulation while still being able to perform post-translational modifications that are necessary for more complex proteins (2). Additionally, *P. pastoris* does not carry known pathogens, viruses or pyrogens and is able to secrete proteins with high product titers (3), while just low levels of endogenous proteins are secreted facilitating purification and making this yeast interesting for the pharmaceutical industry (2).
Very high cell densities of more than 200 g/l cell dry weight can be reached (4) as well as high concentrations of heterologous proteins. Intracellular protein yields of up to 22 g/l expressing *Hevea brasiliensis* hydroxynitrile lyase (5) and over 18 g/l of secreted protein expressing *Trichoderma reseei* cellobiohydrolase 2 (*TrCBH2*) (6) were achieved using *P. pastoris* as expression host.
The expression of the *alcohol oxidase 1* gene (*AOX1*), one of two genes responsible for the oxidation of methanol (7), is transcriptionally regulated by methanol (8). When *P. pastoris* is grown on methanol approximately 30% of the total soluble protein and 5% of the mRNA correspond to *AOX1*, while the mRNA cannot be detected when grown on repressing carbon sources (7). This tight regulation on the transcriptional level as well as the strong expression after induction made the *AOX1* promoter (P_{AOX1}) one of the most interesting and most widely used promoters beside the strong constitutive promoter of the *glyceraldehyde-3-phosphate dehydrogenase* gene (P_{GAP}), which can be used if a separation between growth and production phase is not necessary (2).
In the past years *P. pastoris* was steadily becoming more interesting for industry as well as research institutions. For research it was even much more widely used than *Saccharomyces cerevisiae* or higher eukaryotes when it comes to single protein production (9). The genome sequences of the most commonly used production strains have been published (10–12). Also a wider range of genetic tools, such as a larger set of synthetic as well as natural promoters and terminators have been reported recently (*e.g.* (13) and (Vogl T, Dissertation)). Some novel

promoters show a derepressed regulation pattern, providing the ability to avoid methanol, which is flammable and toxic and not desired in large scale production. Additionally the expression of pathways is facilitated through the availability of different co-regulated promoters with varying expression strength (Vogl T, Dissertation) and the use of self processing 2A sequences (14).

1.2 Genomic and episomal expression systems in *P. pastoris* and their application for protein engineering

Most vector systems that are available for *P. pastoris* are used for genomic integration. Integration into the genome leads to stable strains, but also to some negative effects (15). The transformation efficiency is lower for integrative plasmids, they have to be linearized prior to transformation and also the expression level of primary transformants can be quite variable and low. So many transformants have to be screened (16). The varying expression strength can be beneficial when searching for one strain with elevated expression level. But it is a clear disadvantage for the comparison between different strains during the screening of libraries or comparison of promoters.

An alternative method to the use of often large integrative vectors was applied by using overlap extension PCR to create expression cassettes carrying mutated genes and direct transformation of the linear DNA fragments, which avoids additional cloning steps. This method leads to a larger fraction of clones that is active and also less clonal variation between transformants. But still a fraction of about 20% shows significantly elevated or lower expression levels. This may be due to integration events, which can occur multiple times and also in different loci (17). Even when a specific locus is targeted by the use of large homologous regions the rate of homologous recombination still remains low and random integration still occurs (18).

Despite the efforts in making *P. pastoris* more attractive for the screening of libraries, *E. coli* is still the most widely used host for that purpose. If the protein cannot be expressed in *E. coli*, *S. cerevisiae* is the host of choice in general (15,16).

Even though there has been effort in enzyme engineering trying to minimize the screening effort when using integrative vectors (e.g. (19,20)), *S. cerevisiae* is still superior when it comes

to transformation efficiency, the ease of plasmid recovery and genetic tools like the good homologous recombination machinery (21). Another option is screening the mutant library in *S. cerevisiae* and just try to express the hits in *P. pastoris*, which might show beneficial properties (22).

A possibility to circumvent at least some of these limitations is the use of episomal expression vectors that contain an autonomously replicating sequence (ARS). Two ARSs for *P. pastoris* have been known since 1985, the *Pichia autonomously replicating sequence 1* and *2* respectively (*PARS1* and *PARS2*). Plasmids containing one of these sequences can be transformed circular and replicate autonomously in multiple copies. These plasmids were large due to the auxotrophy selection markers and depended on the use of minimal media for selection, resulting in slow growth of transformants. Higher transformation efficiencies of 10^5 CFU/µg can be reached using episomal vectors (23).

Nevertheless, these vectors were not widely used over the last decade. Early ARS vectors were used as shuttle vectors for *E. coli* and *P. pastoris* allowing easy transfer from one host to another. If the expression did not work in *E. coli* the plasmid could be simply transferred into *P. pastoris* without any further steps necessary since the gene of interest is under the control of promoters for both hosts. These vectors contained inducible promoters like P_{AOX1} (24) or the *S. cerevisiae CUP1* promoter (25) for *P. pastoris* coupled with a short bacterial promoter. The *CUP1* promoter allows a shorter induction protocol, but still the cultivation is much more time consuming than in *E. coli* (25). Vectors sold as pPICHOLI System also contain small selection markers suitable for antibiotic selection.

Later episomal expression vectors were based on P_{GAP}, which allowed shorter and easier screening and most of them were used for library screening (15,16,26).

Lee *et al.* reported the use of a novel expression vector containing *PARS1* and P_{GAP} driven secretion of a xylanase. The library was screened in *P. pastoris* since the enzyme could not be expressed in an active form neither in *E. coli* nor in *S. cerevisiae* (15). The same vector was used for directed evolution of *Candida antarctica* lipase A (*Ca*lA) (26).

Also the successful expression of cellulases was reported using a similar expression vector (27) and the expression of various human proteins using P_{GAP} driven intracellular expression from a *PARS1* containing vector (16).

Nearly all vectors are based on *PARS1* to allow episomal replication. Beside that there is another ARS that was published, which works across several yeast species and has improved stability under the absence of selection pressure (28). Stability of episomal expression strains remains the biggest issue of that system, since already approximately two third of the cells did not contain the plasmid anymore after 10 generations (16,23).
Nevertheless, the system still provides many advantages like the ease of handling, plasmid recovery, the high transformation efficiency, higher expression levels and more uniform expression (16).

1.3 Cloning systems

Most vectors for *P. pastoris* still use 'classical' cloning applying type II restriction enzymes and ligation to insert the gene of interest via a multiple cloning site (MCS). These MCSs can have a significant influence on the expression of the gene of interest. The secondary structure of the mRNA is changed, which can interfere with the ribosome scanning (29) and also the space between the promoter and the coding sequence is changed (30). Additionally to that the scars, which are left after classical cloning, can be part of the translated sequence when tags are used. Additional amino acids are added to the desired protein and can influence its properties (31).
In order to evaluate the influence of MCSs Crook *et al.* cloned various reporter genes with the different restriction sites present in the expression vector. There was a 5-fold difference in reporter protein expression when using different restriction enzymes. The use of restriction sites further downstream of the promoter did not necessarily lead to lower expression levels, indicating that the mRNA secondary structure has a major influence (30). The same might be caused by altering the distance of transcription factor binding sites due to such inserted sequences.

Modern cloning techniques are often designed to allow seamless cloning, *i.e.* fusion of the gene of interest to the promoters or tags without undesired DNA fragments in between. Many alternative methods were developed using PCR based cloning, *e.g.* overlap extension PCR

cloning (32,33) or Circular Polymerase Extension Cloning (CPEC). Methods like CPEC can also be used to join multiple fragments, which is necessary for the assembly of pathways (34).

Another option to allow seamless cloning is the use of homologous recombination *in vivo*. In *E. coli* there are two different ways to perform *in vivo* recombination cloning. Either by co-transforming the PCR product, which contains appropriate homologous overlaps to the vector, and the cut vector backbone (35,36) or by simply transforming the linear PCR product in cells carrying the cloning vector (37). These methods have been successfully applied for the cloning of large libraries of predicted ORFs (38). Similarly, in vitro recombination by methods such as Gibson cloning (39), the NEBuilder System or In-Fusion cloning can be applied efficiently.

An alternative approach, which is still dependent on restriction enzymes, but does not leave scars, is the use of type IIS restriction enzymes. These enzymes recognize defined non-palindromic restriction sites and cut several, defined bp outside of their recognition site (40). There are various type IIS restriction enzymes available leaving either blunt ends or sticky ends with overhangs ranging from 1 to 4 bp. A set of expression vectors designed for the use of the type IIS restriction enzymes *Mly*I and *Bmr*I was reported for blunt end PCR cloning and TA cloning. Although these methods do not allow to control the direction of the gene of interest and show lower transformation rates, they still have the advantage that restriction sites in the insert do not matter and just short primers are needed, since there is no need for overlapping regions (31).

In this thesis the type IIS enzyme *Sap*I was used. This enzyme leaves a 3 bp sticky end, which allows more efficient cloning and proper orientation of the insert. *Sap*I cloning can be performed in a single pot reaction at room temperature using *Sap*I to cut the insert and T4 DNA ligase to join the fragments.
The commercial Electra cloning system (DNA2.0) is also based on *Sap*I cloning, but it is designed for other overhangs, which add an additional triplet upstream of the stop codon.

1.4 Selection systems for *P. pastoris*

Selection markers are needed to select for strains carrying foreign DNA after transformation. In yeast often auxotrophic selection is used, but also antibiotic selection. While auxotrophic selection needs minimal media and the proper strain, antibiotic selection often needs complex media (41).
Different selection systems are available for *P. pastoris*. There is a wide range of antibiotics that can be used for selection, most prominently Zeocin selection using the *Streptoalloteichus hindustanus ble* gene (42). Other genes available provide resistance to blasticidin (43), nourseothricin (44), G418 (45), soraphen A (46) or hygromycin (47).

Additionally, to the antibiotic selection there are also auxotrophic selection systems available. The most widely used auxotrophic selection marker is *HIS4* that can be used with the commercial strain *P. pastoris* GS115. Other common selection markers based on auxotrophy are for example *ARG4*, *URA3*, *URA5* (48) or *GUT1* (18), which was also used for the episomal expression vectors described in this study.
Two other auxotrophic selection markers that can be used if multicopy strains are preferred are *ADE1* and *ADE2*. It was reported that most of the transformants contain at least two copies of the expression cassette (49).
Auxotrophic markers that could be interesting for episomal expression vectors are ones that provide selection pressure on media that induce *P. pastoris* promoters, like methanol or oleate containing media. The *formaldehyde dehydrogenase* gene (*FLD1*) can be used since auxotrophic mutant strains cannot grow on methanol as the sole carbon source (50).
POX1 encodes for the first and rate limiting enzyme in ß-oxidation of fatty acids like oleate (51) and could be a powerful selection marker as well as *PEX19*, which codes for a protein that is crucial for the biogenesis of peroxisomes. In the absence of *PEX19*, peroxisomal membrane structures cannot be build up and therefore also β-oxidation cannot take place (52).

Both systems, the dominant selection with antibiotics and auxotrophic selection, have some disadvantages. In order to keep up selection pressure for episomal expression for example antibiotics have to be added to the culture medium, which is often undesirable. And for auxotrophic selection respective mutant strains have to be made.

Another alternative to overcome these problems are dominant selection markers, which allow the wildtype to grow on media components that cannot be utilized normally. *P. pastoris* is not capable of growing on sucrose, for example. Introduction of an invertase from *S. cerevisiae* (*SUC2*) allows the yeast to utilize sucrose as a carbon source (53).

A similar approach is the use of an amidase from *Aspergillus* (*amdS*). After introduction of *amdS* acetamide can be used as the sole nitrogen source. The same marker also allows counter selection in case fluoroacetamide is added to the media (54).

2 Aim of the thesis

The objective of the thesis was to provide an efficient episomal expression kit for *P. pastoris* based on a new ARS showing beneficial properties compared to the commonly used *PARS1* and in defined applications also better performance than the more widely used integrative vector systems.

The kit should comprise two basic sets of plasmids, one for antibiotic selection and another one for auxotrophic selection, both of them with different promoters and the possibility to use them for intracellular and secretory expression. For auxotrophic selection corresponding knockout strains should be provided in the *P. pastoris* BSYBG10 and BSYBG11 strains of bisy.

Further an easy to use IP-free cloning system should be established that allows state of the art seamless cloning.

Potential ARSs were selected based on previous results. In the Dissertation "Synthetic biology to improve protein expression in *Pichia pastoris*" from Thomas Vogl two potential candidates were found.

The promoter of the *catalase 1* gene (P_{CAT1}) exhibited ARS behavior when the length from the start codon to the next gene upstream (692 bp) was used. The ARS function was lost after deleting the 192 bp long AT-rich region at the 5' end, using the shorter, 500 bp long variant. The expression strength although remained the same. Therefore, the 192 bp long AT-rich region was selected for a comprehensive analysis of different promoter fragments.

A heterologous promoter from the *Candida boidinii methanol oxidase 1* gene (P_{CbAOD1}) also served as origin of replication. Plasmids containing P_{CbAOD1} behaved similar to the $P_{CAT1\text{-}692}$ containing plasmids.

Plasmids with P_{CbAOD1} or $P_{CAT1\text{-}692}$ showed higher transformation efficiencies. Transformation with circular DNA was possible and transformation with linear DNA yielded background colonies, which contained the vector episomally. Seemingly the linearized DNA circularized in the cells by non-homologous end joining (NHEJ) repair after transformation.

Besides the generation of an innovative plasmid vector set a major goal was a detailed evaluation of these 2 DNA parts with promoter and ARS function.

3 Materials and Methods

3.1 Materials

All strains, plasmids, media, chemicals, enzymes, devices and other equipment used during the thesis are listed in Tables 3.1 to 3.6.

The strains generated during the thesis are listed in Section 8.1.

Table 3.1 Strains and plasmids used during the thesis

Strain name	**Source**
Pichia pastoris BSYBG10	IMBT #7287
Pichia pastoris BSYBG11	IMBT #7288
Escherichia coli K12 TOP10F'	IMBT #1482
Vector name (*E. coli* strains)	**Source**
pADH_ADHTT_B1_2xmat2x201_CBH2-V04	IMBT #3892
pPpT4_S_AlphaS_Zeo_P_CAT_500	Julia Pitzer
pPpT4_alpha-CalB_S_KatrinWeinhandl	Katrin Weinhandl
pPpT4mutZeoMlyI-intArg4-EGFP-AOX1BglII	TV0020*
pPpT4mutZeoMlyI-intArg4-EGFP-CAT1 (=pCAT1-692)	TV0038*
pPpT4mutZeoMlyI-intArg4-EGFP-GAP	TV0056*
pPpT4mutZeoMlyI-intArg4-EGFP-CbAOD1	TV0069*
pPpT4mutZeoMlyI-intArg4-EGFP-AOX1BglII-TM20-ScSPG5TT K2	TV0756*
pPpT4mutZeoMlyI-intArg4-EGFP-AOX1BglII-TM22-ScTHI5 stuffer/no terminator + NotI K13	TV0758*
pPpT4mutZeoMlyI-intArg4-EGFP- pCAT1-500 Clone A	TV0973*
pPpT4-bidi-dTomato-EGFP-V38 (<--pCoreHHT2-TATA(55)<-pDAS1-TATA-105->)	TV1195*
pPpT4mutZeoMlyI-intArg4-EGFP-AOX1BglII-CAT1ARS-TT	TV1254*

*TV, internal strain collection Thomas Vogl (for details see Dissertation "Synthetic biology to improve protein expression in *Pichia pastoris*).

P. pastoris BSYBG10 (wildtype) was used for all expression experiments, except for the secretory expression of *Candida antarctica lipase B (CalB)* and *TrCBH2* where *P. pastoris* BSYBG11 (Mut^S) was used, because Mut^S strains had exhibited better performance for secretory protein expression in previous studies if methanol was applied for induction (55).

All the cloning work was performed in *E. coli* K12 TOP10 F'.

Table 3.2 Used media, their composition and stock solutions

Media (recipes for 1 l)	**Components**
BMD1	10 g glucose, 200 ml 10x PPB, 100 ml 10x YNB, 2 ml biotin
BMG1	10 g glycerol, 200 ml 10x PPB, 100 ml 10x YNB, 2 ml biotin
BMM2	10 ml methanol, 200 ml 10x PPB, 100 ml 10x YNB, 2 ml biotin
BMM10	50 ml methanol, 200 ml 10x PPB, 100 ml 10x YNB, 2 ml biotin
YPD	10 g yeast extract, 20 g peptone, 100 ml 10x D
YPM2	10 ml methanol, 10 g yeast extract, 20 g peptone
YPM10	10 ml methanol, 10 g yeast extract, 20 g peptone
LB	10 g tryptone, 5 g yeast extract, 5 g NaCl
BEDS (filter sterilized)	10 mM bicine-NaOH, 30 ml ethylene glycol, 50 ml DMSO, 1 M sorbitol, pH 8.3
SOC	3.46 g glucose, 20 g tryprone, 5 g yeast extract, 0.56 g NaCl, 2 g $MgCl_2$, 0.16 g KCl, 2.46 g $MgSO_4$
Stocks for media (recipes for 1 l)	**Components**
10X YNB	134 g yeast nitrogen base
10X D (glucose)	200 g glucose (220 g glucose monohydrate)
500X Biotin (filter sterilized)	200 mg biotin
10X PPB (potassium phosphate buffer)	30 g K_2HPO_4, 118 g KH_2PO_4, pH 6
Antibiotics	**Concentration in media**
Ampicillin 100 mg/ml	100 µg/ml
Kanamycin 100 mg/ml	100 µg/ml
Zeocin 100 mg/ml	25 µg/ml (*E. coli*), 50 µg/ml (*P. pastoris*)
Plates	**Like liquid media with 15 g/l agar-agar**

Zeocin concentration in yeast media was always 50 µg/ml if not stated otherwise. *E.g.* YPD-Zeo and YPM2-Zeo refer to 50 µg/ml.

Table 3.3 Enzymes and kits used during the thesis

Restriction enzymes	**Company**
EndoH	New England Biolabs, Ipswich, MA, Unites States
FastDigest *Bam*HI	Thermo Scientific – Austria GmbH, Vienna, Austria
FastDigest *Bgl*II	Thermo Scientific – Austria GmbH, Vienna, Austria
FastDigest *Eco*RI	Thermo Scientific – Austria GmbH, Vienna, Austria
FastDigest *Kpn*I	Thermo Scientific – Austria GmbH, Vienna, Austria
FastDigest *Nhe*I	Thermo Scientific – Austria GmbH, Vienna, Austria
FastDigest *Not*I	Thermo Scientific – Austria GmbH, Vienna, Austria
FastDigest *Pst*I	Thermo Scientific – Austria GmbH, Vienna, Austria
FastDigest *Sma*I (*Swa*I)	Thermo Scientific – Austria GmbH, Vienna, Austria
FastDigest *Xba*I	Thermo Scientific – Austria GmbH, Vienna, Austria
FastDigest *Xho*I	Thermo Scientific – Austria GmbH, Vienna, Austria
FastDigest *Sac*I	Thermo Scientific – Austria GmbH, Vienna, Austria

*Kpn*I-HF®	New England Biolabs, Ipswich, MA, Unites States
*Sap*I	New England Biolabs, Ipswich, MA, Unites States
Other enzymes	**Company**
FastAP Thermosensitive Alkaline Phosphatase	Thermo Scientific – Austria GmbH, Vienna, Austria
GoTaq® DNA Polymerase	Promega Madison, WI, USA
Phusion® High Fidelity DNA Polymerase	Thermo Scientific – Austria GmbH, Vienna, Austria
T4 DNA Ligase	New England Biolabs, Ipswich, MA, Unites States
T4 DNA Ligase	Thermo Scientific – Austria GmbH, Vienna, Austria
T5 Exonuclease	New England Biolabs, Ipswich, MA, Unites States
Taq DNA Ligase	New England Biolabs, Ipswich, MA, Unites States
Zymolyase®20T	Gerbu Biotechnik GmbH, Heidelberg, Germany
Kits	**Company**
GeneJET Plasmid Miniprep Kit	Thermo Scientific – Austria GmbH, Vienna, Austria
GeneJET Gel Extraction Kit	Thermo Scientific – Austria GmbH, Vienna, Austria
Wizard® SV Gel and PCR Clean-Up System	Promega Madison, WI, USA
CloneJET PCR Cloning Kit	Thermo Scientific – Austria GmbH, Vienna, Austria

Table 3.4 Chemicals used during the thesis

Reagents/Chemicals	**Company**
4-Nitrophenyl butyrate (pNPB)	Sigma-Aldrich Chemie GmbH, Vienna, Austria
4-Hydroxybenzoic acid hydrazide	Sigma-Aldrich Chemie GmbH, Vienna, Austria
6x Mass Ruler DNA Loading Dye	Thermo Scientific – Austria GmbH, Vienna, Austria
Adenosine triphosphate	New England Biolabs, Ipswich, MA, Unites States
Agar-Agar	Carl Roth GmbH, Karlsruhe, Germany
Ampicillin	Sigma-Aldrich Chemie GmbH, Vienna, Austria
Aqua bidest. "Fresenius"	Fresenius Kabi Austria GmbH, Graz, Austria
Avicel® PH-101	Sigma-Aldrich Chemie GmbH, Vienna, Austria
Bacto™ peptone	Becton, Dickinson and Company, Sparks, USA
Bacto™ yeast extract	Becton, Dickinson and Company, Sparks, USA
Bicine	Fluka Chemia AG, Basel, Switzerland
Biozym LE Agarose	Biozym Biotech Trading GmbH, Vienna, Austria
Citric acid monohydrate	Carl Roth GmbH, Karlsruhe, Germany
D-biotin	Fluka Chemia AG, Basel, Switzerland
D-glucose monohydrate	Carl Roth GmbH, Karlsruhe, Germany
D-sorbitol	Carl Roth GmbH, Karlsruhe, Germany
Deoxyadenosine triphosphate	Thermo Scientific – Austria GmbH, Vienna, Austria
Deoxycytidine triphosphate	Thermo Scientific – Austria GmbH, Vienna, Austria
Deoxyguanosine triphosphate	Thermo Scientific – Austria GmbH, Vienna, Austria
Deoxythymidine triphosphate	Thermo Scientific – Austria GmbH, Vienna, Austria
Di-potassium hydrogen phosphate	Carl Roth GmbH, Karlsruhe, Germany
Difco™ yeast nitrogen base w/o amino acids	Becton, Dickinson and Company, Sparks, USA

Dimethyl sulphoxide	Carl Roth GmbH, Karlsruhe, Germany
Dithiothreitol	Carl Roth GmbH, Karlsruhe, Germany
Ethanol absolute	Merck KGaA, Darmstadt, Germany
Ethidium bromide (1%)	Carl Roth GmbH, Karlsruhe, Germany
Ethlenediaminetetraacetic acid	Carl Roth GmbH, Karlsruhe, Germany
Ethylene glycol	Sigma-Aldrich Chemie GmbH, Vienna, Austria
Glycerol	Carl Roth GmbH, Karlsruhe, Germany
Hydrochloric acid (37%)	Merck KGaA, Darmstadt, Germany
Kanamycin	Carl Roth GmbH, Karlsruhe, Germany
LB-media	Carl Roth GmbH, Karlsruhe, Germany
LDS Sample buffer 4x	Thermo Scientific – Austria GmbH, Vienna, Austria
Liquid nitrogen	Air Liquide Austria GmbH, Graz, Austria
Magnesium chloride	Promega Madison, WI, USA
Methanol	Carl Roth GmbH, Karlsruhe, Germany
Potassium dihydrogen phosphate	Carl Roth GmbH, Karlsruhe, Germany
Potassium hydroxide	Carl Roth GmbH, Karlsruhe, Germany
Roti® phenol/chloroform/isoamylalcohol	Carl Roth GmbH, Karlsruhe, Germany
Sodium acetate	Carl Roth GmbH, Karlsruhe, Germany
Sodium chloride	Carl Roth GmbH, Karlsruhe, Germany
Sodium dodecyl sulphate	Carl Roth GmbH, Karlsruhe, Germany
Sodium hydroxide	Carl Roth GmbH, Karlsruhe, Germany
Tris	Carl Roth GmbH, Karlsruhe, Germany
Triton X-100	Carl Roth GmbH, Karlsruhe, Germany
Zeocin™	InvivoGen-Eubio, Vienna, Austria

Table 3.5 Instruments, devices and single use materials used during the thesis

Instruments	**Company**
2720 Thermal Cycler	Applied Biosystems, Foster City, USA
Arium® basic ultrapure water system	Sartorius Stedim Biotech GmbH, Göttingen, Germany
Benchtop 2 UV Transilluminator	UVP®, Cambridge, UK
Binder drying oven	Binder GmbH, Tuttlingen, Germany
BioRad Gene Pulser 1652076	Bio-Rad Laboratories Inc., Hercules, CA, United States
BioRad PowerPac Basic Power Supply	Bio-Rad Laboratories Inc., Hercules, CA, United States
Capicitance Extender 1652087	Bio-Rad Laboratories Inc., Hercules, CA, United States
Certoclav LVEL 12L	CertoClav GmbH, Traun, Austria
Eppendorf BioPhotometer plus	Eppendorf AG, Hamburg, Germany
Eppendorf Centrifuge 5415 D	Eppendorf AG, Hamburg, Germany
Eppendorf Centrifuge 5415 R	Eppendorf AG, Hamburg, Germany
Eppendorf Centrifuge 5810 R	Eppendorf AG, Hamburg, Germany
Eppendorf Thermomixer comfort	Eppendorf AG, Hamburg, Germany
GelDoc-It™ Imaging System	UVP®, Cambridge, UK
Hamilton® Polyplast lab pH electrode	Sigma-Aldrich Chemie GmbH, Vienna, Austria

Heidolph MR 3000 Magnetic Stirrer	Heidolph Instruments, Schwabach, Germany
HT Infors Orbitron shaker	Infors AG, Bottmingen, Switzerland
inoLab® pH 720 pH Meter	WTW GmbH, Weilheim, Germany
MT PG12001-S DeltaRange Balance	Mettler Toledo Inc., Greifensee, Switzerland
NanoDrop 2000c Spectrophotometer	Peqlab Biotechnologie GmbH, Polling, Austria
PowerEase® 500 Power Supply	Invitrogen™, Lofer, Austria
Pulse Controller P/N 1652098	Bio-Rad Laboratories Inc., Hercules, CA, United States
Sartorius Analytical BL120S	Sartorius Stedim Biotech GmbH, Göttingen, Germany
SynergyMxPlate reader	BioTek Instruments, Winooski, USA
Vortex Genie 2	Scientific Industries, New York, USA
Pipettes	**Company**
Acura® manual 855 Multichannel Micropipette, 5-50µl	Socorex Isba S.A., Ecublens, Swiss
Biohit Proline® Multichannel Electronic Pipettor, 8 channels, 5- 100 µL	Biohit Oyj, Helsinki, Finnland
Biohit Proline® Multichannel Electronic Pipettor, 8 channels, 50-1200 µL	Biohit Oyj, Helsinki, Finnland
Denvillle XL3000i™ Single Channel Pipette 100-1000 µl	Denville Scientific Inc., Holliston, USA
Denvillle XL3000i™ Single Channel Pipette 20-200 µl	Denville Scientific Inc., Holliston, USA
Denvillle XL3000i™ Single Channel Pipette 2-20 µl	Denville Scientific Inc., Holliston, USA
Eppendorf Research® Series 2100 Adjustable Volume, Single Channel Pipette, 0.5 -10 µL	Eppendorf AG, Hamburg, Germany
Eppendorf Research® Series 2100 Adjustable Volume, Single Channel Pipette, 0.1-2.5 µL	Eppendorf AG, Hamburg, Germany
Plates	**Company**
96-Well Deep Well Plate	Bel-Art Products, Wayne, NJ, United States
96-Well PS Microplate	Greiner Bio-One GmbH, Frickenhausen, Germany
MicroAmp® Adhesive covers	Applied Biosystems, Foster City, USA
MicroAmp® Optical 96-Well Reaction Plate	Applied Biosystems, Foster City, USA
Nunc™ MicroWell™ 96-Well Optical-Bottom Plates	Thermo Scientific – Austria GmbH, Vienna, Austria
Electrophoresis markers and gels	**Company**
GeneRuler™ 1kb DNA-Ladder	Thermo Scientific – Austria GmbH, Vienna, Austria
PageRuler™ Prestained Protein Ladder	Thermo Scientific – Austria GmbH, Vienna, Austria
NuPAGE™ 4-12% Bis-Tris Protein Gel, 15 slots	Thermo Scientific – Austria GmbH, Vienna, Austria
Single use materials	**Company**
15 ml PP-Tube, sterile	Greiner Bio-One GmbH, Frickenhausen, Germany
50 ml PP-Tube, sterile	Greiner Bio-One GmbH, Frickenhausen, Germany
Biohit Tips 1200 µl	Biohit Oyj, Helsinki, Finnland
Biohit Tips 300 µl	Biohit Oyj, Helsinki, Finnland
Electroporation cuvettes EP-102	Cell Projects Ltd, Kent, United Kingdom
Microcentrifuge tubes, 1.5 mL with lid	Greiner Bio-One GmbH, Frickenhausen, Germany
Petri Dishes 94x16 mm	Greiner Bio-One GmbH, Frickenhausen, Germany
Pipette tips, 1000 µl	Greiner Bio-One GmbH, Frickenhausen, Germany
Pipette tips, 200 µl	Greiner Bio-One GmbH, Frickenhausen, Germany
Pipette tips, 10 µl	Greiner Bio-One GmbH, Frickenhausen, Germany

Semi-micro cuvette 10x4x45 mm, PS	Sarstedt Aktiengesellschaft & Co., Nümbrecht, Germany
Square Petri Dishes 120x120x17 mm	Greiner Bio-One GmbH, Frickenhausen, Germany
Ultra Yield Flasks^{TM} 250 ml	Thomson Instrument Company, California, USA

Table 3.6 Software used during the thesis

Software	Company
Gene Designer	DNA2.0 Inc., Menlo Park, CA, USA
CLC Main Workbench 6	QIAGEN N.V., Spoorstraat, Netherlands
RNA secondary structure prediction	http://www.genebee.msu.su/services/rna2_reduced.html
EMBOSS 6.3.1: freak	http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::freak
BitGene Gene Analysis	http://www.bitgene.com/cgi/gene_analysis.cgi
Multiple Primer Analyzer	https://www.thermofisher.com/at/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html
T_M calculator	https://www.thermofisher.com/at/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html
Basic Local Alignment Search tool	http://blast.ncbi.nlm.nih.gov/Blast.cgi
Clustal Omega	http://www.ebi.ac.uk/Tools/msa/clustalo/

3.2 Genomic DNA isolation from yeast

Yeast genomic DNA (gDNA) for the generation of the knockout strains was isolated according to the Bust n' Grab protocol reported by Harju *et al.* (56) with slight changes.
The cell material was scratched off agar plates. Instead of the dry ice and ethanol bath liquid nitrogen was used and instead of the water bath a thermomixer.
All centrifugation steps were performed for 5 min at maximal speed and the DNA pellet was dissolved in 20 µl ddH_2O.

3.3 Plasmid DNA isolation

The isolation of plasmid DNA from *E. coli* and also from *P. pastoris* (see section 3.4) was performed with the GeneJET Plasmid Miniprep Kit (Thermo Scientific) according to the manufacturer's manual.

3.4 *Pichia* plasmid isolation

Plasmids were isolated from *P. pastoris* using an adapted protocol of the one, which was previously described in literature for *S. cerevisiae* (57). A 5 ml ONC was harvested by centrifugation (500 rcf, 5 min) and the pellet was washed with 1 ml ddH_2O.
The washed cell pellet was resuspended in 1 ml yeast lysis buffer (1 M Sorbitol, 100 mM EDTA, 14 mM β-Mercaptoethanol) and 100 µl of a Zymolyase stock solution (1000 U/ml) were added. The reaction mix was incubated at 30°C for 1 h without shaking and the generated spheroblasts were harvested by centrifugation for 5 min at maximal speed.
Afterwards the GeneJET Plasmid Miniprep Kit (Thermo Scientific) was applied according to the manufacturer's protocol. The isolated DNA can be used as a PCR template and for the transformation of *E. coli* or *P. pastoris*.

3.5 PCR and gel purification

For the purification of PCR products and of DNA from agarose gel slices the Wizard® SV Gel and PCR Clean-Up System (Promega) or the GeneJET Gel Extraction Kit (Thermo Scientific) were used.

3.6 Polymerase chain reactions (PCRs)

3.6.1 Standard PCR

Standard PCRs for the amplification of DNA fragments, *e.g.* ARSs and promoters were carried out using the Phusion® High-Fidelity DNA Polymerase (Thermo Scientific) in a 50 µl reaction mix, which is shown in Table 3.7 and the temperature profile shown in Table 3.8.

Table 3.7 Standard PCR composition

Component	Amount (µl)
ddH_2O	28.5
Phusion® HF Buffer (5x)	10
dNTPs (2 mM)	5
Forward primer (10 µM)	2.5
Reverse primer (10 µM)	2.5
Template DNA	1
Phusion® Polymerase	0.5

Table 3.8 Temperature profile for standards PCRs

Step	T (°C)	t (s)	Cycle number
Initial denaturation	98	180	1
Denaturation	98	10	30
Annealing	*	30	
Extension	72	30 s/kb	
Final extension	72	420	1
Hold	4	∞	

* Annealing temperature was 3°C above the lower T_M of the two primers for primers >20 nt and equal to that T_M for primers up to 20 nt.

3.6.2 Colony PCR

GoTaq® Flexi DNA polymerase was used for colony PCRs in order to check the presence of the correct vector in *E. coli* colonies.

One colony was resuspended in 20 µl ddH_2O and boiled for 10 min at 98°C. Afterwards the cell debris was spun down for 5 min at maximal speed and 2 µl of the supernatant were used for the PCR in a total volume of 20 µl per reaction with the composition shown in Table 3.9 and the temperature profile shown in Table 3.10.

Table 3.9 Colony PCR composition

Component	Amount (µl)
ddH_2O	9.9
GoTaq® Green Buffer (5x)	4
$MgCl_2$ (25 mM)	1.2
dNTPs (2 mM)	2
Forward primer (10 µM)	0.4
Reverse primer (10 µM)	0.4
Template DNA	2
GoTaq® Polymerase	0.1

Table 3.10 Colony PCR temperature profile

Step	T (°C)	t (s)	Cycle number
Initial denaturation	95	300	1
Denaturation	95	30	30
Annealing	*	30	
Extension	72	60 s/kb	
Final extension	72	300	1
Hold	4	∞	

* Annealing temperature was 5°C below the lower T_M of the two primers for *Taq* polymerase

3.6.3 Overlap extension PCR (oePCR)

Overlap extension PCR was used to join DNA fragments for building the knock out cassette, the final expression vectors containing the *CbARS* ($P_{CbAOD1\text{-}F1}$) and joining the different promoters with the signal sequence.

The reaction was carried out in two steps with a volume of 25 µl each. The composition is shown in Table 3.11 and the time profile in Table 3.12. For the fist step without primers 25 cycles were run and the amplification of full length fragments was carried out using 15 cycles in a reaction mix containing outer primers.

Equimolar amounts of the DNA (5 ng of the largest fragment) were used as template for the first PCR round.

Table 3.11 Overlap extension PCR composition

	Component	**Amount (µl)**
PCR1	ddH_2O	17.2 - x
	Phusion® HF Buffer (5x)	5
	dNTPs (2 mM)	2.5
	Template DNA	x
	Phusion® Polymerase	0.3
PCR2	**Component**	**Amount (µl)**
	ddH_2O	12.2
	Phusion® HF Buffer (5x)	5
	dNTPs (2 mM)	2.5
	Forward primer (10 µM)	2.5
	Reverse primer (10 µM)	2.5
	Phusion® Polymerase	0.3

Table 3.12 Overlap extension PCR cycling profile

	Step	**T (°C)**	**t (s)**	**Cycle number**
PCR1	Initial denaturation	98	30	1
	Denaturation	98	10	25
	Annealing	*	30	
	Extension	72	30 s/kb	
	Final extension	72	420	1
	Hold	4	hold	
PCR2	**Step**	**T (°C)**	**t (s)**	**Cycle number**
	Initial denaturation	98	30	1
	Denaturation	98	10	15
	Annealing	*	30	
	Extension	72	30 s/kb	
	Final extension	72	420	1
	Hold	4	∞	

* Annealing temperature was 3°C above the lower T_M of the two primers for primers >20 nt and equal to that T_M for primers up to 20 nt. In the first step it 3°C above the T_M of the overlapping region.

3.7 Gibson cloning

Gibson cloning (39) was used to join overlapping DNA fragments for the construction of test vectors. Fifteen µl of an assembly mastermix (Table 3.13), 50-100 ng of cut vector backbone and insert with a molar ratio of 1:2 and ddH_2O to a total volume of 20 µl were mixed and incubated for 1 h at 50 °C. Two µl of the reaction were used for the transformation of *E. coli*.

Table 3.13 Gibson assembly mastermix composition

Component	Amount (µl)
ISO reaction buffer (5x)	220
T5 exonuclease (10 U/µl)	0.64
Phusion® Polymerase	20
Taq DNA ligase (40 U/µl)	160
ddH_2O	700

3.8 Restriction digests

Preparative restriction digests for the linearization of expression vectors and for classical cloning were performed with 0.5 µl restriction enzyme per µg DNA in a total volume of 50 µl and appropriate conditions according to the manufacturer. The cut DNA fragments were gel or column purified afterwards.
Control cuts were made with 2 µl miniprep and 0.5 µl enzyme in a total volume of 20 µl.

3.9 Ligation

Ligation reactions were set up using 50 ng of the largest DNA fragment and a 3:1 molar ratio (insert:vector) when small inserts, *e.g.* promoters were used, and for building the final expression vectors equimolar amounts were used. The reaction mix shown in Table 3.14 was incubated for 1 h at room temperature and heat inactivated at 70°C for 5 min prior the transformation.

Table 3.14 Ligation reaction composition

Component	Amount
DNA fragment 1	50 ng
DNA fragment 2	1:1 or 3:1 molar ratio
T4 DNA ligase buffer (10x)	2 µl
T4 DNA ligase (5 Weiss U/µl)	1 µl
ddH_2O	to 20 µl

3.10 pJET cloning

The knockout cassettes for the generation of *GUT1* and *KU70* knockout strains were stored by cloning the cassettes in two parts into pJET1.2 with the CloneJET PCR Cloning Kit (Thermo scientific) using a 3:1 molar excess of the insert. The ligation mix was incubated for 1 h at room temperature.

3.11 *Sap*I cloning

The vectors containing a *Sap*I cloning stuffer were linearized using *Sap*I and dephosphorylated simultaneously with FastAP, gel purified and diluted to 5 ng/µl.

The cloning was performed using the reaction set up shown in Table 3.15 and incubated for 1 h at room temperature. Afterwards the cloning mix was desalted with a nitrocellulose membrane and 10 µl of the desalted reaction mix were transformed into electrocompetent *E. coli* TOP10F' cells or the whole reaction mix into electrocompetent *P. pastoris* cells.

Amplification of the gene of interest was performed using TACACGTACTTAGTCGCTGAAGCTCTTCT**ATG** as overhang for the forward primer and AGGTACGAACTCGATTGACGGCTCTTCT**TTA** for the reverse primer. These overhangs can be used for cloning into all intracellular expression vectors. For cloning into secretory expression vectors a different forward primer has to be used since there is not an ATG in every secreted protein. The primer contains an overlap to the last three bp of the signal sequence. The overhang TACACGTACTTAGTCGCTGAAGCTCTTCTGCT should be used for cloning into secretory expression vectors.

Table 3.15 *Sap*I cloning reaction mix

Component	Amount (µl)
CutSmart buffer (10x)	2
ATP (10 mM)	2
*Sap*I (10 U/µl)	0.25
T4 DNA ligase (400 U/µl)	0.25
Cut vector (5 ng/µl)	2
PCR product (60 ng/µl)	2
ddH_2O	to 20 µl

3.12 Sequencing

All constructs made during the thesis were sequence verified using Sanger sequencing. The samples were sent to LGC Genomics GmbH, Germany or Microsynth AG, Austria.

3.13 Agarose gel electrophoresis

Standard DNA electrophoresis was performed in 1% agarose gels containing ethidium bromide. For smaller fragments, e.g. PCR products of the ARSs, higher agarose concentrations were used.
Preparative agarose gels were run with 90 V and control gels with 120 V in 1x TAE buffer.

3.14 Methanol/Chloroform precipitation

Total protein precipitation was performed using methanol/chloroform precipitation. Two-hundred µl protein containing culture supernatant were collected and mixed with 320 µl methanol and 107 µl chloroform. After vortexing 360 µl ddH_2O were added, the samples vortexed again and centrifuged for 5 min at maximal speed.

The supernatant was carefully removed without touching the protein layer and 200 µl methanol were added followed by vortexing and centrifugation for 30 min at 4°C and maximal speed. The supernatant was carefully removed and the pellet dried at 60°C for 10 min.

3.15 Protein deglycosylation

The protein pellet was dissolved in 10 µl 1x Glycoprotein Denaturing buffer and boiled at 99°C for 10 min. Afterwards 2 µl 10x GlycoBuffer 3, 7.5 µl ddH_2O and 0.5 µl EndoH were added and the denaturated proteins deglycosylated for 1 hour at 37°C.

3.16 SDS-PAGE

For the SDS-PAGE of *Tr*CBH2p containing supernatants ready to use NuPAGE™ 4-12% Bis-Tris Protein Gels with 15 slots were applied.
NuPAGE® LDS Sample buffer was added to the deglycosylated proteins and 10 µl sample per well and 4 µl of the PageRuler™ Prestained Protein Ladder were loaded onto the gel, which was run for 1 h with 200 V in 1x MOPS buffer.

3.17 Protein staining

After the separation via electrophoresis the proteins were stained with Coomassie overnight. Three destaining steps with 20 min were performed after washing the gel with water.

3.18 Knockout strain generation

The *ΔGUT1* strains and the *KU70* knockout cassette were built for using the FLP recombinase system reported in (18). The knockout cassettes were designed to delete the whole ORFs by

adding 1000 bp and 1500 bp long flanking homologous regions for the *GUT1* and *KU70* locus respectively (for homologous sequences see section 7.5).

The homologous regions were amplified from genomic DNA using the primers from Table 3.16 and the 5' and 3' parts were attached to the 5' and 3' parts of the knockout cassette by oePCR. In a second round of oePCR the two resulting parts were combined to obtain the full length knockout cassette.

The *ΔGUT1* strains were generated by transforming *P. pastoris* BSYBG10 and BSYBG11 with 100-300 ng of the knockout cassette. The transformants were cultivated in 96-well deep well plates (DWPs) containing BMD1 and pinned onto BMD1 and BMG1 plates after 48 h of cultivation.

Clones showing the correct phenotype were confirmed by PCR using primers binding outside of the homologous regions.

Positive clones were cultivated again in DWPs for 48 h in BMD1 and induced for 72 h with BMM2 and BMM10 for excision of the cassette, *i.e.* marker recycling.

The clones were pinned onto BMM2 plates and subsequently re- streaked for single colonies on BMM2 afterwards.

Fifty-four BSYBG10 clones and 27 BSYBG11 clones were picked and re-streaked on YPD and YPD-Zeo plates to check if they still show Zeocin resistance. Clones without Zeocin resistance were confirmed by PCR with the same primers as mentioned above.

Table 3.16 Primers used for the knockout strain generation

Name	Sequence
5prFlipper_fwd	GAAGTTCCTATACTTTCTAGAGAATAGGAAC
5prFlipper_rev	TTATATGCGTCTATTTATGTAGGATGAAAG
3prFlipper_fwd	CTTTCATCCTACATAAATAGACGCATATAAACAGGCCCCTTTTCCTTTG
3prFlipper_rev	GAAGTTCCTATTCTCTAGAAAGTATAGGAAC
5prGUT1_fwd	GCAGCCTGCATTCTCTTG
5prGUT1_rev	GTTCCTATTCTCTAGAAAGTATAGGAACTTCTATAGTAGATATATCTGTGGTATAGTGTGAAAAAG
3prGUT1_fwd	GTTCCTATACTTTCTAGAGAATAGGAACTTCGAAGAGCAGCTGTAATTATATTATCATG
3prGUT1_rev	TTTTGACAGACTCATTGGTCAAG
5prKU70_fwd	GATCCAGAAGATCAGCGGTTTC
5prKU70_rev	GTTCCTATTCTCTAGAAAGTATAGGAACTTCCTTGCAATGCTTTTTATTATTCTCTGTGTTAG
3prKU70_fwd	GTTCCTATACTTTCTAGAGAATAGGAACTTCTTCGTGTTCCTTACTTTTTCCTC
3prKU70_rev	GGATGAAGGAGAAACTTGGC
5prGUT1outside	TGGGAGATTTGTTGGCAC
3prGUT1outside	TCAACCTGAAAAACGTTAAAATG

3.19 Electrocompetent *Pichia pastoris* cells

The electrocompetent *P. pastoris* cells were made according to the condensed protocol from Lin Cereghino (58). The cells were immediately used for the transformation.

3.20 *Pichia pastoris* transformation

Eighty µl electrocompetent *P. pastoris* cells were transformed with 1 µg linearized DNA for integration into the genome or with 1-10 ng circular plasmid DNA for episomal expression from ARS vectors.
After the cells were pulsed with 2.0 kV/25 µF/200 Ω, 500 µl 1 M sorbitol and 500 µl YPD were added immediately and the transformants were regenerated at 28°C and 600 rpm for 1 h.
When transforming *ΔGUT1* strains with *GUT1* plasmids regeneration is not necessary and YPD must not be added. The transformants were resuspended in 10% (w/v) glycerol or 1 M sorbitol and plated onto BMG plates.

3.21 *E. coli* transformation

Forty µl electrocompetent *E. coli* cells, prepared according to the protocol of Seidman *et al*. (59), were mixed with DNA and pulsed with 2.5 kV/25 µF/200 Ω. After the pulse 1 ml SOC media was added and the cells regenerated for 1 h at 37°C and 650 rpm before plating onto selective media.

3.22 Transformation rate evaluation

The transformation rate for *P. pastoris* was determined by transforming 100 ng plasmid (1 µl) and plating 100 µl of appropriate dilution for ARS plasmids and 100 µl and the rest of the undiluted, regenerated transformants for the integrative control.

3.23 P_{GAP} mutation / Site directed mutagenesis

Site directed mutagenesis was used for the mutation of P_{GAP} in order to remove the *Sap*I site from the promoter. Outer primers were used for amplifying parts of the vector backbone and eGFP and the inner primers were designed to target the position 239 harboring one single bp mutation (Table 3.17). The position was changed to all three other nucleotides. The 5' and 3' parts were amplified in two separate PCRs using pPpT4mutZeoMlyI-intArg4-EGFP-GAP (TV0056) as template and cloned into the *Swa*I/*Not*I linearized starting ARS vector containing the *CbARS*, which is described in section 3.30.3, using recombinase cloning.

Table 3.17 Primers used for generating P_{GAP} mutants and cloning of them and the wildtype promoter

Name	Sequence
pGAP-Mut239A fwd	GCAGCAATGATCTTCCCAGC
pGAP-Mut239A rev	GCTGGGAAGATCATTGCTGC
pGAP-Mut239T fwd	GCAGCAATGTTCTTCCCAGC
pGAP-Mut239T rev	GCTGGGAAGAACATTGCTGC
pGAP-Mut239G fwd	GCAGCAATGGTCTTCCCAGC
pGAP-Mut239G rev	GCTGGGAAGACCATTGCTGC
pUCori-SwaI-pGAP	CCTTTTGCTCACATGTATTTAAATTTTTTGTAGAAATGTCTTGGTGTCCTC
AOXTT-NotI-rev	GCATTCTGACATCCTCTTGAGC

3.24 Primer design

The primers were designed using the Multiple Primer Analyzer (Thermo Scientific), which shows melting temperatures optimized for Phusion DNA polymerase and also possible self- and hetero dimers that were tried to be avoided.

3.25 Glycerol stocks

Glycerol stocks of *E. coli* and *P. pastoris* strains were made either by adding 60% (w/v) glycerol to a liquid culture to a final concentration of 20% (w/v) or by resuspending cell material from agar plates with 30% (w/v) glycerol. The glycerol stocks were frozen at -80°C.

3.26 Plasmid loss (shake flask cultivation)

To test the plasmid stability of the *CbARS* containing expression vectors, strains bearing the Zeocin resistance and the *GUT1* based plasmids, integrated into the genome and as episomal expression vectors, were cultivated in 5 ml YPD-Zeo and BMG respectively. Fifty ml main cultures in 250 ml Ultra Yield Flasks™ with and without selection pressure (YPD and BMD respectively) were inoculated to an OD_{600} of 0.2 and cultivated for 168 h.
Samples were taken every 24 h for the measurement of the OD_{600}, eGFP expression and for plating appropriate dilutions onto selective media to determine the number of cells, which can still grow on selective media, *i.e.* still contain the plasmid.

3.27 Reporter protein measurements

3.27.1 Fluorescence measurement

An enhanced version of green fluorescent protein (eGFP) and a red fluorescent protein variant (dTomato) were used as reporter proteins. The measurements were performed in Nunc™ MicroWell™ 96-Well Optical-Bottom Plates and read out in a Synergy MX plate reader. In order to stay in the linear range the samples were diluted with ddH_2O to a total volume of 200 µl and the relative fluorescence units (RFU) were normalized to the OD_{600}. Excitation and emission wavelengths were 488/507 nm for eGFP and 554/581 nm for dTomato respectively.

3.27.2 *CalB* activity assay

CalB activity was determined using 4-Nitrophenyl butyrate (pNPB) according to the method described in literature (55). A 400 mM stock solution of pNPB in DMSO was prepared and frozen at -20°C. The working solution with a final concentration of 4 mM pNPB was prepared freshly before each measurement by diluting the stock solution with 300 mM Tric/HCl buffer

pH 7.4. The absorbance at 405 nm was measured over a period of 3 minutes every 11 seconds and the maximal velocity in mAU/min was calculated.
The volumetric activity in U/ml was calculated using the following formula.

$$\frac{U}{ml} = \frac{\Delta E * V * f}{v * \varepsilon * d}$$

ΔE… velocity (mAU/min) ε… extinction coefficient (9.594 ml*µmol^{-1}*cm^{-1}) V… total volume
v… sample volume f… dilution factor d… pathlength in cm (0.54)

3.27.3 *CBH2* activity measurement

The *CBH2* activity was determined by using microcrystalline cellulose (Avicel®) as a substrate and detecting the resulting reducing sugars with 4-hydroxybenzoic acid hydrazide (pHBAH) as reported in literature (60). The conversion reaction was carried out in 96 well microtiter plates. Therefore 20 µl of the culture supernatant were transferred to a microtiter plate, 150 µl of 1% Avicel in 50 mM citrate buffer pH 4.8 were added and the reaction mixture was incubated at 50°C and 1000 rpm for 3 h. Afterwards the remaining substrate was centrifuged and 20 µl supernatant, 40 µl buffer and 120 µl of the pHBAH working solution (prepared freshly by diluting a 5% (w/v) stock solution in 0.5% HCl 1:4 with 0.5 M NaOH) were mixed and incubated in a thermo cycler at 95°C for 5 min. The resulting conversion reaction was diluted with ddH_2O to a final volume of 200 µl to stay in the linear range of the plate reader and the absorbance was measured at 410 nm.
A calibration curve was made using different concentrations of glucose diluted with 100 mM sodium acetate buffer pH 5 and the absorbance was measured as described for the samples.

3.28 Deep well plate cultivation

The small scale cultivations were carried out in 96-well deep well plates according to the protocol described in literature by Weis *et al*. (61).

Buffered minimal glucose (BMD1) media was used for the cultivation of strains with integrated expression cassettes. Induction was performed over 48 h with BMM2 and BMM10.
Expression strains carrying ARS plasmids were cultivated using selective media to minimize plasmid loss, *i.e.* YPD-Zeo and YPM2-Zeo (and YPM10) for strains harboring the ARS plasmids for Zeocin selection and BMG1 when the *GUT1* based plasmids were used. The Zeocin concentration was 50 µg/ml if not indicated.

3.29 Codon optimization

The *glycerol kinase* (*GUT1*) gene was codon optimized using a codon usage from proteins that can be produced in high quantities using the strong P_{AOX1} (62). Additionally to the codon usage the mRNA sequence was predicted using the average length of the 5' and 3' UTRs (63), the resulting secondary structures of the mRNA were predicted using a webtool (Table 3.6) and strong secondary structures were avoided as well as local GC maxima or minima (above 70% or below 30%). Commonly used restriction sites, homopolymers and other undesired motifs were also avoided.

3.30 Vector construction

3.30.1 ARS test plasmids

The plasmid pPpT4mutZeoMlyI-intArg4-eGFP-pCAT1-500 (TV0973) obtained from Thomas Vogl was used as a reporter plasmid for ARS function. The different fragments of the *Candida boidinii AOD1* promoter (P_{CbAOD1}), *PARS1* and *CAT1-ARS* were cloned into the *Pst*I linearized vector after PCR amplification with the primers from Table 3.18 using recombinase cloning.

Table 3.18 Primers for cloning the different ARSs into the ARS reporter vector

	Name	Sequence
PARS1	AODTT-PstI-ARS1-Gib	GAAGATTGGGGAAACTTGGATCTGATTACCTTAGCTGCAGTCGAGATAAGCTGGGGGAACATTCG
	pUC Ori-KpnI-ARS1-Gib	CTACGGGGTCTGACGCTCAGTGGTACCTCGACAATTAATATTTACTTATTTTGGTCAACCCCAAATAG
CAT1-ARS	AODTT-PstI-pCAT1ARS-Gib	GAAGATTGGGGAAACTTGGATCTGATTACCTTAGCTGCAGAGTGTGTAATCATATATATAATAAATGAGGAATAATAATTGAATAGAGATTTAAC
	pUC Ori-KpnI-pCAT1ARS-Gib	CTACGGGGTCTGACGCTCAGTGGTACCCGTAGAAAAAAATGTGGTGAAACAGTTTCATAAGAG
AOD-F1	AODTT-PstI-CbAOD1ARS-F1-Gib	GAAGATTGGGGAAACTTGGATCTGATTACCTTAGCTGCAGGGAGTATACGTAAATATATAATTATATATAATCATATATATGAATACAATGCAATG
	pUC Ori-KpnI-CbAOD1ARS-F1-Gib	CTACGGGGTCTGACGCTCAGTGGTACCAAAATAAATTAAATAAGTTAAATAAAATTAAGTGAATAAAGTTTCAGAATTGTTATTAAG
AOD-F2	AODTT-PstI-CbAOD1ARS-F2-Gib	GAAGATTGGGGAAACTTGGATCTGATTACCTTAGCTGCAGTACCCCAGTTTTTCAGTACAATGCAGC
	pUC Ori-KpnI-CbAOD1ARS-F2-Gib	CTACGGGGTCTGACGCTCAGTGGTACCTGCGGAGTGGGGCGTG
AOD-F3	AODTT-PstI-CbAOD1ARS-F3-Gib	GAAGATTGGGGAAACTTGGATCTGATTACCTTAGCTGCAGTAAACATCCCCAGCAGTTTCCCCAG
	pUC Ori-KpnI-CbAOD1ARS-F3-Gib	CTACGGGGTCTGACGCTCAGTGGTACCATTTTAATTAAGCGAATATAAATTAATATTATAATATGAATTTATTTATAGATAGTAAATATAG
AOD-F4	AODTT-PstI-CbAOD1ARS-F4-Gib	GAAGATTGGGGAAACTTGGATCTGATTACCTTAGCTGCAGGCTCTTTTCCATCATCATCATCATCATCATCATC
	pUC Ori-KpnI-CbAOD1ARS-F4-Gib	CTACGGGGTCTGACGCTCAGTGGTACCTGATGATACTTAATTTACGTATATACATATATGAAAATAGAATAAAAAATGC
AOD-F5	AODTT-PstI-CbAOD1ARS-F5-Gib	GAAGATTGGGGAAACTTGGATCTGATTACCTTAGCTGCAGCGCTGCTTACTGTACGTTTAAAATGTGG
	pUC Ori-KpnI-CbAOD1ARS-F5-Gib	CTACGGGGTCTGACGCTCAGTGGTACCTATTGAAAAATAATTTTGTTTTTTTTTTTTTTGTTTTTTAAAAGTTCGTTAAAATTC

AOD refers to P_{CbAOD1} and F to the fragments thereof.

3.30.2 Transcription terminator test plasmids

The transcription terminator reporter plasmid was obtained from Thomas Vogl as well as the constructs containing *ScSPG5* (TV0756), *CAT1-ARS* (TV1254) and *Not*I*AOX*TT* (TV0020) as terminators. The reporter plasmid is based on the pPpT4_S shuttle vector (18). An integration site for the *ARG4* locus and eGFP were cloned into the vector (64) and the *AOX1* terminator replaced by a stuffer fragment containing a part of the *THI5* gene from *Saccharomyces cerevisiae*, which was also used in other vectors before as stuffer fragment (31).
The other ARSs and terminators respectively were cloned into the *Not*I and *Bam*HI linearized vector after PCR amplification with the primers from Table 3.19 using recombinase cloning.

Table 3.19 Primers used for cloning of ARS and transcription terminator elements into the terminator reporter vector

	Name	Sequence
PARS1	eGFP-ARS1-Gib	CACATGGCATGGATGAATTGTACAAGTAATCGAGATAAGCTGGGGGAACATTC
	pILV5-BamHI-ARS1-Gib	CACTGCAACAAAAGAAACAAGACATTACTGAAGGATCCTCGACAATTAATATTTACTTATTTTGGTCAACCCCAAATAG
AOD-TT	eGFP-AOD_TT-Gib	CACATGGCATGGATGAATTGTACAAGTAAAATTGACACCTTACGATTATTTAGAGAGTATTTATTAG
	pILV5-BamHI-AOD_TT-Gib	CACTGCAACAAAAGAAACAAGACATTACTGAAGGATCCTGCAGCTAAGGTAATCAGATCCAAG
GUT1-TT	eGFP-GUT1_TT-Gib	CATGGCATGGATGAATTGTACAAGTAAGAAGAGCAGCTGTAATTATATTATCATGTTAGG
	pILV5-BamHI-GUT1_TT-Gib	CACTGCAACAAAAGAAACAAGACATTACTGAAGGATCCTGCCAGAGCTGTCACATACTTG
AOD-F1	eGFP-CbAOD1-F1-Gib	CACATGGCATGGATGAATTGTACAAGTAAGGAGTATACGTAAATATATAATTATATATAATCATATATATGAATACAATGCAATG
	pILV5-BamHI- CbAOD1-F1-Gib	CACTGCAACAAAAGAAACAAGACATTACTGAAGGATCCAAAATAAATTAAATAAGTTAAATAAAATTAAGTGAATAAAGTTTCAGAATTG
$P_{CAT1-692}$	eGFP-pCAT1-Gib	GCATGGATGAATTGTACAAGTAAAGTGTGTAATCATATATATAATAAATGAGG
	pILV5-BamHI-pCAT1-Gib	CAAAAGAAACAAGACATTACTGAAGGATCCTTTAATTGTAAGTCTTGACTAGAGC
$P_{CAT1noCore}$	pILV5-BamHI-pCAT1noCore-Gib	CAAAAGAAACAAGACATTACTGAAGGATCCTGATTTTGGCCTGATGAG

3.30.3 Bifunctional *Sap*I cloning vectors

The *PARS1* and *CbARS* were cloned as a transcription terminator of the Zeocin resistance gene to test whether the ARSs are suitable for the bifunctional use as replication origin and transcription terminator. Therefore, the *AOD* terminator and the 3' end of the Zeocin resistance gene, which contains one *Sap*I recognition site, present in the pPpT4_S plasmid were cut out with *Sap*I and *Kpn*I.

The ARSs were amplified with the primers listed in Table 3.20 providing overlaps to the vector backbone and changed codons for the Zeocin resistance gene and cloned into the vector with recombinase cloning.

Afterwards the multiple cloning site (MCS) was exchanged by a stuffer fragment enabling *Sap*I cloning. The vector was *Eco*RI/*Not*I linearized, a shorter part of the stuffer fragment from the transcription terminator reporter vectors was PCR amplified with the primers shown in Table 3.20 adding appropriate restriction site and overlaps to the backbone and cloned into the vector using recombinase cloning.

Table 3.20 Primers used for cloning of the ARSs instead of the *AOD* terminator and modifying the vector for *Sap*I cloning

	Name	Sequence
*Sap*I cloning stuffer	pCAT1-SapCloning-ScTHI5-Gib	CATAACACTTGCTCTAGTCAAGACTTACAATTAAAATGAGAAGAGCGAATTCGGCGCGC CGGTAAGATCCAAATCGATGAATTGACCAAG
	AOX1_TT-SapCloning-ScTHI5-Gib	GCAAATGGCATTCTGACATCCTCTTGAGCGGCCGCTTATGAAGAGCTGTTCAATTGAGG CTTGAAGTCGATG
eGFP *Sap*I cloning	SapI-eGFP fwd	TACACGTACTTAGTCGCTGAAGCTCTTCTATGGCTAGCAAAGGAGAAGAACTTTTCAC
	eGFP-SapI rev	AGGTACGAACTCGATTGACGGCTCTTCTTTACTTGTACAATTCATCCATGCCATGTG
PARS1	ZeoMutSapI-PARS1-GIB	GCTGGTAACTGCGTGCATTTCGTCGCAGAGGAACAGGACTAATCGAGATAAGCTGGGG GAACATTCG
	KpnI-T4 backbone-PARS1-GIB	CTACGGGGTCTGACGCTCAGTGGTACCTAAAAAAAATCCTTAGCTTTCGCTAAGGATTC GACAATTAATATTTACTTATTTTGGTCAACCCCAAATAG
CbARS	ZeoMutSapI-CbAOD1-F1-GIB	GCTGGTAACTGCGTGCATTTCGTCGCAGAGGAACAGGACTAAGGAGTATACGTAAATA TATAATTATATATAATCATATATATGAATACAATGCAATG
	KpnI-T4 backbone-CbAOD1-F1-GIB	CTACGGGGTCTGACGCTCAGTGGTACCTAAAAAAAATCCTTAGCTTTCGCTAAGGATAA AATAAATTAAATAAGTTAAATAAAATTAAGTGAATAAAGTTTCAGAATTG

3.30.4 Zeocin^R ARS vectors

The final ARS vectors for Zeocin selection were built from synthetic DNA fragments. The *CbARS* was synthesized as gene from DNA2.0 and amplified using PCR, while the other parts of the vector were ordered as double stranded DNA fragments (gBlocks). The parts were joined with overlap extension PCR followed by restriction digest and ligation.

After PCR of the *CbARS* (oePrimer1-ZEO and oePrimer2-ZEO) the product was joined with gBlock2-Zeo using oePCR (Primers 3 & 4). The second part of the vector was produced by PCR amplification of gBlock1-Zeo (Primers 5 & 6) and the resulting fragments were *Kpn*I/*Bam*HI digested and ligated to obtain the full vector.

Table 3.21 Primers for generating Zeocin resistance based ARS vectors suitable for *Sap*I cloning from synthetic DNA fragments

Name	Sequence
oePrimer1-ZEO	CGTCGCAGAGGAACAGGACTAAGGAGTATACGTAAATATATAATTATATATAATCATATATATGAATAC AATGC
oePrimer2-ZEO	CAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGTACCTAAAAAAAATCCTTAGCTTTC GCTAAGGATAAAATAAATTAAATAAGTTAAATAAAATTAAGTGAATAAAGTTTCAG
oePrimer3-ZEO	CAAGAAGATCCTTTGATCTTTTCTAC
oePrimer4-ZEO	GAAGATTAAGTGAGACCTTCGTTTG
oePrimer5-ZEO	TTTTAGGTACCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATC
oePrimer6-ZEO	CTGAAGGATCCGCACAAACGAAGGTCTCAC

The different promoters were cloned into the *Smi*I and *Eco*RI linearized vector with classical cloning, *i.e.* restriction digest followed by ligation.

The promoter fragments were PCR amplified amplified using the primers from Table 3.22. For the secretory expression vectors an improved, mutated version of the *S. cerevisiae* α-mating factor signal sequence (65) was amplified and joined with the promoter fragments by using overlap extension PCR prior the restriction digest and ligation. Therefore the promoters were amplified with reverse primers adding an overlap to the signal sequence and the oePCR was performed with the primers P-SS-oePCR fwd and P-SS-oePCR rev.

Table 3.22 Primers used for cloning of different promoters with and without signal sequence into the generated ARS vectors

Name	Sequence
pAOX1 fwd	CCTTTTGCTCACATGTATTTAAATAACATCCAAAGACGAAAGGTTG
pAOX1 rev	GATTTGGATCTTACCGGCGCGCCGAATTCGCTCTTCTCATCGTTTCGAATAATTAGTTGTTTTTTG
pAOX1-SS rev	GGTGAAAATAGATGGGAATCTCATCGTTTCGAATAATTAGTTGTTTTTTG
pHTA1 fwd	CCTTTTGCTCACATGTATTTAAATGGGGGGAGGACTCTCGTTTC
pHTA1 rev	GATTTGGATCTTACCGGCGCGCCGAATTCGCTCTTCTCATTGTTGTAGTTTTAATATAGTTTGAGTATGA GATG
pHTA1-SS rev	GGTGAAAATAGATGGGAATCTCATTGTTGTAGTTTTAATATAGTTTGAGTATGAGATG
Dalpha fwd	ATGAGATTCCCATCTATTTTCACC
Dalpha rev	GATTTGGATCTTACCGGCGCGCCGAATTCGCTCTTCTAGCTTCGGCCTCTCTCTTCTC
pUCori-SwaI-pGAP	CCTTTTGCTCACATGTATTTAAATTTTTTGTAGAAATGTCTTGGTGTCCTC
pGAP rev	GATTTGGATCTTACCGGCGCGCCGAATTCGCTCTTCTCATTGTGTTTTGATAGTTGTTCAATTGATTG
pGAP-SS rev	GGTGAAAATAGATGGGAATCTCATTGTGTTTTGATAGTTGTTCAATTGATTG
pCAT1-SS rev	GGTGAAAATAGATGGGAATCTCATTTTAATTGTAAGTCTTGACTAGAGCAAG
P-SS-oePCR fwd	CCTTTTGCTCACATGTATTTAAAT
P-SS-oePCR rev	GATTTGGATCTTACCGGC

Sequences of the gBlocks are available in the supplementary information.

3.30.5 *GUT1* ARS vectors

The *GUT1* based expression vectors were also built similarly to the Zeocin^R based vectors described in section 3.30.4 using overlap extension PCR and classical cloning. The gBlock3-GUT had to be ordered as gene due to synthesis problems and was therefore also PCR amplified (oePrimer3-GUT1 & gBlock3-GUT-fwd).

The *CbARS* was amplified (oePrimer1-GUT1 & oePrimer2-GUT1) and joined with the PCR product of gBlock3-GUT with oePCR (oePrimer3-GUT1 & oePrimer4-GUT1). gBlock 1-GUT and gBlock2-GUT were also assembled together with oePCR (oePrimer4-ZEO & oePrimer6-GUT1) and the two oePCR products were subsequently joined in another round of oePCR (oePrimer3-GUT1 & oePrimer4-ZEO). The resulting fragment was ligated to the PCR product of gBlock1-Zeo (oePrimer7-GUT1 & oePrimer8-GUT1) after digesting the two parts with *Xba*I and *Sac*I.

Table 3.23 Primers used for building *GUT1* based ARS vectors from synthetic DNA fragments

Name	Sequence
oePrimer1-GUT1	CAAAGGGTTGGTTGAAAGATACTGCTTAAGGAGTATACGTAAATATATAATTATATATAATCATATATATGAATACAATGC
oePrimer2-GUT1	CGAAAGCTAAGGATTTTTTTTACTCGAGAAAATAAATTAAATAAGTTAAATAAAATTAAGTGAATAAAGTTTCAG
oePrimer3-GUT1	CAAGAAGATCCTTTGATCTTTTCTAC
oePrimer4-GUT1	CAAAGGGTTGGTTGAAAGATAC
oePrimer4-ZEO	GAAGATTAAGTGAGACCTTCGTTTG
oePrimer6-GUT1	TTAAGCAGTATCTTTCAACCAACC
oePrimer7-GUT1	AACACTCTAGACGCGTCCCGTAGAAAAGATCAAAGGATC
oePrimer8-GUT1	GGTATGAGCTCGCACAAACGAAGGTCTCAC
gBlock3-GUT-fwd	CTCGAGTAAAAAAAATCCTTAGCTTTC

The different promoters were cloned in the same way as for the Zeocin resistance based vector (described in section 3.30.4).
Sequences of the gBlocks are available in the supplementary information.

3.30.6 *Sap*I cloning vectors without *Not*I recognition site

All *Sap*I cloning vectors were also build without the *Not*I recognition site to improve transformation efficiency in *E. coli*. Therefore, the vectors were linearized with *Eco*RI and *Not*I, which removes the stuffer fragment, and the stuffer fragment was replaced by the same fragment without the *Not*I site. The stuffer fragments providing overlaps to the different promoters or the signal sequence were amplified from the corresponding vectors containing the *Not*I site using the primers from Table 3.24 and cloned into the linearized vectors using recombinase cloning.

Table 3.24 Primers used for *Not*I removal from *Sap*I cloning vectors

Primer name	Sequence
SapInoNotI rv	GCATTCTGACATCCTCTTGATTATGAAGAGCTGTTCAATTGAGG
pCAT1-SapI fw	CAAGACTTACAATTAAAATGAGAAGAGC
pGAP-SapI fw	CTATCAAAACACAATGAGAAGAGC
pHTA1-SapI fw	CTATATTAAAACTACAACAATGAGAAGAGC
pAOX1-SapI fw	CTAATTATTCGAAACGATGAGAAGAG
SS-SapI fw	GAGAGGCCGAAGCTAGAAGAG

4 Results and Discussion

4.1 ARS function test

Three different autonomously replicating sequences (ARSs) and fragments thereof were tested for their function as ARS by evaluating the expression levels of eGFP under selective and non-selective conditions as well as their transformation rate.

The well known *PARS1* (23) (NCBI Accession number M11199), a 192 bp long AT-rich fragment, directly upstream of $P_{CAT1\text{-}500}$ (*CAT1-ARS*) and the promoter region of the *Candida boidinii methanol oxidase* (*AOD1*) gene (P_{CbAOD1}), which exhibited ARS function, but no expression in *P. pastoris* in previous tests (Vogl T., Dissertation chapters 3.2 and 3.3), were tested.

P_{CbAOD1} was amplified from *Candida boidinii* DSM 70026 genomic DNA and showed sequence differences compared to the gene reported in literature (NCBI Accession number M81702) from the *Candida boidinii* S2 strain. A detailed alignment of the two sequences is shown in Section 7.1.

In order to keep the vector as small as possible different fragments of P_{CbAOD1} were designed. Searching for the consensus motif WTTTAYRTTTW, which is present in all ARSs (66), yielded many hits. The motif was found twice with 100% identity, 14 times with at least 90% and 71 times with at least 80% identity.

Because of the many hits the fragments were designed based on the AT-content (Figure 4.2). Also the GC-rich fragment was included into the tests, since ARSs are not always AT-rich (67).

The different ARSs and fragments thereof were cloned into the reporter plasmid (an example containing *PARS1* is shown in Figure 4.1) between the *AOD* terminator and the *E. coli* origin of replication.

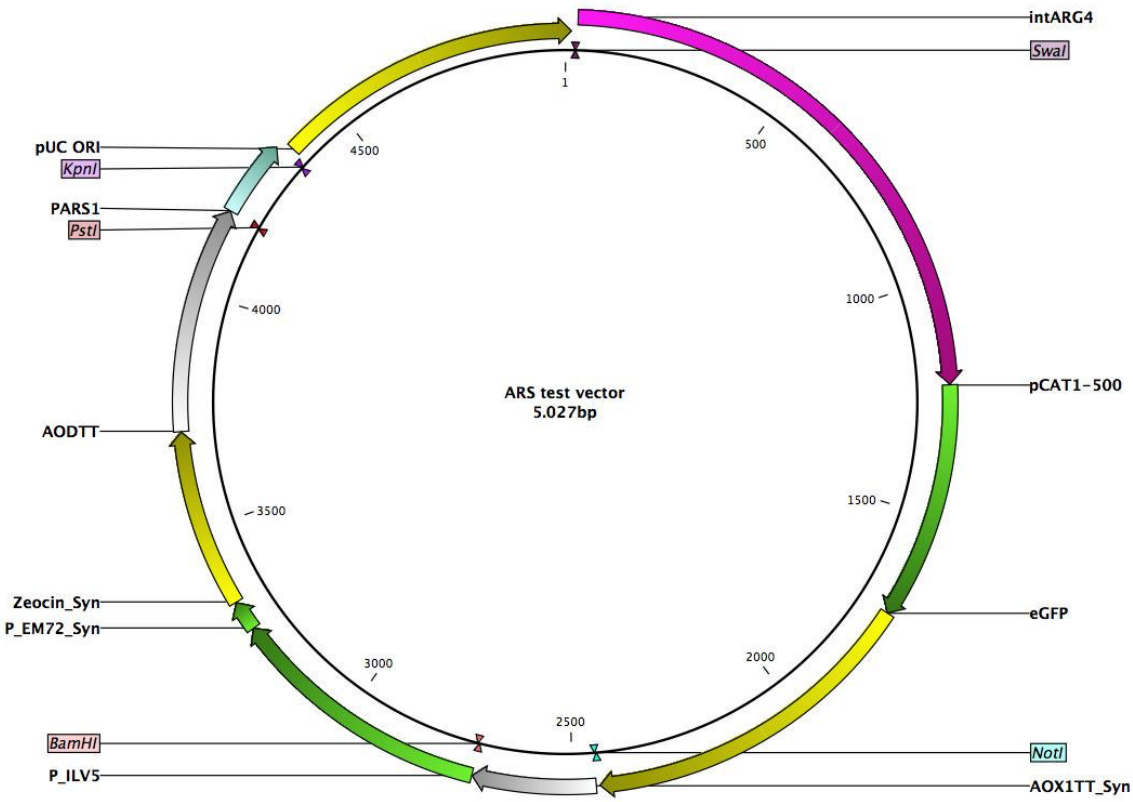

Figure 4.1 ARS reporter plasmid example containing *PARS1*

All of the designed fragments (Figure 4.2) showed similar transformation efficiencies as well as *PARS1* and *CAT1-ARS*, except for the GC-rich (F2), which only gave a few transformants that were also growing slower compared to the other constructs.

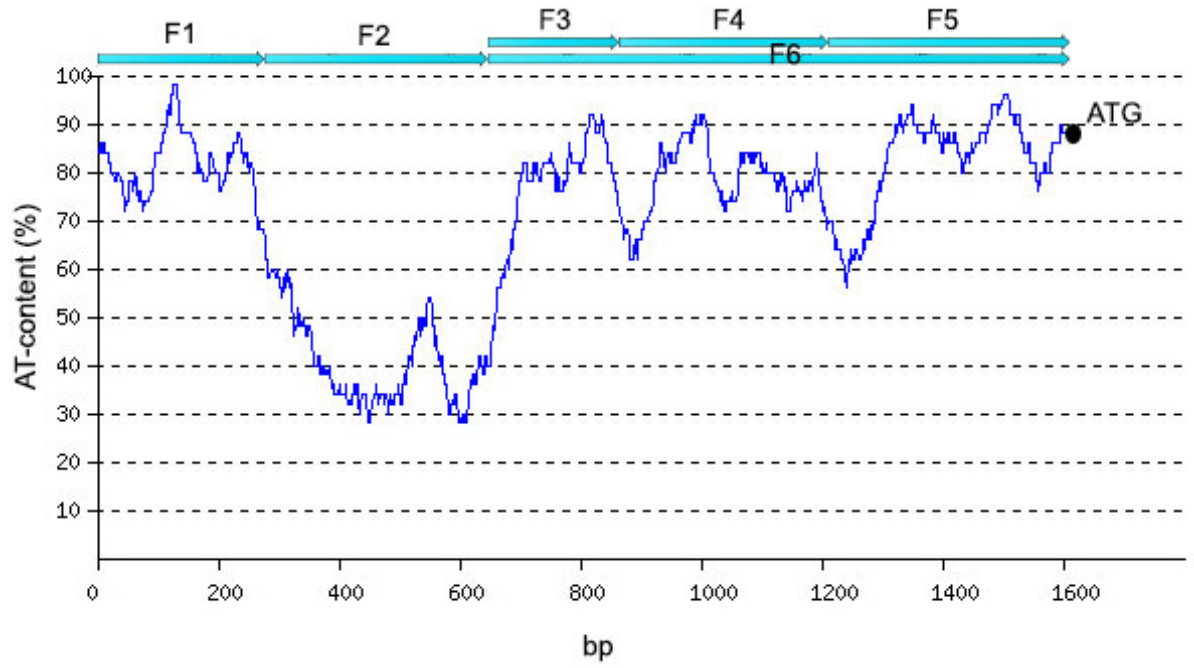

Figure 4.2 AT-content of the Cb_{AOD1} promoter region and designed fragments for ARS test. Analysis made with BitGene Gene Analysis using a 50 bp window.

Seven individual transformants per construct were chosen for the screening under selective and non-selective conditions and compared to an integrated control (Figure 4.3). Most of the constructs yielded eGFP expression levels similar to the integrated control when cultivated under non-selective conditions and significantly higher expression levels under selective conditions.

P_{CbAOD1} fragment 2 showed no expression under non-selective conditions and only very weak fluorescence with selection pressure. Also *CAT1-ARS* yielded only weak fluorescence, but in contrast to P_{CbAOD1} fragment 2 many transformants could be obtained and the expression was also higher under selective conditions. That could be a valuable tool for proteins that need weaker transcription to be efficiently expressed.

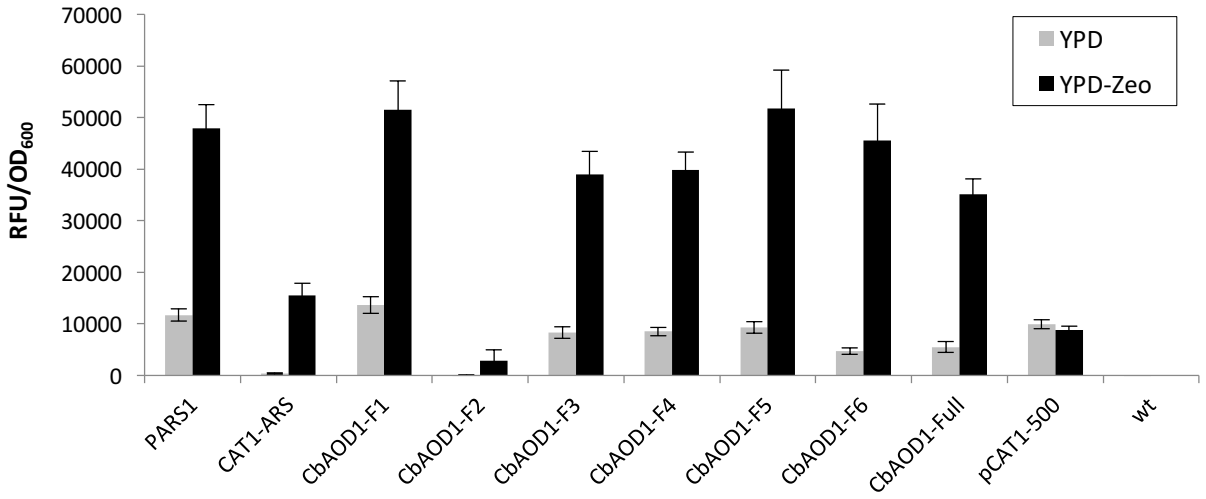

Figure 4.3 ARS function test using P_{CAT1} driven eGFP expression values after 60 h of cultivation. Mean values and standard deviations from biological seven-fold replicates are shown. AOD, P_{CbAOD1}; F, fragment number; CAT1-500, integrated control without ARS ($P_{CAT1\text{-}500}$).

The raised expression levels under selective conditions indicate the presence of the plasmids in multiple copies, which was already reported for *PARS1* containing plasmids (23).

We hypothesized that the copy number can be influenced by the strength of the selection pressure applied and maybe also by using different promoters to drive the expression of the marker gene. Therefore the same clones were tested under different Zeocin concentrations ranging from 0 to 300 µg/ml after pre-growth in media with 50 µg/ml Zeocin (Figure 4.4).

This time also the 692 bp long P_{CAT1} version was included, since the *CAT1-ARS* maybe needs some additional bp downstream to work efficiently.

As expected the expression levels were raised even further when applying higher Zeocin concentrations. The integrated control showed constant expression under all different conditions.

Interestingly the expression levels under non-selective conditions were higher compared to the previous screening indicating higher plasmid stability and/or copy number after pre-growth under selection pressure.

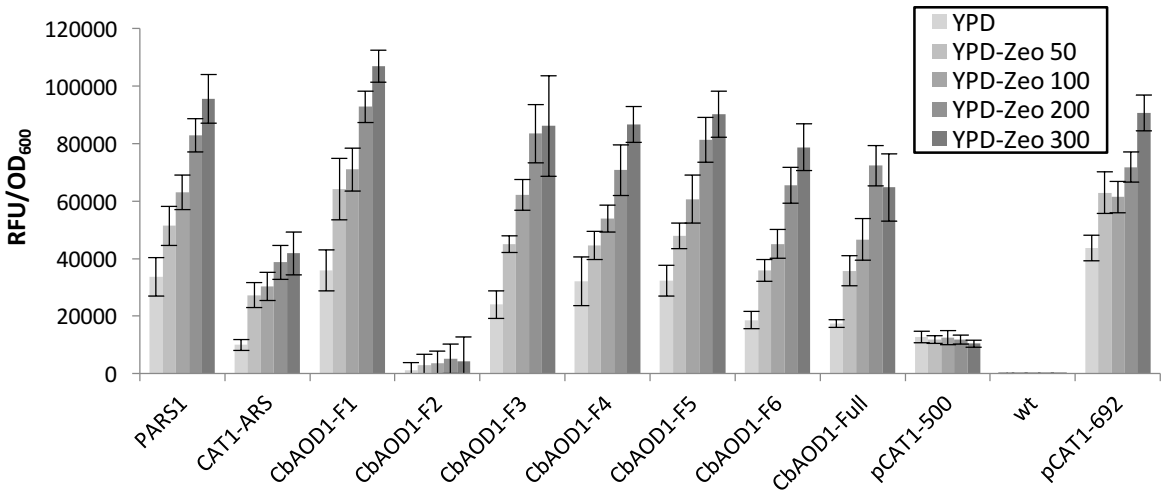

Figure 4.4 ARS function test with different Zeocin concentrations using P_{CAT1} driven eGFP expression values after 60 h of cultivation. A pre-culture with 50 µg/ml Zeocin was made and the DWPs with different Zeocin concentrations inoculated with 2 µl from the pre-culture. The numbers after Zeo indicate the Zeocin concentration present in the media in µg/ml. Mean values and standard deviations from biological seven-fold replicates are shown. AOD, CbAOD1; F, fragment number; CAT1-500, integrated control without ARS.

The fragment 1 of P_{CbAOD1} showed the most promising results, since it is smaller than most of the other fragments and also showed higher expression. Therefore, further work was just continued with this fragment (called *CbARS* in the following sections) and the other 5 fragments as well as the full length P_{CbAOD1} were not used for other tests.

Nevertheless, it has to be mentioned that after 60 h cultivation time in glucose limited media all cells are in a starvation phase and no further C-source was provided for energy production to support the protein expression. Therefore, no conclusions can be drawn about specific productivities of the different clones with different ARSs. Further experiments employing methanol induction or a constitutive promoter for eGFP expression could provide valuable information about the full potential and differences between these tested vectors.

4.2 Transcription terminator function test

Since ARSs and transcription terminators are often overlapping in yeast (66), the different ARSs were also tested for their transcription termination function and compared to the strongest homologous and heterologous terminator (*Not*I*AOX*TT* and *S. cerevisiae SPG5TT*) as well as the terminators present in the pPpT4_S and pPpGUT1 plasmids (18) (*AOD_TT* and *GUT1_TT*). Additionally to the *CbARS* also *PARS1*, *CAT1-ARS* and $P_{CAT1\text{-}692}$ were tested. $P_{CAT1\text{-}692}$

was tested with and without core promoter to avoid transcription run through. The 3' end of the promoter beginning with the TATA box was deleted, because transcription would directly run into the promoter and coding sequence of the selection marker, which could interfere with the plasmid functionality.

An example for the transcription terminator reporter plasmid containing *PARS1* as terminator is shown in Figure 4.5.

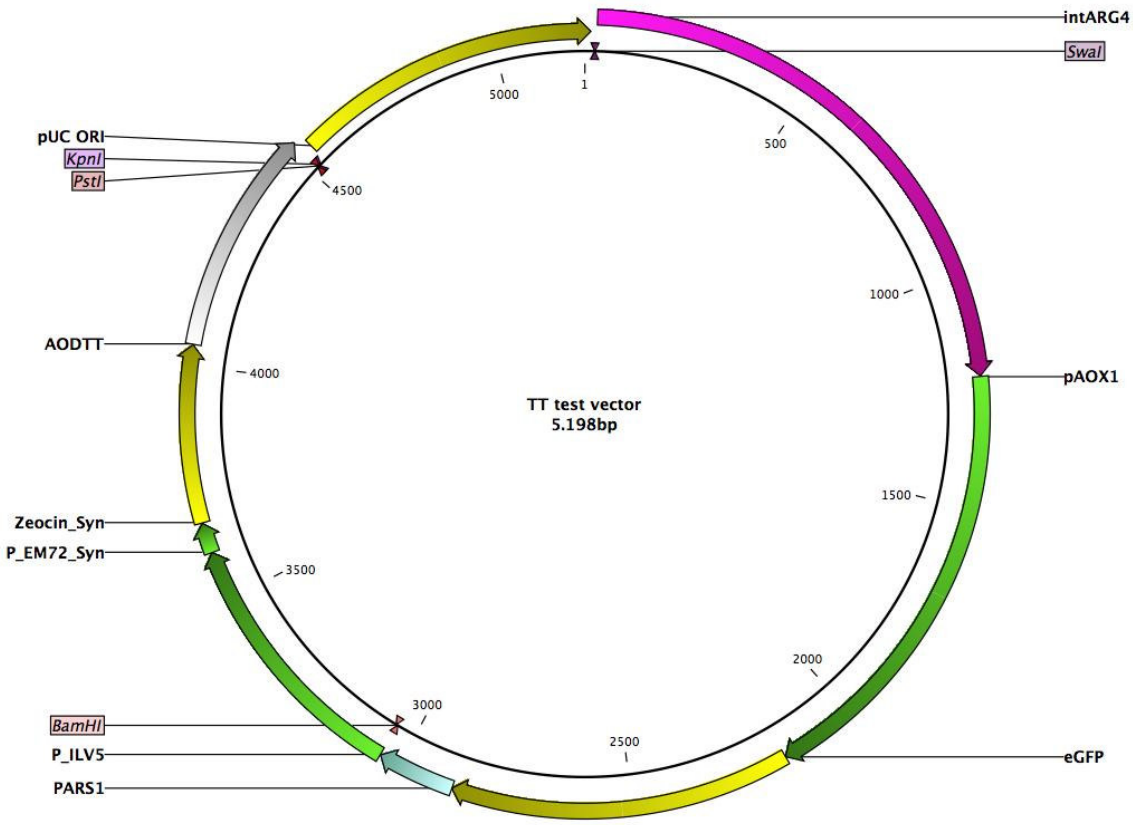

Figure 4.5 Transcription terminator reporter plasmid containing *PARS1* as transcription terminator

42 individual transformants were screened per construct, four representative clones chosen for a re-screening and cultivated in seven-fold replicates. In the end one representative clone was chosen for a final screening in which every single construct was cultivated on the same plates in seven-fold replicates.

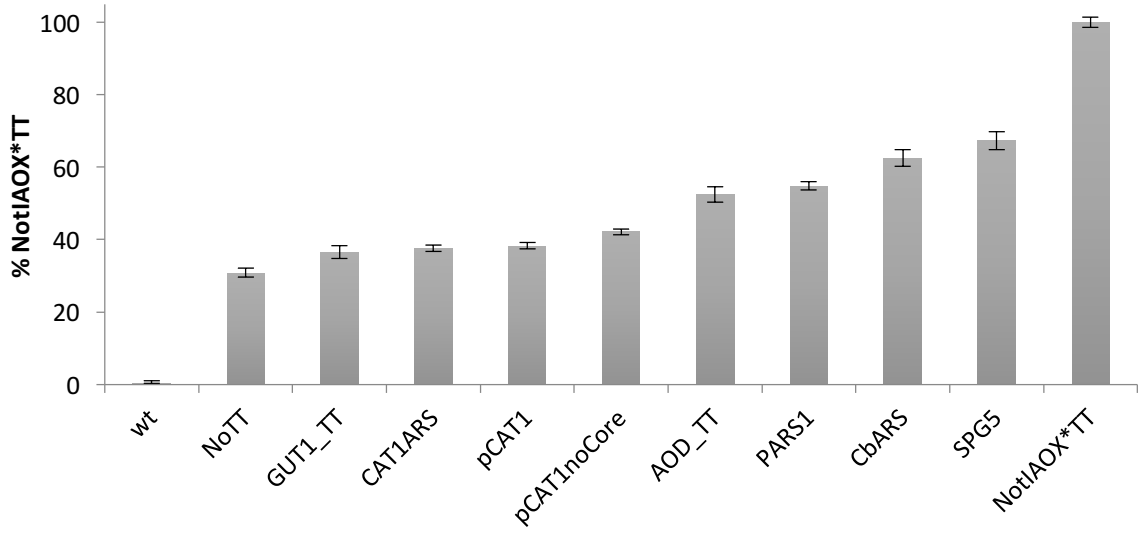

Figure 4.6 Transcription terminator test. P_{AOX1} driven eGFP expression values (in % eGFP fluorescence of the strongest terminator) after 108 h of cultivation (60 h + 48 h induction) are shown. Representative clones identified by re-screening were cultivated in seven-fold replicates. Mean values and standard deviations are shown.

The *CbARS* performed also well as transcription terminator yielding 62.5% of the strongest terminator (*Not*I*AOX*TT*) and almost as much as the strongest heterologous terminator (*ScSPG5TT*) with 67.3%. The ARS was even outperforming the terminators that are present in pPpT4_S (*AOD_TT*) and pPpGUT1 (*GUT1_TT*) plasmids.

The *Not*I site increased terminator efficiency in previous tests. The expression level of the reporter gene was raised by 37% for the *AOX1* terminator with *Not*I site compared to without the restriction site (Vogl *et al.*, ACS Synbio submitted, minor revisions). Therefore, the *CbARS* could reach similar levels to the *AOX1* terminator without the *Not*I site.

All variants of P_{CAT1} and *CAT1-ARS* exhibited significantly weaker signals, just 6-12% above the control strain without a terminator. Since *PARS1* also performed well it was also used in addition to *CbARS* to replace the *AOD_TT* in the pPpT4_S based plasmids.

4.3 Bifunctional vectors using the ARSs as terminator simultaneously

Since both ARSs, the *PARS1* and *CbARS*, exhibited high signals in the ARS and transcription terminator tests, they were cloned instead of *AOD_TT* to be used as a replication origin and

transcription terminator of the marker gene simultaneously. This design minimized the vector size and should help to improve the transformation efficiency even further.
The design of the vector (containing *PARS1*) is shown in Figure 4.7.

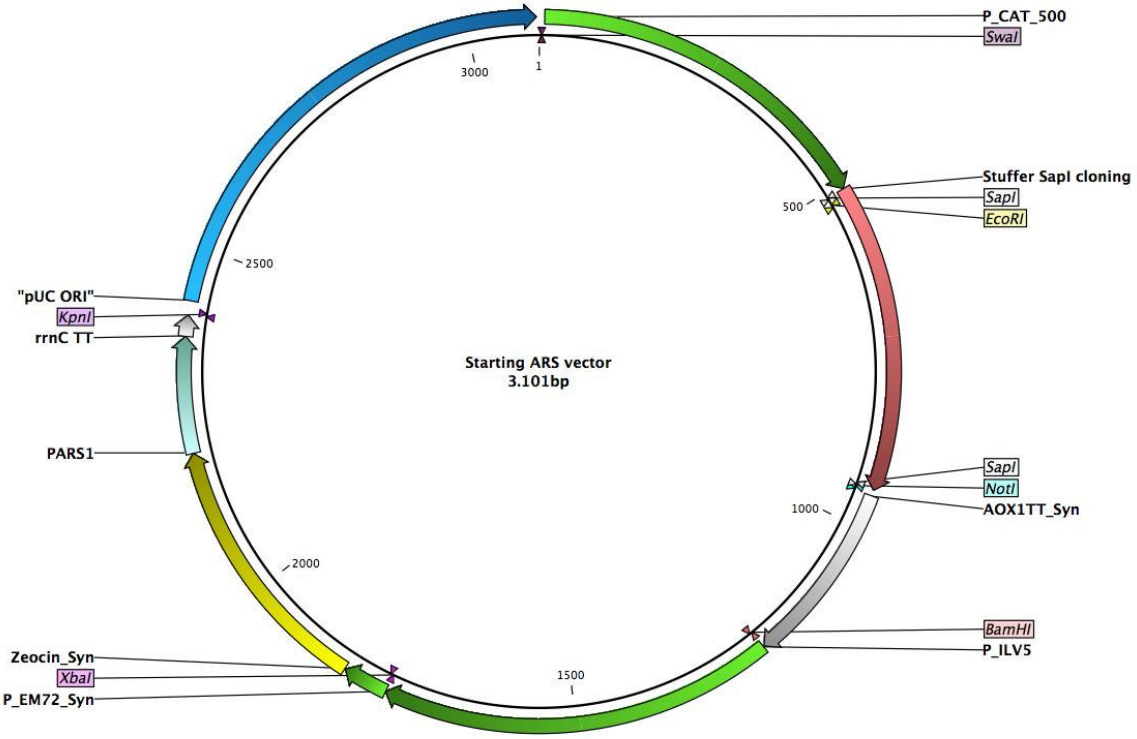

Figure 4.7 Bifunctional *Sap*I cloning ARS vector

The vectors were modified for *Sap*I cloning and eGFP was cloned downstream of the P_{CAT1}. The transformants were cultivated under selective and non-selective conditions (Figure 4.8). Even though *PARS1* performed well when used as ARS or transcription terminator separately, it showed clear deficiencies compared to *CbARS* when used as both simultaneously.
Under non-selective conditions the *PARS1* containing expression constructs only yielded low eGFP expression (mean value approx. 1300 RFU/OD_{600}) with a high relative standard deviation of 46.3% indicating plasmid instability. *CbARS* plasmid showed 13-fold higher and 3-fold more uniform expression.

Under selection pressure the *PARS1* plasmids were also more stable exhibiting roughly the same clonal variation as *CbARS* plasmids, but still 4-fold lower expression levels.

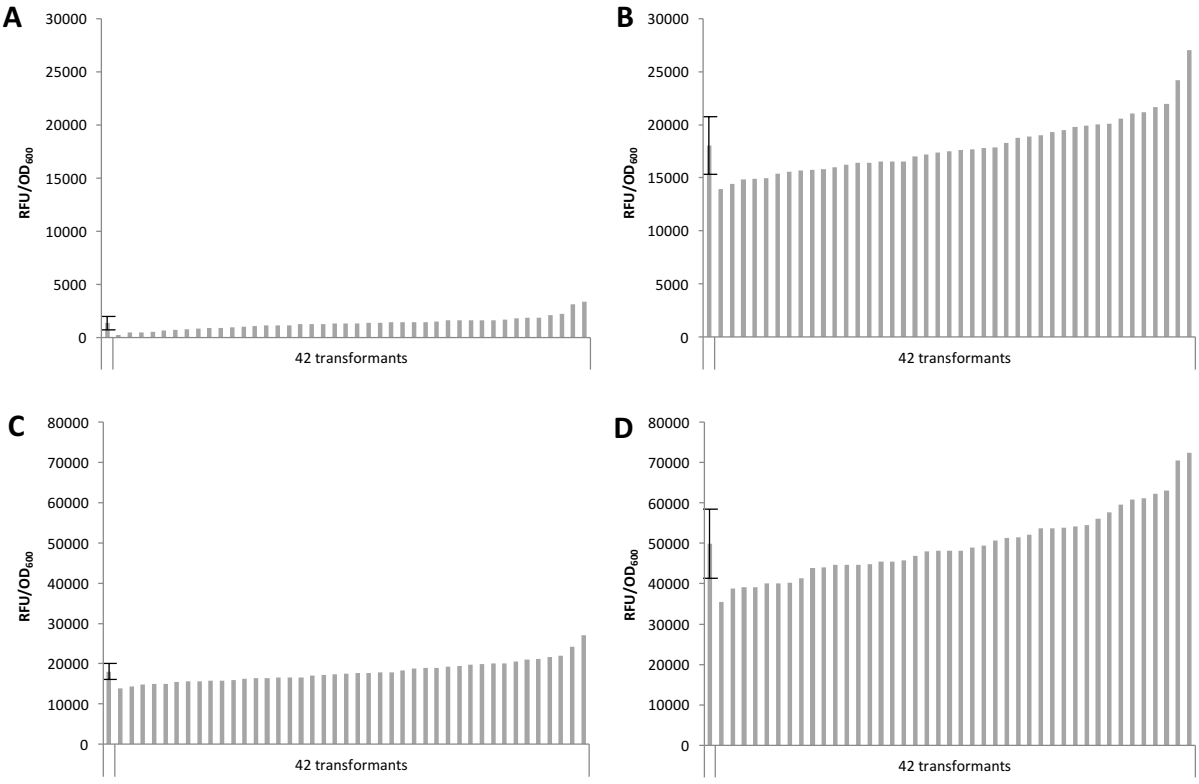

Figure 4.8 *PARS1* and *CbARS* used as transcription terminator and origin of replication simultaneously cultivated without (YPD) and with (YPD-Zeo) selection pressure. P_{CAT1} driven eGFP expression values after 60 h of cultivation.are shown. **A**, *PARS1* without selection pressure; **B**, *CbARS* without selection pressure; **C**, *PARS1* with selection pressure; **D**, *CbARS* with selection pressure. The first bar on the left indicates the mean value and the error bar represents the standard deviation.

Since the *CbARS* turned out to be more suitable for the intended vector design, further *Sap*I cloning vectors were build using only *CbARS*.

However, these landscapes might look different if other promoters are used to drive transcription of the gene of interest. In case of the *CAT1* promoter and without any further C-source feeding protein expression levels after 12 – 24 h of carbon source depletion are measured, which might also contribute to the low standard deviations observed in these experiments.

4.4 *Sap*I cloning

The final expression vectors were designed to use the IP-free *Sap*I cloning system, which is based on the type IIS restriction enzyme *Sap*I, and allows seamless cloning (Figure 4.9).
*Sap*I recognizes the 7 base recognition site 5' GCTCTTC 3' and cuts 1 base after it, leaving a 3 base overhang on the other strand.

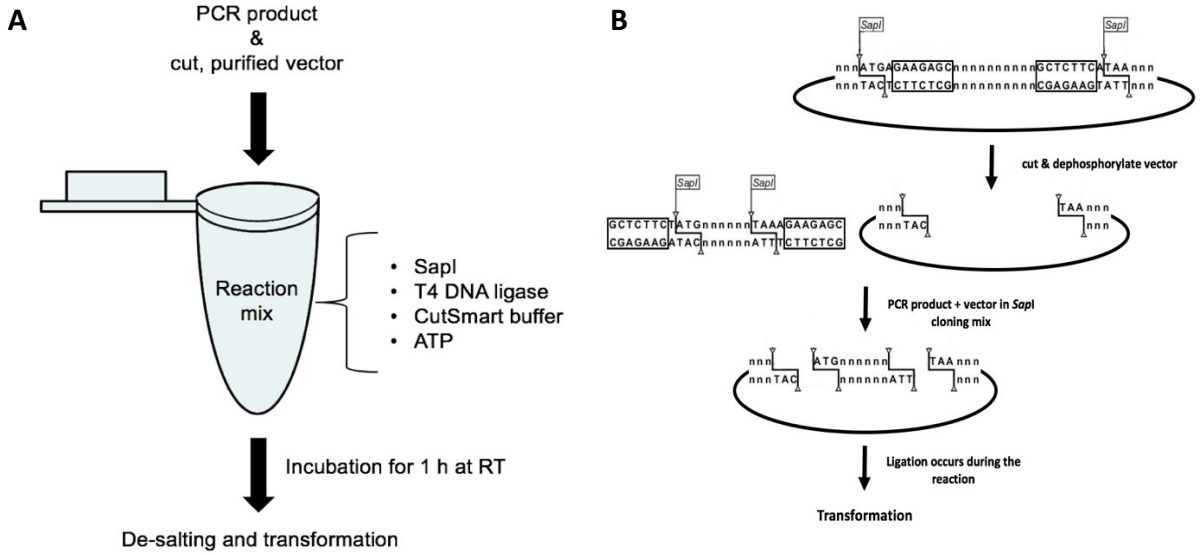

Figure 4.9 *Sap*I cloning set-up (**A**) and reaction principle (**B**). The boxes represent the *Sap*I recognition sites.

The cloning system is easy to set up, since it is carried out in a one pot reaction and incubation at room temperature. This is possible due to the fact that the *Sap*I recognition site is a long non palindromic sequence and if the recognition sites are not created on DNA parts of the final construct (vector & insert) there is no way to cut the final product by SapI, while the starting material is in an equilibrium between cut and uncut status.
E. coli transformation yielded just approximately 10^3-10^4 CFU/µg of vector backbone, but still most of the colonies were positive when the vector was dephosphorylated. Dephosphorylation of the vectors led to a lower transformation efficiency, but removed most of the background colonies. Nearly all transformants tested contained the correct insert.
Since there is a *NotI* site in front of the terminator and directly after the stop codon, where annealing takes place, it could have an influence on the cloning efficiency because of the high probability to form secondary structures.

The *Not*I site was removed from the expression vectors and cloning was tested using P_{GAP} with and without *Not*I and also with and without dephosphorylation (Figure 4.10).
The transformation efficiency was generally higher using the vector without *Not*I, but since the *Not*I site increases the terminator efficiency and there were still enough colonies on the plate, I would recommend to use the vectors containing the *Not*I site.

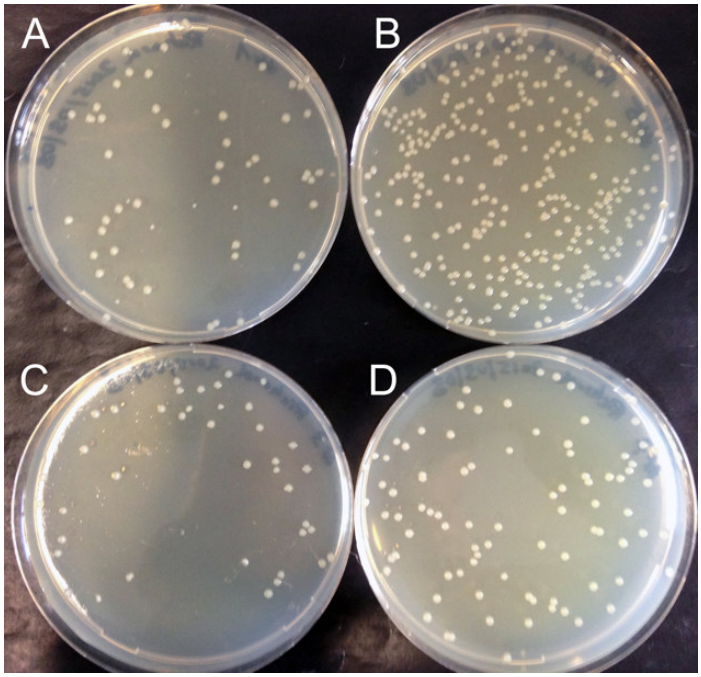

Figure 4.10 *Sap*I cloning (*TinselPurple* into intracellular ZEO-P_{GAP} expression vectors) *E. coli* transformation. **A**, with *Not*I; **B**, without *Not*I; **C**, with *Not*I, dephosphorylated; **D**, without *Not*I, dephosphorylated.

4.5 P_{GAP} engineering

The final vectors should provide a range of promoters offering a different regulatory profile including P_{GAP}, which inconveniently contained a *Sap*I site. One position inside the *Sap*I recognition site was targeted and exchanged to all other nucleotides. The substitution to a T is also present among other additional mutations in a P_{GAP} variant that did not show weaker expression regarding mRNA concentration and reporter protein measurement (68).
The three different mutants and the wildtype P_{GAP} promoter were tested in a Zeocin selection vector containing *CbARS* and the expression was measured after 24, 48 and 60 h (Figure 4.11).

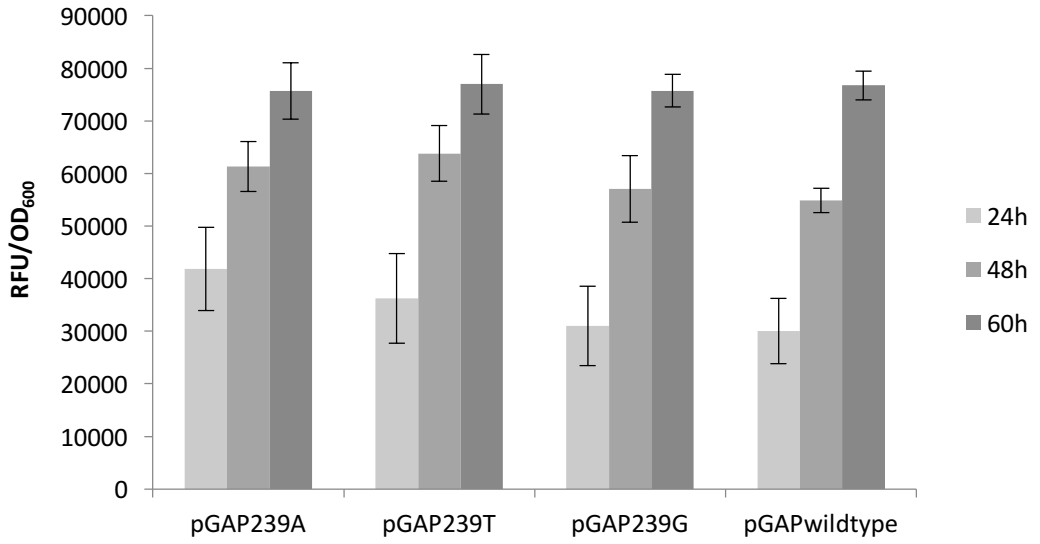

Figure 4.11 P_{GAP} engineering. eGFP expression values after 24, 48 and 60 h of cultivation in selective media. Mean values and standard deviations from 21 individual clones are shown.

Small differences between the variants were only visible at the earlier time points, which might have arisen because of growth differences and different inoculation. These differences are also smaller than the deviations in individual experiments and therefore not significant. After 60 h all tested variants reached the same expression level and $P_{GAP239T}$ was chosen for further work and cloning into the final expression vectors.

4.6 Characterization of final expression vectors

4.6.1 Expression uniformity

All intracellular expression vectors (containing P_{CAT1}, P_{GAP}, P_{HTA1} and P_{AOX1} for Zeocin resistance based vectors and P_{CAT1} and P_{HTA1} for *GUT1* based vectors) were tested for their expression uniformity using eGFP and comparing the expression landscapes of 42 clones. The plasmids were transformed circular and also *Swa*I linearized and gel purified for integration. The results for the derepressed P_{CAT1} driven expression (Zeocin resistance based) are shown in Figure 4.12, the results for all other expression vectors and methanol induced expression from the P_{CAT1} vector are shown in the figures 4.13 to 4.18. For some of the vectors it was hard to discriminate big colonies that have integrated the expression cassette from small colonies, *i.e.*

background colonies that can still maintain the expression vector as episomal element after linearization.

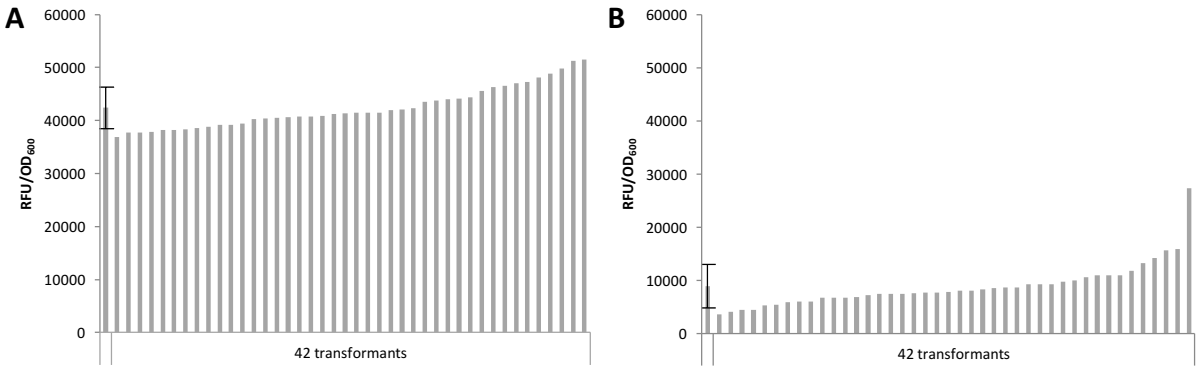

Figure 4.12 P_{CAT1} driven eGFP expression landscape comparison of episomal and integrated expression strains. Expression values from Zeo^R vectors after 60 h of cultivation are shown. **A**, episomal expression; **B**, genomic expression. The first bar on the left indicates the mean value and the error bar represents the standard deviation.

The strains generated by integrating the vector showed large clonal variety and lower expression levels. The mean expression was approximately 8800 RFU/OD_{600} with a relative standard deviation of 46%.

In contrast to that the episomal expression strains showed about 4.8-fold higher expression (42000 RFU/OD_{600}) and 5-times less clonal variation (9% relative standard deviation).

The results for the P_{CAT1} driven expression were the most reliable, since no background colonies were picked according to the fluorescence levels. Beside the fact that not only integrative expression strains were picked for the other vectors, the results were similar. Higher and more uniform expression levels were also obtained with all the other vectors (Figures 4.14 to 4.18) for episomal expression.

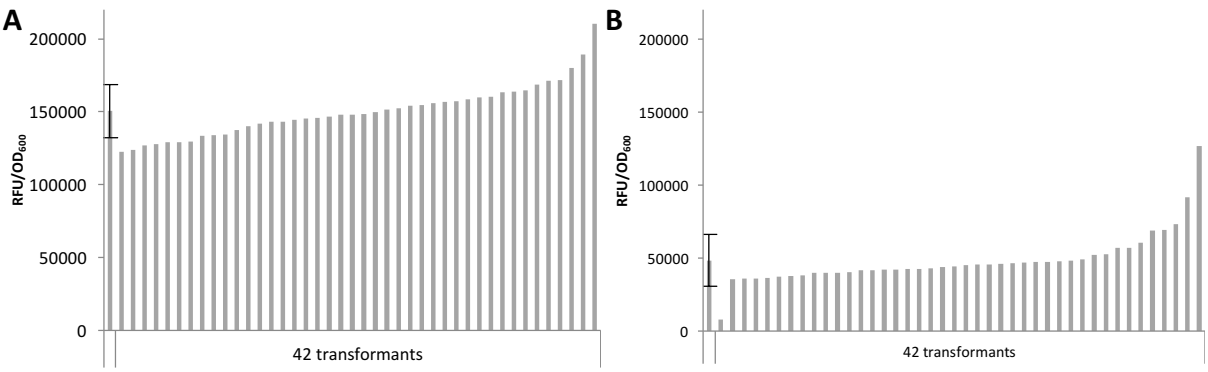

Figure 4.13 P_{CAT1} driven eGFP expression landscape comparison of episomal and integrated expression strains. Expression values from Zeo^R vectors after 108 h (60 h + 48 h methanol induction) of cultivation in selective media are shown. **A**, episomal expression; **B**, genomic expression. The first bar on the left indicates the mean value and the error bar represents the standard deviation.

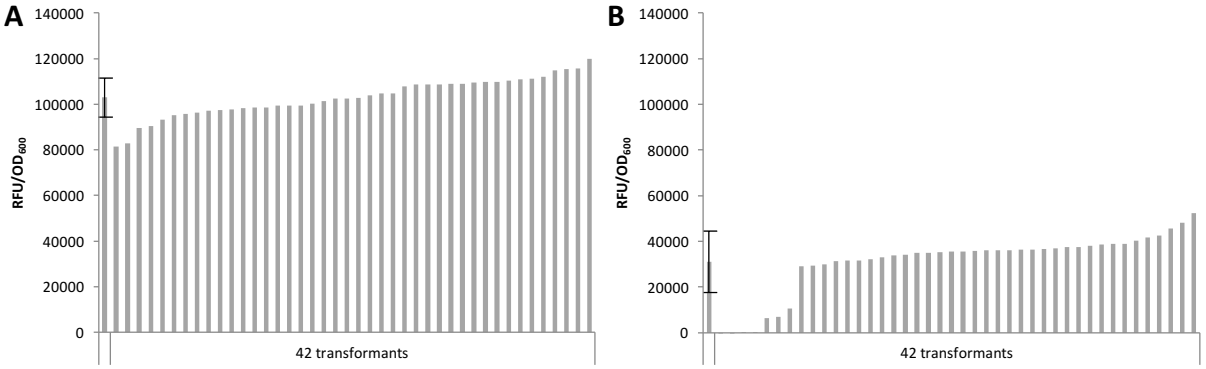

Figure 4.14 P_{GAP} driven eGFP expression landscape comparison of episomal and integrated expression strains. Expression values from Zeo^R vectors after 60 h of cultivation in selective media are shown. **A**, episomal expression; **B**, genomic expression. The first bar on the left indicates the mean value and the error bar represents the standard deviation.

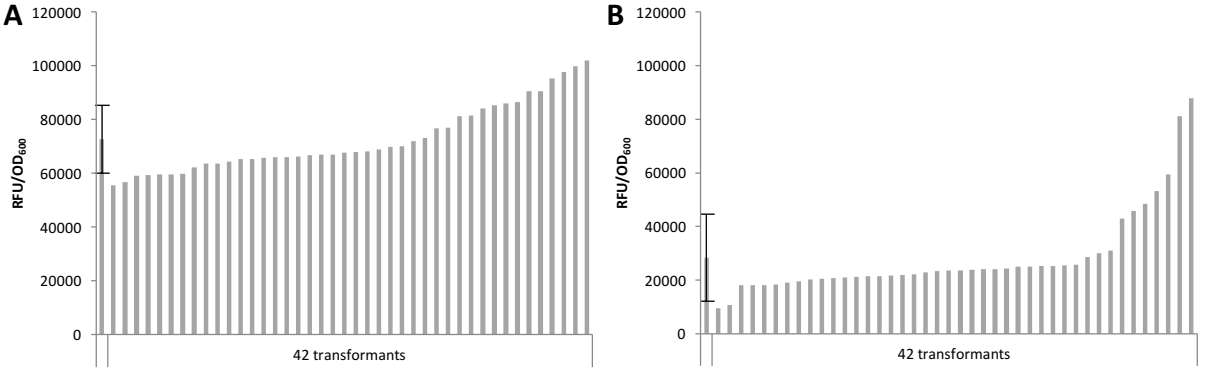

Figure 4.15 P_{HTA1} driven eGFP expression landscape comparison of episomal and integrated expression strains. Expression values from Zeo^R vectors after 60 h of cultivation in selective media are shown. **A**, episomal expression; **B**, genomic expression. The first bar on the left indicates the mean value and the error bar represents the standard deviation.

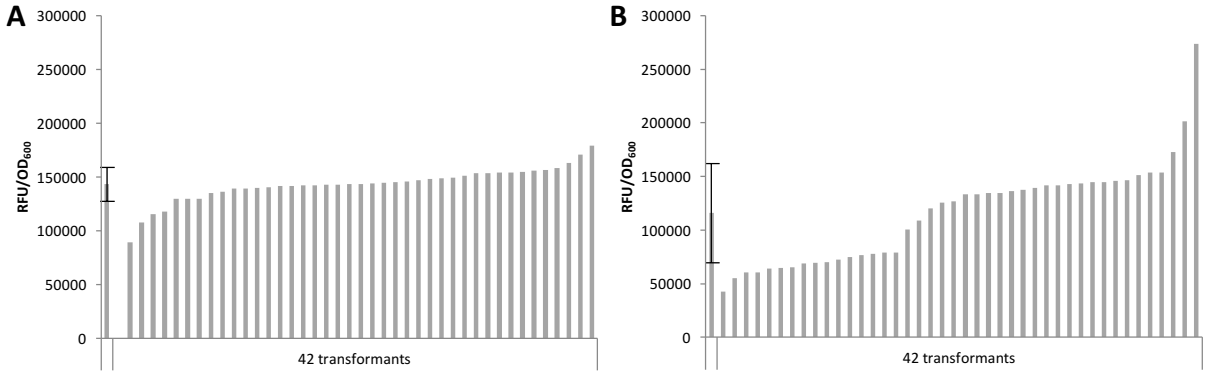

Figure 4.16 P_{AOX1} driven eGFP expression landscape comparison of episomal and integrated expression strains. Expression values from Zeo^R vectors after 108 h (48 h induction) of cultivation in selective media are shown. **A**, episomal expression; **B**, genomic expression. The first bar on the left indicates the mean value and the error bar represents the standard deviation.

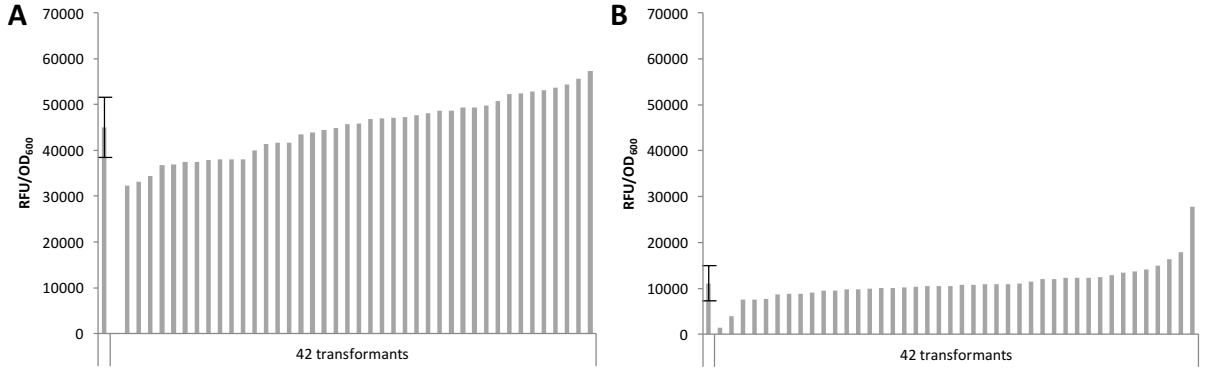

Figure 4.17 P_{CAT1} driven eGFP expression landscape comparison of episomal and integrated expression strains. Expression values from *GUT1* expression vectors after 60 h of cultivation in BMG are shown. **A**, episomal expression; **B**, genomic expression. The first bar on the left indicates the mean value and the error bar represents the standard deviation.

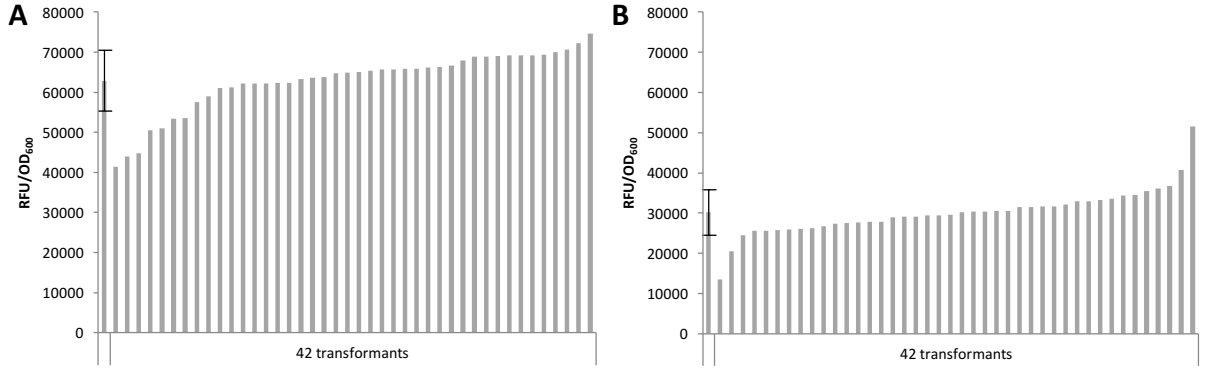

Figure 4.18 Expression landscape comparison of episomal and integrated expression strains. P_{HTA1} driven eGFP expression values from *GUT1* expression vectors after 60 h of cultivation in BMG are shown. **A**, episomal expression; **B**, genomic expression. The first bar on the left indicates the mean value and the error bar represents the standard deviation.

4.6.2 eGFP expression

In order to compare the vectors with each other some representative clones from the middle of the landscapes were chosen and re-streaked for single colonies. These clones were cultivated in seven-fold replicates on the same DWP to make it more comparable and rule out differences arising from cultivation.

An example for the Zeocin resistance based epression vectors was shown in Figure 4.7 and the design of the *GUT1* based vectors is shown in Figure 4.19. The *CbARS* ($P_{CbAOD1\text{-}F1}$) served as autonomously replicating sequence and also as transcription terminator for the *GUT1* gene.

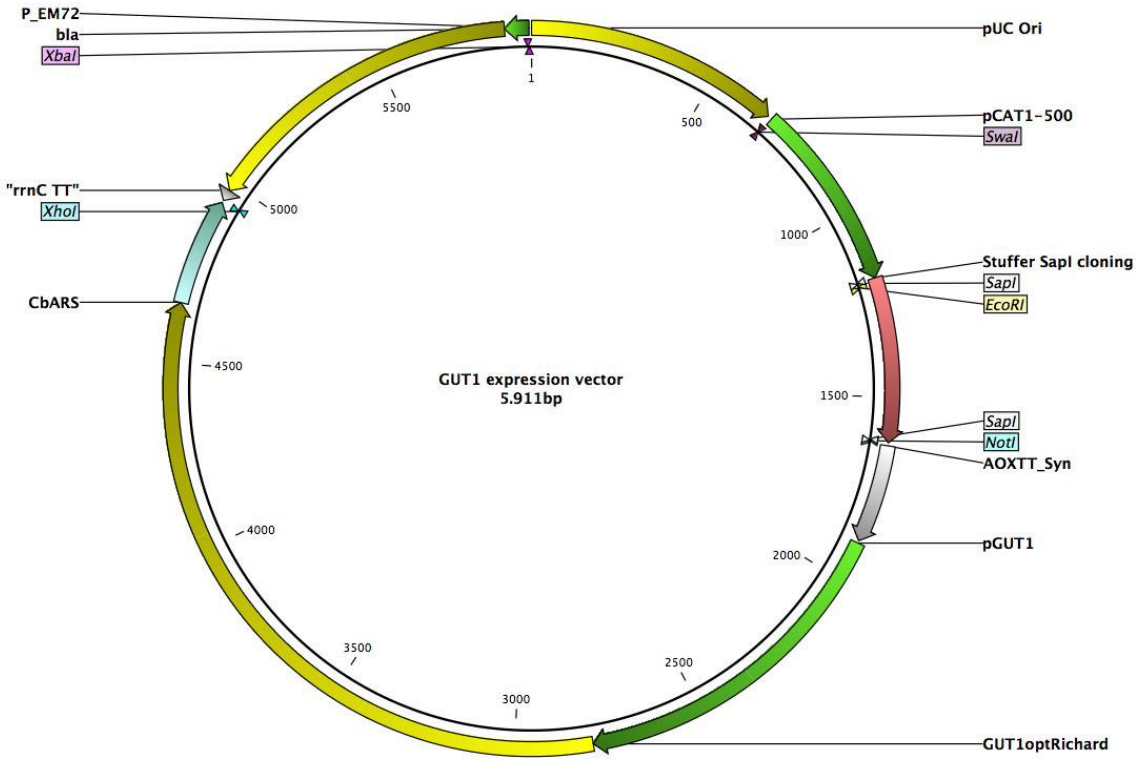

Figure 4.19 *GUT1* based *Sap*I cloning vector

Figure 4.20 shows the different range of expression levels and regulatory profiles obtained when using the different vectors for genomic integration or episomal expression. Induction

with methanol should not be performed for the *GUT1* based vectors since the expression was weaker for episomal constructs, because of the loss of selection pressure. The cells can grow on methanol and have no need for the plasmid anymore. A possibility to avoid that would be using a *Mut⁻ ΔGUT1* strain as host or using genomic integration, since the selection pressure has no effect on integrated constructs.

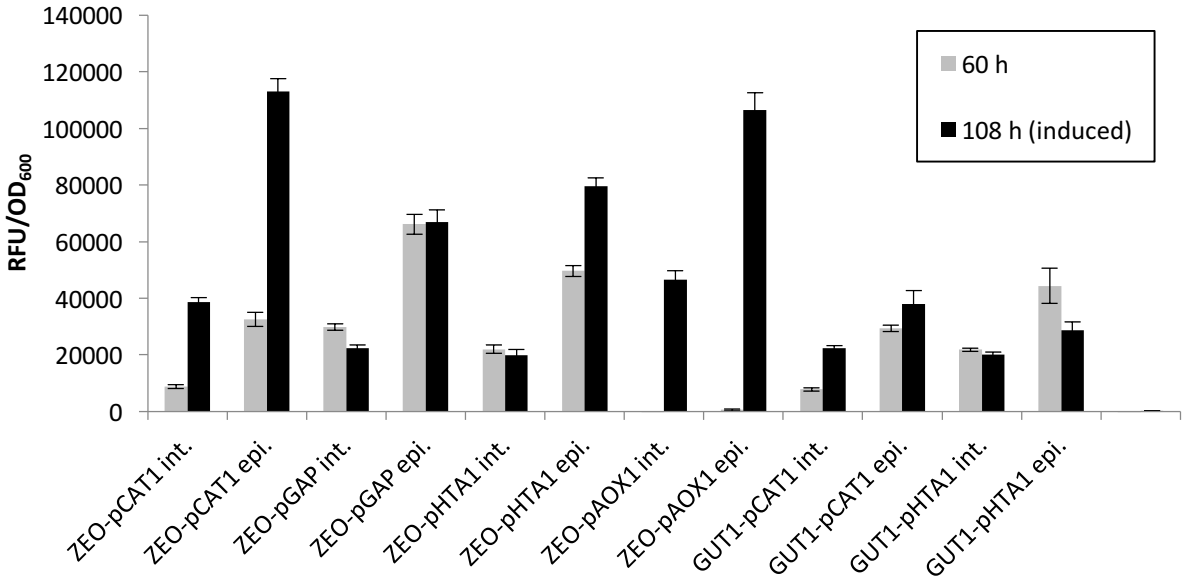

Figure 4.20 Expression profile characterization of the different ARS expression vectors. Derepressed expression values after 60 h of cultivation and induced expression levels after 48 h of induction deriving from biological seven-fold replicates from 96 deep well plate experiments are shown. *GUT1* expression constructs were cultivated in BMG and induced with BMM2 and BMM10. Zeocin expression constructs were cultivated in YPD-Zeo and induced with YPM2-Zeo and YPM10.

4.6.3 Transformation rate

ARS plasmids show much higher transformation efficiencies than their integrated counterpart. The transformation efficiency for the Zeocin resistance based vector enabling P_{CAT1} driven eGFP expression and the two different ARSs (*PARS1* and *CbARS*) were compared to *Sac*I linearized pPpT4_S plasmid, which served as a genomic integration control.

Three transformations per construct using 100 ng were performed and appropriate dilutions plated onto selective media.

The transformation efficiencies calculated as CFU/µg (Table 4.1) were much higher for the episomal expression vectors and also a difference between the two ARSs was observed. The

transformation efficiency of the *CbARS* based vector was more than 1.5-fold higher than of the *PARS1* based vector and approximately 146 times higher than for the integrative vector.

Table 4.1 Transformation rates using episomal expression vectors and integrative plasmids. Mean values and standard deviations were calculated from three individual transformations.

Construct	CFU/µg	MV	SD
pPpT4-S *Sac*I	6700	4900	1600
	4300		
	3700		
PARS1-ZEO-pCAT1-eGFP	455000	460000	68000
	530000		
	395000		
CbARS-ZEO-pCAT1-eGFP	645000	715000	64000
	770000		
	730000		

Another difference between the two ARSs was observed regarding the size of the colonies formed after transformation. The colonies obtained from transformants harboring the *PARS1* based expression vector were smaller and also less uniform compared to the *CbARS* vector containing transformants (Figure 4.21). This also supports the theory that the *PARS1* based vectors are less stable when *PARS1* is used as terminator and replication origin, which was indicated by the eGFP expression values shown in section 4.3.

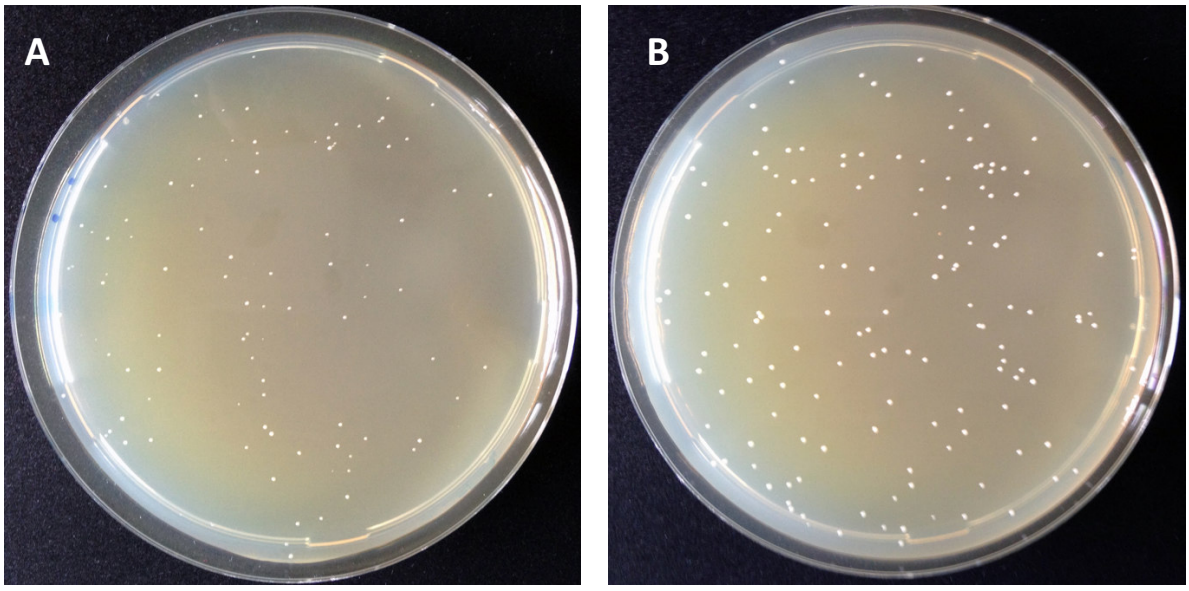

Figure 4.21 Morphology of transformants carrying *PARS1* (**A**) and *CbARS* (**B**) based episomal expression constructs when the ARSs are used as terminators. Pictures were taken 48 h after plating on YPD-Zeo (50 µg/ml).

4.6.4 Uptake of multiple plasmids per cell

With the high transformation efficiency and the uniform expression levels two important features for enzyme engineering were already provided by the new ARS vectors. For engineering purposes it is also important that there is only one variant per transformant and not a mixture of different ones. Only if just one variant is entering one cell improved variants can be clearly identified.

This was tested by transforming a 1:1 mixture of ARS plasmids containing either eGFP or dTomato under the control of P_{GAP}. Different amounts ranging from 1 to 100 ng of plasmid DNA and 80 µl cells were used for transformation and 84 clones from each transformation picked for screening (Figure 4.22).

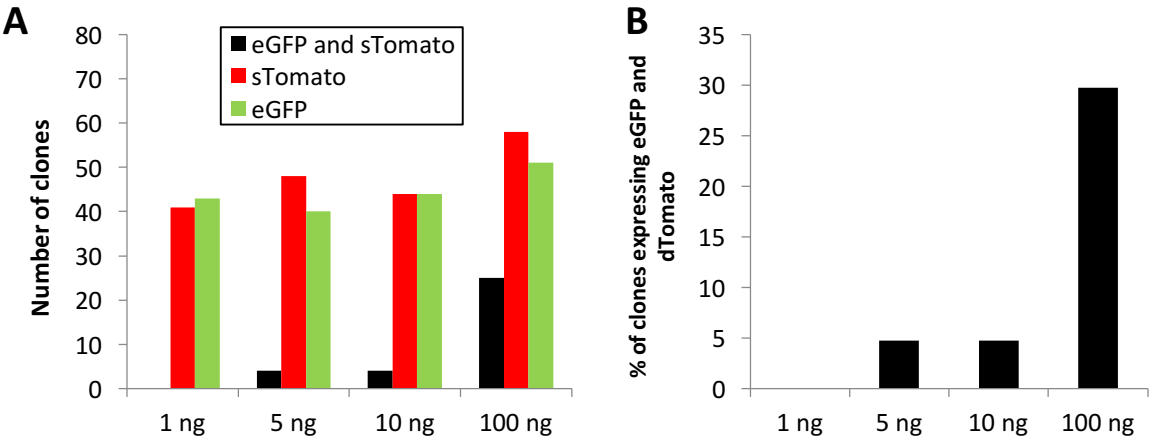

Figure 4.22 Number of transformants carrying more than one plasmid type after transformation with a 1:1 mixture of eGFP and dTomato expression plasmids. Different amounts of DNA were transformed and 84 transformants from each transformation picked for screening.

When transforming just 1 ng of plasmid DNA none of the picked transformants expressed both fluorescent proteins. Even with higher amounts of DNA up to 10 ng just 4 transformants (<5%) harbored both plasmids. But when 100 ng were transformed already approximately 30% of all transformants carried both plasmids.

The 25 transformants that carried both plasmids after transformation with 100 ng DNA were analyzed regarding their plasmid distribution. The plasmids were not distributed equally. There was a clear trend that strains with low eGFP expression exhibited high dTomato expression and vice versa (Figure 4.23). Transformants that contain two enzyme variants could

contain fewer copies of the improved variant and more copies of a neutral or negative variant. So positive clones could be missed during screening.

Even with 1 ng up to 1000 transformants could be obtained. Therefore, DNA amounts as low as possible should be chosen to assure that in most cases only one variant is present. If more transformants are needed, they can still be generated by performing multiple transformations with low DNA amounts.

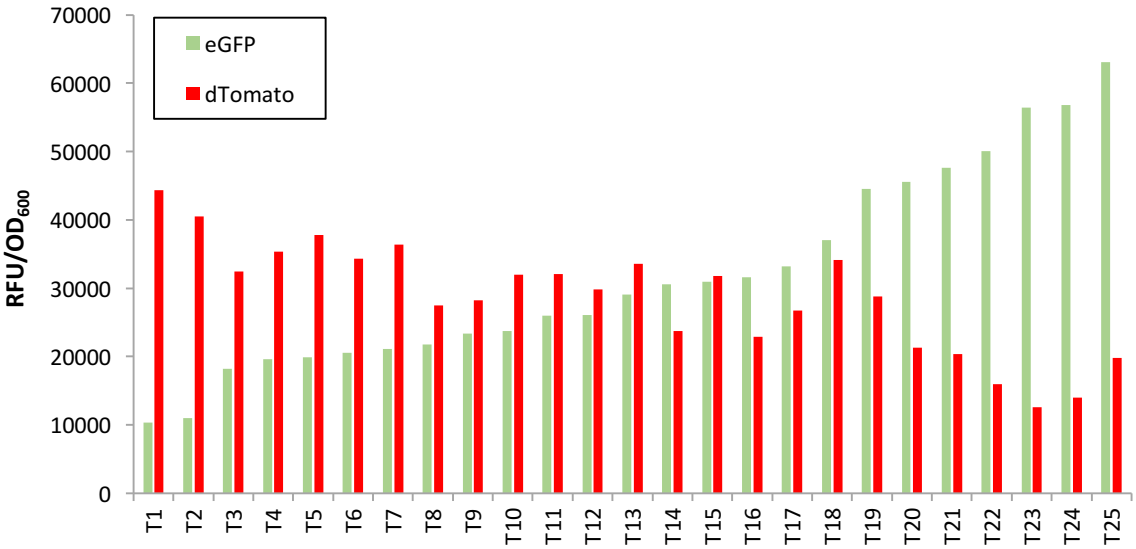

Figure 4.23 Expression levels of the two different reporter genes (eGFP and dTomato) after uptake of multiple plasmids per cell. P_{GAP} driven expression values after 48 h of cultivation in YPD-Zeo from 25 individual transformants (T1 – T25) carrying both plasmid variants are shown.

Multiple plasmids entering a single cell could also have beneficial effects. It enables fast and easy screening of large libraries of helper proteins, such as chaperons, or transcription factors (*e.g. HAC1*), which might improve the expression of the gene of interest.

Therefore, an expression strain containing the desired expression cassette simply has to be transformed with a mixture of plasmids containing different helper proteins or transcription factors. High amounts of DNA should be used to ensure that most of the transformants contain more than one type of plasmid.

After screening for improved expression strains the plasmids can be isolated from these improved strains and analyzed. Suitable candidates showing additive or even synergistic effects can be identified.

Many different random combinations and also different expression strengths can be tested in a short amount of time using this method. Once suitable genes for co-expression are

identified, the expression levels can be fine-tuned using different bidirectional promoters and stable expression strains can be generated.

4.6.5 Plasmid stability

Another important aspect of ARS plasmids is their stability. The plasmids containing eGFP under the control of P_{CAT1} (Zeocin resistance and *GUT1* vectors) were tested using different conditions. Strains carrying the expression cassette integrated into the genome and strains carrying the plasmid for episomal expression (reported in section 4.6.2) were cultivated.
Pre-cultures of the four different strains (Zeocin resistance and *GUT1* based vectors integrated and episomal) with selection pressure were made. Two main cultures for each strain, one with and one without selection pressure, were started with an OD_{600} of 0.2. Every 24 h samples were taken and the eGFP expression as well as the CFU/OD_{600}/ml were measured after plating on selective media (Figure 4.24).
Already after 24 h strains carrying the expression vector episomally yielded fewer colonies than their integrative counterpart. The plasmids seemed to get lost during cell division indicating seggregational instability. After the stationary phase was reached the number of cells carrying the plasmid stayed constant, even after 7 days of cultivation no further plasmid loss was observed.
About one third of the cells (CFU/OD_{600}/ml of the episomal expression strains compared to the genomic expression strains) kept the plasmid when cultivated with Zeocin and one quarter when cultivated without Zeocin.
Interestingly also under non-selective conditions the plasmid seemed to be quite stable. A larger fraction lost the plasmid in the beginning compared to the cultivation with selection pressure, but afterwards the number of cells harboring the plasmid also stayed constant.

Also the eGFP fluorescence was steadily increasing until the end of the cultivation and was higher for the episomal expression strains compared to the genomic expression strains, even under non-selective conditions. It remains to be shown if this is an amplification effect of the expression cassette, or the absorption properties change with cultivation time or any other reason, such as long maturation time of eGFP. Although episomal expression strains showed

elevated eGFP fluorescence with and without selection pressure, the level was even higher under selective conditions.

Interestingly also transformants with integrated expression cassettes showed only 50% CFU/OD_{600} after 24 h compared to longer cultivation times.

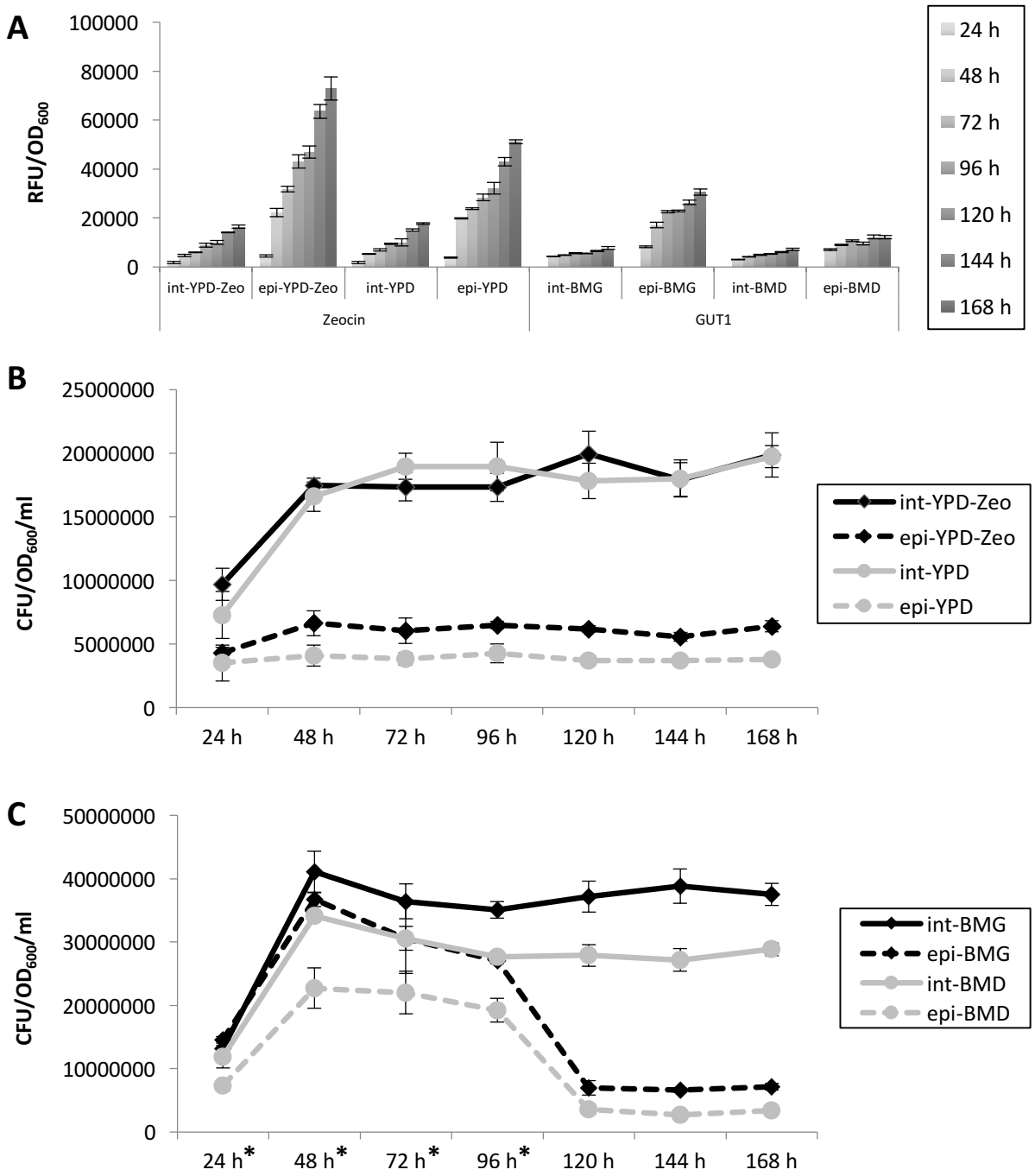

Figure 4.24 Plasmid stability of ARS vectors containing *CbARS*. **A**, eGFP expression values; **B**, Number of clones that can still grow on selection plates (carrying the Zeocin resistance plasmid) normalized to the OD; **C**, Number of clones that can still grow on selection plates (carrying the *GUT1* plasmid) normalized to the OD. Data of technical triplicates is shown. * Background colonies were also counted for episomal *GUT1* expression constructs. The CFU for the episomal *GUT1* vectors are not accurate from 24 h to 96 h.

In an alternative strategy cells could also be plated onto agar plates with non-selective media to normalize colony numbers instead of OD_{600} values. Alternatively, real cell counts could be used. This also means that no clear conclusion about the fraction of cells still containing the plasmid can be made at this stage without such additional experiments.

The results for the *GUT1* vectors were not clear since BMG and BMD cultivation also led to other CFU/OD_{600}/ml values for the integrated vectors. This also indicated that OD_{600} values allow no direct conclusion about cell numbers (possible CFUs). Also background colonies without the plasmid were counted for the first 4 days, so the number of CFU/OD_{600}/ml for the episomal *GUT1* expression strains was not accurate until the 5^{th} day (120 h). Nevertheless, the number of cells carrying the plasmid stayed constant over the last three days.

To obtain more reliable results the strains carrying episomal expression vectors should be cultivated in biological replicates and plated onto selective and non-selective media to calculate the fraction of cells without the plasmid. Shorter times of cultivation should be sufficient, since plasmid loss did not occur during the last days of cultivation.
Alternatively, plasmid loss should be calculated from cultures in bioreactors with constant growth rate (either fed batch or chemostat).

4.6.6 Direct *Pichia* transformation

The high transformation efficiency suggested that direct transformation into *P. pastoris* could be possible avoiding the cloning step in *E. coli* and thus saving time.
Direct transformation was evaluated using the chromogenic protein TinselPurple (DNA2.0) and *Sap*I cloning into a Zeocin based expression vector with P_{GAP} in order to obtain constitutive expression and colored colonies on the agar plate. The whole de-salted cloning reaction containing 10 ng vector backbone and 120 ng insert was used for the transformation.
The transformation yielded approximately 1500 transformants (Figure 4.25), most of them colored, *i.e.* containing the insert. Only <10% did not contain the correct insert.

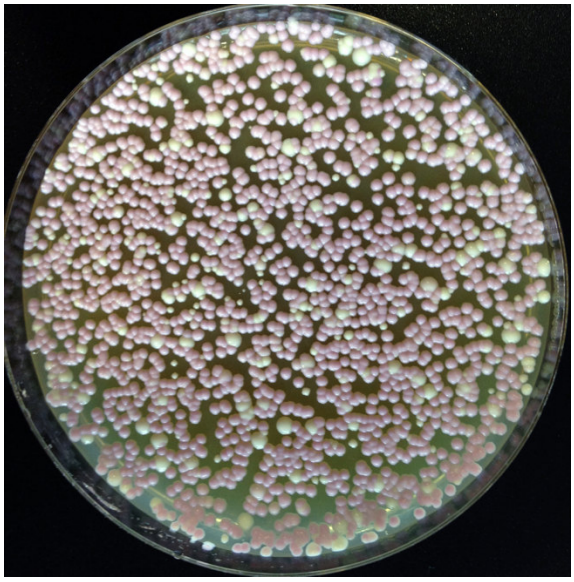

Figure 4.25 Direct *P. pastoris* transformation with a *Sap*I cloning reaction mix containing *TinselPurple* as insert and a Zeocin resistance based ARS vector with P_{GAP}. The whole de-salted cloning reaction containing 10 ng vector backbone and 120 ng insert was used for the transformation. Purple colonies contain the vector with the insert and the white colonies contain the religated vector backbone.

Direct *Pichia* transformation could be used to save time, but also negative clones could be picked and therefore more colonies should be screened.

The transformation was also tried with the vector containing no *Not*I site, but in contrast to *E. coli* (Figure 4.10) there was no effect on the transformation efficiency (data not shown).

White colonies without the insert could be due to genomic integration of the vector backbone or more likely due to relegation of the vector backbone. Genomic integration is unlikely, because only 10 ng plasmid are used for cloning. Also the white colonies seemed to have the same size and shape as the pink colonies indicating presence of the vector episomally. The use of more insert could further decrease the fraction of negative clones, *i.e.* transformants carrying the religated vector backbone, even though it was already below 10 %.

Additionally, it would be interesting to see if *in vivo* recombination in *Pichia* could be applied to get the insert into the vector, or if this is too inefficient due to the relegation of the plasmid by NHEJ repair.

4.6.7 *Pichia* plasmid isolation and re-transformation

Another important part in enzyme engineering is the recovery of improved variants. When integrated into the genome the procedure to recover the improved gene is laborious since it involves gDNA extraction, PCR amplification and re-cloning.

ARS plasmids can be easily recovered by isolating them directly from an overnight culture with Zymolyase digestion followed by application of a plasmid isolation kit.

This was made with the same vector described in section 4.6.6 and 20 µl or 5 µl of the miniprep were used for re-transformation of *P. pastoris* and *E. coli* respectively.

E. coli transformation yielded approximately 10^3 to 10^4 transformants (not shown) and approximately 300 to 400 transformants could be obtained for *P. pastoris* (Figure 4.26).

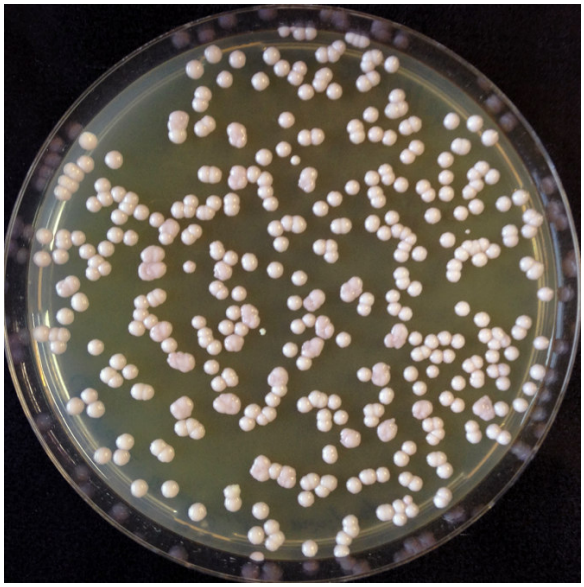

Figure 4.26 *P. pastoris* transformants after plasmid isolation from *P. pastoris* and re-transformation using a *TinselPurple* expression vector. One plasmid isolation batch (20 µl) from 5 ml ONC was used for the transformation.

All of the transformants showed the pink color, thus carried the correct plasmid. The transformation step into *E. coli* allows recovery of the plasmids without any gDNA extraction or PCR and therefore storage of the plasmid and the possibility to obtain it in sufficient amounts for further work.

Also the re-transformation into *P. pastoris* can be useful allowing one plasmid to be tested in different strains without laborious methods needed. Many different expression strains can be tested in a short amount of time.

4.7 Expression of secretory proteins

Secretory expression with ARS vectors can be a problem because too strong transcription and translation can overwhelm the secretory pathway and leads to lower final expression of active, secreted protein.

Since different ARSs with different expression strengths were identified during this thesis, the weaker ones might be beneficial to drive expression of proteins that are hard to secrete like *CalB*. The strong ARSs should be suitable to express easy to secrete proteins like *TrCBH2* where high protein titers in the media can be reached (6).

CalB and *TrCBH2* were expressed using *CAT1-ARS* and *CbARS* using the native signal sequence for *TrCBH2* and the alpha factor signal sequence with the deleted alpha helix for *CalB* (65).

Expression was driven by P_{CAT1} and the derepressed values after 60 h are shown for *CalB* (Figure 4.27) and *TrCBH2* (Figure 4.28). *CalB* expression was also measured after 48 h of induction (Figure 7.1).

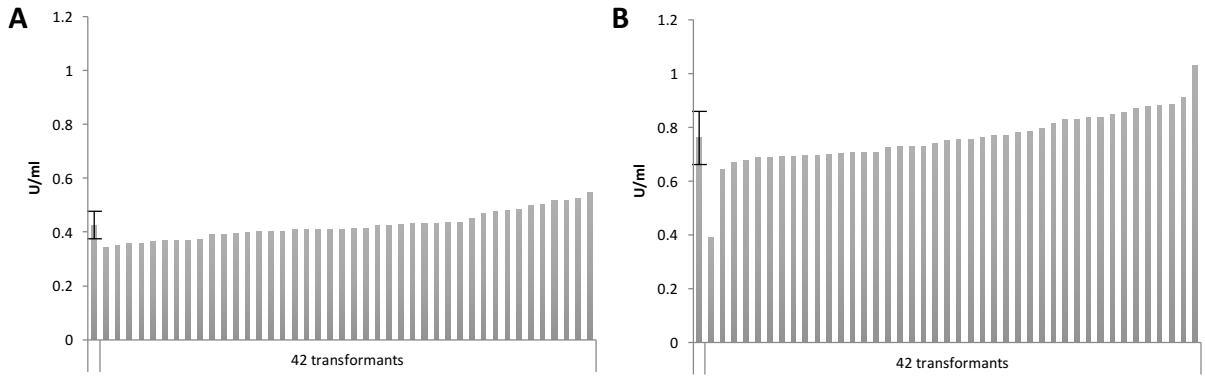

Figure 4.27 *CalB* expression with different ARSs. **A**, *CAT1-ARS*; **B**, *CbARS*. P_{CAT1} driven expression values after 60 h of cultivation in selective media from 42 transformants are shown. The first bar on the left indicates the mean value and the error bar represents the standard deviation.

Even though lower enzyme activity in the culture supernatant was expected for the strong *CbARS* it yielded more active *CalB* than the *CAT1-ARS*. But still the difference between the two ARSs was just approximately 1.6-fold, which is rather low compared to the difference using eGFP as a reporter protein (approximately 3.4-fold).

The difference was clearer for *TrCBH2* expression. The pHBAH assay did not work well and therefore the *TrCBH2* levels were determined with SDS-PAGE (Figure 4.28).

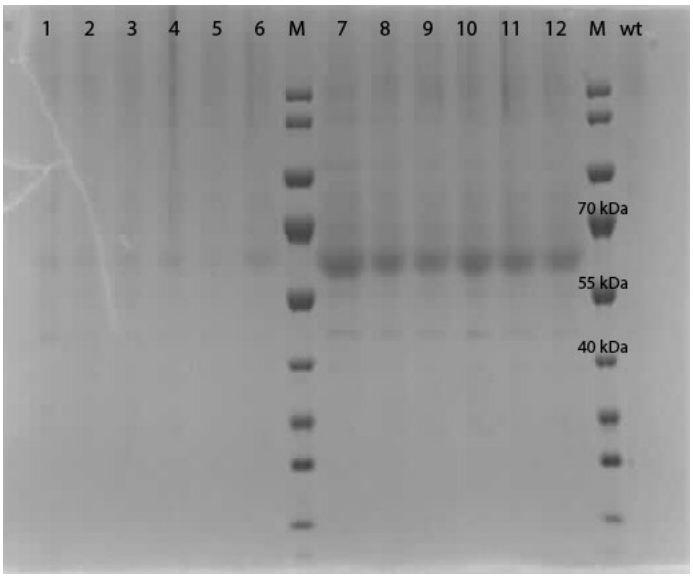

Figure 4.28 P_{CAT1} driven *TrCBH2* expression with *CAT1-ARS* (1-6) and *CbARS* (7-12) containing vectors. The strains were cultivated for 60 h in selective media and 200 µl of the culture supernatant were used for analysis. M, marker; wt, wildtype. 10 µl of EndoH digested methanol/chloroform precipitated supernatants were loaded in each slot. Proteins were stained with Coomassie and the image taken using the G:BOX system.

*Tr*CBH2p carries multiple glycosylation sites, O- and N-linked (69). Since the performed EndoH digestion just removes the N-linked glycosyl moieties most of the protein still ran at a higher molecular weight. *TrCBH2* should run at 49.6 kDa.

The bands with a molecular weight of slightly above 40 kDa, which are very weak for the *CbARS* expression clones (7-12) might correspond to enzyme without the cellulose binding domain. The majority of the protein ran at approximately 60 kDA were clear bands were visible for *CbARS* and just very faint bands for *CAT1-ARS*.

This confirmed that easily secreted proteins could be produced in much higher titers with the stronger ARSs using their full potential. While the expression of other more difficult to secrete proteins could not profit as much from the higher copy number and would possibly need further helper proteins to assure proper folding and secretion. Nevertheless, we were surprised to see any improvement at all for such proteins.

4.8 Generation of *glycerol kinase* (*GUT1*) knockout strains

GUT1 knockout strains were generated in *P. pastoris* BSYBG10 and BSYBG11 for the use as host strains for the *GUT1* based expression plasmids.
After transformation of the knockout cassette, cultivation of the transformants in DWPs and pinning them onto BMD and BMG plates (Figure 4.29) the phenotype was checked.
Some clones per plate (approximately 5%) showed clearly abolished growth on glycerol, while they grew normally on BMD (not shown).

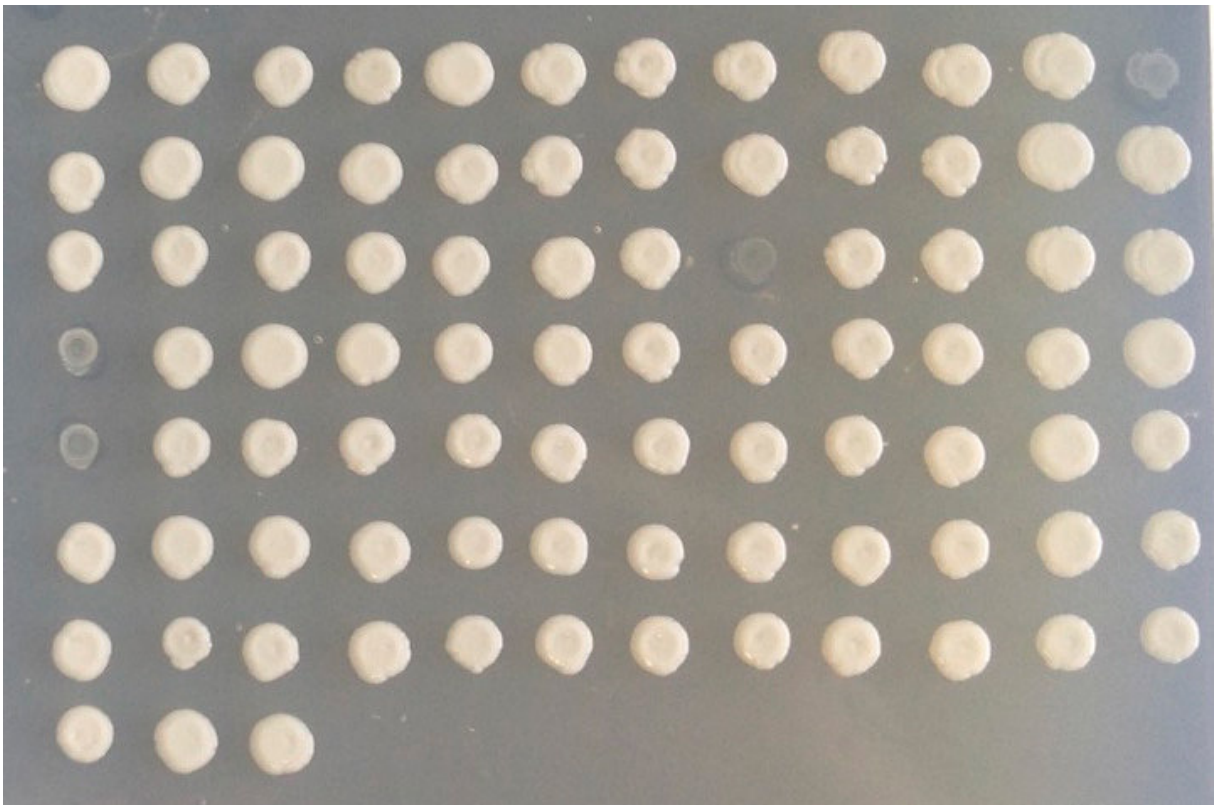

Figure 4.29 Transformants after transformation with the *GUT1* knockout cassette. They were pinned onto BMG plates after 48 h of cultivation in BMD.

The positive clones, which showed the desired phenotype were further analyzed using PCR with primers outside of the target locus and observing the length of the PCR product, which was longer when the cassette was integrated correctly (not shown).

All clones with the correct phenotype also showed the expected PCR bands and were cultivated again in DWPs and induced as well as pinned and re-streaked on BMM2 plates to facilitate the FLP recombinase expression and to excise the knockout cassette.

Strains without the cassette should not be able to grow on Zeocin containing media anymore and therefore the excision was checked by re-streaking single colonies on YPD and YPD-Zeo plates (Figure 4.30).

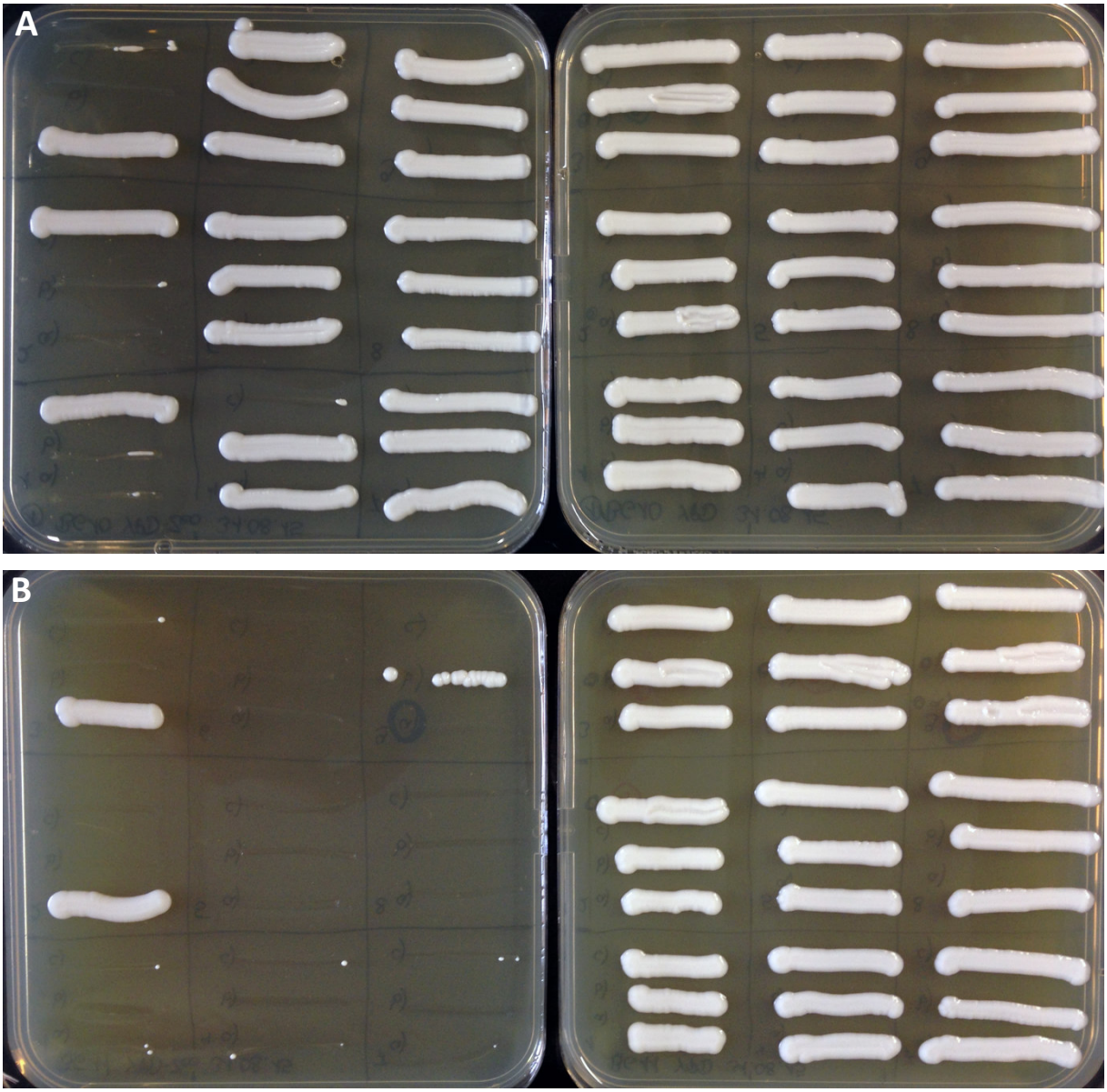

Figure 4.30 Clones re-streaked on YPD-Zeo (left) and YPD (right) plates after induction to screen for correctly excised knockout cassettes. **A**, BSYBG10 ΔGUT1; **B**, BSYBG11 ΔGUT1.

Many of the *P. pastoris* BSYBG10 *ΔGUT1* strains still carried the knockout cassette even after several rounds of induction, while the *P. pastoris* BSYBG11 *ΔGUT1* (Mut^S) strains were mostly without Zeocin resistance. Flipping the cassette out of the genome worked much better with the Mut^S strain, probably because of the more efficient induction.

5 Conclusion and Outlook

5.1 Conclusion

In this study an ARS was found that provides some beneficial effects compared to the commonly used *PARS1*. It enables stronger expression especially when used as terminator and origin of replication simultaneously, which helps to minimize the plasmid size.
Vectors carrying the *CbARS* also displayed a higher transformation rate and are therefore more suitable for library screening where the loss of diversity should be kept minimal.

Not only the high expression and transformation rate are valuable tools in enzyme engineering, also the very uniform expression levels among different clones is very important. Uniform expression minimizes the screening effort that has to be taken and also minimizes the chance to get false positives or to miss positive clones.
The high transformation efficiency also gives the chance to avoid the cloning step in *E. coli* by directly transforming *P. pastoris* with the cloning mix. Also plasmid recovery from yeast is easy and can be performed very quickly in sufficient amounts for the use as PCR template, to transform *E. coli* or re-transform *P. pastoris*.
It was also proven that in most of the cases only one plasmid variant is present in one transformant when low amounts of DNA are transformed.
These facts make ARS plasmids the ideal candidates for enzyme engineering or promoter engineering. Many different variants can be reliably screened within a short amount of time and also in different strains.
In general, the ARS plasmids generated in that study are easier to handle than standard expression vectors for genomic integration. Only very low amounts of DNA have to be prepared and no linearization, gel purification or other modification is required.

With that new ARS a set of plasmids was designed for state of the art seamless cloning with an IP-free method using the type IIS restriction enzyme *Sap*I. The set of plasmids developed in this study does not only offer a range of different promoters and therefore also different expression profiles and strengths, but also different selection systems.

One based on the use of the antibiotic Zeocin and the other one on auxotrophic selection to avoid the need for antibiotics in the cultivation media.

ARS plasmids based on the Zeocin resistance are in principle not suitable for production processes in larger scale the, since the selection pressure, *i.e.* Zeocin, has to be present all the time limiting the application to small scale cultures. The main application for these plasmids should be the generation and screening of libraries, because they provide very homogenous and reliable expression. Once an improved variant is found it can be either integrated into the genome using the same vector or transferred to other vector systems like the *GUT1* vectors. Nevertheless, relatively high expression with *CbARS* based plasmids even without selection pressure might allow scale up for some cases.

The *GUT1* based plasmids might be especially suitable for production processes in larger scale, since the selection pressure can be maintained without the addition of antibiotics. Only glycerol has to be present as sole carbon source, which can be easily achieved using a continuous glycerol feed in the bioreactor. As expected, growth seems to be necessary to maintain the plasmid.

5.2 Outlook

There are several steps that could be done to further optimize the episomal expression system. Different promoters for the expression of the selection marker could be tested. That might give the possibility to control the plasmid copy number and therefore also the expression strength. Since it was shown that different Zeocin concentrations had an effect on the expression levels of the reporter protein, different expression levels of the marker gene should also have a similar effect. Using weak promoters to drive the expression of the marker gene should maximize the plasmid copy number while strong promoters respectively should weaken the expression of the gene of interest by minimizing the plasmid copy number.

Not only the promoter of the selection marker, but also the selection system itself could have an effect on the expression strength. Therefore, additional knockout strains providing the use

of additional marker genes should be tested to give alternatives to the *GUT1* system for antibiotic free cultivation. One example could be the use of isocitrate lyase knockout as demonstrated before for *S. cerevisiae*. Also dominant markers allowing antibiotic free selection could be tried.

Also *POX1* or *PEX13* could be very interesting options, because selection pressure can be kept up using oleate as the sole carbon source. Similar to the *GUT1* system a continuous feed of oleate can provide selection pressure throughout the whole cultivation. These two genes are particularly interesting, because oleate also induces P_{CAT1} (Vogl T, Dissertation), which could lead to very high expression levels.

There are also many other possibilities to provide auxotrophic selection. The different systems would have to be tested regarding the background, which can occur due to the fact that knockout strains could still grow, although they show abolished growth, or because of cross feeding. Problems with cross feeding were reported for sucrose selection using *SUC2* as marker. The resulting loss of selection pressure also led to a higher plasmid loss (53).

The stability in general is a property, which should be investigated further. To make it more comparable the method described in literature (16,23) should be used. The strains should be cultivated in biological replicates and plated onto selective and non-selective media. Cultivation should also be performed for a longer time inside the exponential growth phase using an appropriate feeding strategy.

The *Sap*I based cloning system used in this study already led to a large number of colonies when the mix was transformed directly into *P. pastoris*. Since only 10 ng of vector backbone were currently used for the cloning, that number could be increased even further.

If the efficiency is high enough single genes or a pool of mutated genes could be cloned into the vector using *Sap*I cloning and the cloning step in *E. coli* completely avoided.

This leads to the possibility to build a vector without any parts for *E. coli*. The replication origin could be placed inside the stuffer fragment, in order to provide sufficient amounts of vector backbone for cloning. The *E. coli* origin of replication would be excised prior cloning since the stuffer is removed. That would make the plasmids even smaller and thus could help to improve the transformation efficiency. Although the transformation efficiency for *P. pastoris* could be increased, the plasmids could not be re-transformed in *E. coli* anymore, but isolation from *Pichia*, re-transformation and PCR amplification from yeast minipreps should be simple.

6 Literature

1. Kurtzman C. Biotechnological strains of Komagataella (Pichia) pastoris are Komagataella phaffii as determined from multigene sequence analysis. J Ind Microbiol Biotechnol [Internet]. Springer-Verlag; 2009;36(11):1435–8. Available from: http://dx.doi.org/10.1007/s10295-009-0638-4

2. Cregg JM, Cereghino JL, Shi J, Higgins DR. Recombinant Protein Expression in Pichia pastoris. Mol Biotechnol. 2000;16(1):23–52.

3. Gellissen G. Heterologous protein production in methylotrophic yeasts. Appl Microbiol Biotechnol. 2000;54(6):741–50.

4. Heyland J, Fu J, Blank LM, Schmid A. Quantitative physiology of Pichia pastoris during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. Biotechnol Bioeng. 2010;107(2):357–68.

5. Hasslacher M, Schall M, Hayn M, Bona R, Rumbold K, Lückl J, et al. High-level intracellular expression of hydroxynitrile lyase from the tropical rubber tree Hevea brasiliensis in microbial hosts. Protein Expr Purif. 1997;11(1):61–71.

6. Mellitzer A, Ruth C, Gustafsson C, Welch M, Birner-Grünberger R, Weis R, et al. Synergistic modular promoter and gene optimization to push cellulase secretion by Pichia pastoris beyond existing benchmarks. J Biotechnol [Internet]. Elsevier B.V.; 2014;191:187–95. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0168165614008402

7. Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA. Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. Mol Cell Biol [Internet]. 1989;9(3):1316–23. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=362724&tool=pmcentrez&rendertype=abstract

8. Ellis SB, Brust PF, Koutz PJ, Waters AF, Harpold MM, Gingeras TR. Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast Pichia pastoris. Mol Cell Biol

[Internet]. 1985;5(5):1111–21. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=366829&tool=pmcentrez&rendertype=abstract

9. Bill RM. Playing catch-up with Escherichia coli: using yeast to increase success rates in recombinant protein production experiments. Front Microbiol [Internet]. 2014 Jan [cited 2014 Sep 3];5(March):85. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3942658&tool=pmcentrez&rendertype=abstract

10. Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, et al. High-quality genome sequence of Pichia pastoris CBS7435. J Biotechnol [Internet]. 2011;154(4):312–20. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0168165611002306

11. De Schutter K, Lin Y-C, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, et al. Genome sequence of the recombinant protein production host Pichia pastoris. Nat Biotechnol. 2009;27(6):561–6.

12. Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, et al. Genome, secretome and glucose transport highlight unique features of the protein production host Pichia pastoris. Microb Cell Fact [Internet]. 2009;8(1):29. Available from: http://www.microbialcellfactories.com/content/8/1/29

13. Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, et al. Promoter library designed for fine-tuned gene expression in Pichia pastoris. Nucleic Acids Res [Internet]. 2008 Jul [cited 2014 Jul 22];36(12):e76. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2475614&tool=pmcentrez&rendertype=abstract

14. Geier M, Fauland P, Vogl T, Glieder A. Compact multi-enzyme pathways in P. pastoris. Chem Commun [Internet]. Royal Society of Chemistry; 2015;51(9):1643–6. Available from: http://xlink.rsc.org/?DOI=C4CC08502G

15. Lee CC, Williams TG, Wong DWS, Robertson GH. An episomal expression vector for screening

mutant gene libraries in Pichia pastoris. Plasmid. 2005;54(1):80–5.

16. Hong IP, Anderson S, Choi SG. Evaluation of a new episomal vector based on the GAP promoter for structural genomics in pichia pastoris. J Microbiol Biotechnol. 2006;16(9):1362–8.

17. Liu Z, Pscheidt B, Avi M, Gaisberger R, Hartner FS, Schuster C, et al. Laboratory evolved biocatalysts for stereoselective syntheses of substituted benzaldehyde cyanohydrins. ChemBioChem. 2008;9(1):58–61.

18. Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A. Deletion of the pichia pastoris ku70 homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One. 2012;7(6).

19. Kenzom T, Srivastava P, Mishra S. Simplified high-throughput screening of AOX1-expressed laccase enzyme in Pichia pastoris. Anal Biochem [Internet]. Elsevier Inc; 2015;489:59–61. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0003269715003929

20. Zhang S-B, Pei X-Q, Wu Z-L. Multiple amino acid substitutions significantly improve the thermostability of feruloyl esterase A from Aspergillus niger. Bioresour Technol [Internet]. Elsevier Ltd; 2012;117:140–7. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0960852412006475

21. Gonzalez-perez D, Garcia-ruiz E, Alcalde M. Saccharomyces cerevisiae in directed evolution An efficient tool to improve enzymes © 2012 Landes Bioscience . Do not distribute . © 2012 Landes Bioscience . Do not distribute . 2012;3(3):172–7.

22. Krondorfer I, Lipp K, Brugger D, Staudigl P, Sygmund C, Haltrich D, et al. Engineering of pyranose dehydrogenase for increased oxygen reactivity. PLoS One. 2014;9(3):1–9.

23. Cregg JM, Barringer KJ, Hessler a Y, Madden KR. Pichia pastoris as a host system for transformations. Mol Cell Biol. 1985;5(12):3376–85.

24. Lueking A, Holz C, Gotthold C, Lehrach H, Cahill D. A System for Dual Protein Expression in Pichia pastoris and Escherichia coli. Protein Expr Purif [Internet]. 2000;20(3):372–8. Available

from: http://linkinghub.elsevier.com/retrieve/pii/S1046592800913172

25. Lueking A, Horn S, Lehrach H, Cahill DJ. A dual-expression vector allowing expression in E. coli and P. pastoris, including new modifications. Methods Mol Biol [Internet]. 2003;205(11):31–42. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12491878

26. Sandström AG, Engström K, Nyhlén J, Kasrayan A, Bäckvall JE. Directed evolution of Candida antarctica lipase A using an episomaly replicating yeast plasmid. Protein Eng Des Sel. 2009;22(7):413–20.

27. Uchima CA, Arioka M. Expression and One-Step Purification of Recombinant Proteins Using an Alternative Episomal Vector for the Expression of N-Tagged Heterologous Proteins in Pichia pastoris. Biosci Biotechnol Biochem. 2012;76(2):368–71.

28. Liachko I, Dunham MJ. An autonomously replicating sequence for use in a wide range of budding yeasts. FEMS Yeast Res [Internet]. 2014;14(2):364–7. Available from: http://femsyr.oxfordjournals.org/cgi/doi/10.1111/1567-1364.12123

29. Kozak M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene [Internet]. 2005;361:13–37. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0378111905004348

30. Crook NC, Freeman ES, Alper HS. Re-engineering multicloning sites for function and convenience. Nucleic Acids Res. 2011;39(14).

31. Vogl T, Ahmad M, Krainer FW, Schwab H, Glieder A. Restriction site free cloning (RSFC) plasmid family for seamless, sequence independent cloning in Pichia pastoris. Microb Cell Fact [Internet]. BioMed Central; 2015;14(1):103. Available from: http://www.microbialcellfactories.com/content/14/1/103

32. Bryksin A V., Matsumura I. Overlap extension PCR cloning: A simple and reliable way to create recombinant plasmids. Biotechniques. 2010;48(6):463–5.

33. Zuo P, Rabie BM. One-step DNA fragment assembly and circularization for gene cloning. Curr Issues Mol Biol. 2010;12(1):11–6.

34. Quan J, Tian J. Circular Polymerase Extension Cloning. Methods Mol Biol [Internet]. 2014;1116(7):103–17. Available from: http://link.springer.com/10.1007/978-1-62703-764-8\nhttp://link.springer.com/10.1007/978-1-62703-764-8_8

35. Bubeck P, Winkler M, Bautsch W. Rapid cloning by homologous recombination in vivo. Nucleic Acids Res. 1993;21(15):3601–2.

36. Oiiner JD, Kinzler KW, Vogelstein B. PCR products. 1993;21(22):5192–7.

37. Lee J, Rha E, Yeom SJ, Lee DH, Choi ES, Lee SG. Generating in vivo cloning vectors for parallel cloning of large gene clusters by homologous recombination. PLoS One. 2013;8(11):1–8.

38. Parrish JR, Limjindaporn T, Hines J a., Liu J, Liu G, Finley RL. High-throughput cloning of Campylobacter jejuni ORFs by in vivo recombination in Escherichia coli. J Proteome Res. 2004;3(3):582–6.

39. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Meth [Internet]. Nature Publishing Group; 2009 May;6(5):343–5. Available from: http://dx.doi.org/10.1038/nmeth.1318

40. Bath AJ, Milsom SE, Gormley NA, Halford SE. Many Type IIs Restriction Endonucleases Interact with Two Recognition Sites before Cleaving DNA. J Biol Chem [Internet]. 2002;277(6):4024–33. Available from: http://www.jbc.org/cgi/doi/10.1074/jbc.M108441200

41. Benko Z, Zhao R. Zeocin for selection of bleMX6 resistance in fission yeast. Biotechniques [Internet]. 2011;51(1):1–4. Available from: http://www.biotechniques.com/article/000113706

42. Drocourt D, Calmels T, Reynes JP, Baron M, Tiraby G. Cassettes of the Streptoalloteichus hindustanus ble gene for transformation of lower and higher eukaryotes to phleomycin resistance. Nucleic Acids Res. 1990;18(13):4009.

43. Kimura M, Kamakura T, Zhou Tao Q, Kaneko I, Yamaguchi I. Cloning of the blasticidin S deaminase gene (BSD) from Aspergillus terreus and its use as a selectable marker for Schizosaccharomyces pombe and Pyricularia oryzae. Mol Gen Genet MGG [Internet]. Springer-Verlag; 1994;242(2):121–9. Available from: http://dx.doi.org/10.1007/BF00391004

44. Chen M-T, Lin S, Shandil I, Andrews D, Stadheim TA, Choi B-K. Generation of diploid Pichia pastoris strains by mating and their application for recombinant protein production. Microb Cell Fact [Internet]. Microbial Cell Factories; 2012;11(1):91. Available from: http://www.microbialcellfactories.com/content/11/1/91

45. Lin-Cereghino J, Hashimoto MD, Moy A, Castelo J, Orazem CC, Kuo P, et al. Direct selection of Pichia pastoris expression strains using new G418 resistance vectors. Yeast [Internet]. 2008;25(4):293–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2504081&tool=pmcentrez&rendertype=abstract

46. Wan H, Sjölinder M, Schairer HU, Leclerque A. A new dominant selection marker for transformation of Pichia pastoris to soraphen A resistance. J Microbiol Methods [Internet]. 2004;57(1):33–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15003686\nfile:///C:/Users/uizcf/AppData/Local/Mendeley Ltd./Mendeley Desktop/Downloaded/Wan et al. - 2004 - A new dominant selection marker for transformation of Pichia pastoris to soraphen A resistance.pdf

47. Yang J, Nie L, Chen B, Liu Y, Kong Y, Wang H, et al. Hygromycin-resistance vectors for gene expression in Pichia pastoris. Yeast [Internet]. 2014 Apr 1;31(4):115–25. Available from: http://dx.doi.org/10.1002/yea.3001

48. Lin-Cereghino J, Lin-Cereghino GP. Vectors and strains for expression. Methods Mol Biol. 2007;389(4):11–26.

49. Du M, Battles MB, Nett JH. A color-based stable multi-copy integrant selection system for Pichia pastoris using the attenuated ADE1 and ADE2 genes as auxotrophic markers. Bioeng Bugs [Internet]. 2012;3(1):32–7. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/22126802

50. Sunga AJ, Cregg JM. The Pichia pastoris formaldehyde dehydrogenase gene (FLD1) as a marker for selection of multicopy expression strains of P. pastoris. Gene [Internet]. 2004;330:39–47. Available from: http://linkinghub.elsevier.com/retrieve/pii/S037811190301151X

51. Koller a, Spong a P, Lüers GH, Subramani S. Analysis of the peroxisomal acyl-CoA oxidase gene product from *Pichia pastoris* and determination of its targeting signal. Yeast [Internet]. 1999;15(11):1035–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10455228

52. Sacksteder K a., Jones JM, South ST, Li X, Liu Y, Gould SJ. PEX19 binds multiple peroxisomal membrane proteins, is predominantly cytoplasmic, and is required for peroxisome membrane synthesis. J Cell Biol. 2000;148(5):931–44.

53. Sreekrishna K, Tschopp JF, Fuke M. Invertase gene (SUC2) of Saccharomyces cerevisiae as a dominant marker for transformation of Pichia pastoris. Gene. 1987;59(1):115–25.

54. Read JD, Colussi P a., Ganatra MB, Taron CH. Acetamide selection of Kluyveromyces lactis cells transformed with an integrative vector leads to high-frequency formation of multicopy strains. Appl Environ Microbiol. 2007;73(16):5088–96.

55. Krainer FW, Dietzsch C, Hajek T, Herwig C, Spadiut O, Glieder A. Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway. Microb Cell Fact [Internet]. BioMed Central Ltd; 2012;11(1):22. Available from: http://www.microbialcellfactories.com/content/11/1/22

56. Harju S, Fedosyuk H, Peterson KR. Rapid isolation of yeast genomic DNA: Bust n' Grab. BMC Biotechnol. 2004;4:8.

57. Singh M V., Weil PA. A method for plasmid purification directly from yeast. Anal Biochem. 2002;307(1):13–7.

58. Lin-Cereghino J, Wong WW, Xiong S, Giang W, Luong LT, Vu J, et al. Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris.

Biotechniques. 2005 Jan;38(1):44, 46, 48.

59. Seidman CE, Struhl K, Sheen J, Jessen T. Introduction of Plasmid DNA into Cells. In: Current Protocols in Molecular Biology [Internet]. John Wiley & Sons, Inc.; 2001. Available from: http://dx.doi.org/10.1002/0471142727.mb0108s37

60. Mellitzer A, Glieder A, Weis R, Reisinger C, Flicker K. Sensitive high-throughput screening for the detection of reducing sugars. Biotechnol J. 2012;7(1):155–62.

61. WEIS R, LUITEN R, SKRANC W, SCHWAB H, WUBBOLTS M, GLIEDER A. Reliable high-throughput screening with by limiting yeast cell death phenomena. FEMS Yeast Res [Internet]. 2004;5(2):179–89. Available from: http://femsyr.oxfordjournals.org/cgi/doi/10.1016/j.femsyr.2004.06.016

62. Abad S, Nahalka J, Bergler G, Arnold SA, Speight R, Fotheringham I, et al. Stepwise engineering of a Pichia pastoris D-amino acid oxidase whole cell catalyst. Microb Cell Fact. 2010;9:24.

63. Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S. Structural and functional features of eukaryotic mRNA untranslated regions. Gene. 2001;276(1-2):73–81.

64. Vogl T, Ruth C, Pitzer J, Kickenweiz T, Glieder A. Synthetic core promoters for Pichia pastoris. ACS Synth Biol [Internet]. 2014 Mar 21;3(3):188–91. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3964828&tool=pmcentrez&rendertype=abstract

65. Lin-Cereghino GP, Stark CM, Kim D, Chang J, Shaheen N, Poerwanto H, et al. The effect of α-mating factor secretion signal mutations on recombinant protein expression in Pichia pastoris. Gene [Internet]. Elsevier B.V.; 2013;519(2):311–7. Available from: http://dx.doi.org/10.1016/j.gene.2013.01.062

66. Chen S, Reger R, Miller C, Hyman LE. Transcriptional terminators of RNA polymerase II are associated with yeast replication origins. Nucleic Acids Res. 1996;24(15):2885–93.

67. Liachko I, Youngblood R a., Tsui K, Bubb KL, Queitsch C, Raghuraman MK, et al. GC-Rich DNA

Elements Enable Replication Origin Activity in the Methylotrophic Yeast Pichia pastoris. PLoS Genet. 2014;10(3).

68. Qin X, Qian J, Yao G, Zhuang Y, Zhang S, Chu J. GAP promoter library for fine-tuning of gene expression in Pichia pastoris. Appl Environ Microbiol. 2011;77(11):3600–8.

69. Koivula a, Reinikainen T, Ruohonen L, Valkeajärvi a, Claeyssens M, Teleman O, et al. The active site of Trichoderma reesei cellobiohydrolase II: the role of tyrosine 169. Protein Eng [Internet]. 1996;9(8):691–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8875646

7 Supplementary information

7.1 P_{CbAOD1} sequence alignment

```
pCbAOD1-amplified       GGAGTATACGTAAATATATAATTATATATAATCATATATATGAATACAATGCAATGAAAG
gbM81702.1              GGAGTATACGTAAATATATAATTATATATAATCATATATATGAATA-----CAATGAAAG
                        ******************************************     *********

pCbAOD1-amplified       TGAATATGATAAGATTGAAATAATAACAAACAGCGATAAATATATCTCAAAATGGAGTTA
gbM81702.1              TAAATATGATAAGATTGAAATAATAACAAACAGCGATAAATATATCTCAAAATGGAGTTA
                        * ********************************************************

pCbAOD1-amplified       CACAACAAATAATAATAAAATATAAATTATAAATTATAAATTATAAAAGAATAAAAAATA
gbM81702.1              CACAACAAATAATAATAAAATATAAATTATAAAATATAAA---------GGAATAAAATA
                        ********************************* ******           * ******

pCbAOD1-amplified       AACCCCACTAATTTATTTTATTAAAAGATAGATTGGTATCTTTACTTAATAACAATTCTG
gbM81702.1              AACCCCACTAATTTATTTTATTAAAAGATAGATTGGTATCTTTACTTAATAACAATTCTG
                        ************************************************************

pCbAOD1-amplified       AAACTTTATTCACTTAATTTTATTTAACTTATTTAATTTATTTTTACCCCAGTTT-TTCA
gbM81702.1              AAACTTTATTCACTTAATTTTATTTAACTTATTTAATTTATTTTTTACCCCAGTTTTTTCA
                        ************************************************************ ****

pCbAOD1-amplified       GTACAATGCAGCTCCGAAACTTTATTTGGCTGTGATTTGGCTGTGATTTGGCTGTGATTT
gbM81702.1              GTACAGTGCAGCTCCGAAACTTTATTTGGCT-----------GTGATTTGGCTGTGATTT
                        ***** ************************           ******************

pCbAOD1-amplified       GGCTTGGCTTGGCTGGCTGGAATTGTCTCCTGCAGGAATTGCTCGGGGTCCGGTTCTCCC
gbM81702.1              GGCTTGGCTTGGCTGGCTGGAATTGTCTCCTGCAGGAATTGCTCGGGGTCCGGTTCTCCC
                        ************************************************************

pCbAOD1-amplified       GCTGGCTGGCTATTTGGCGGGCTGGCTATTTGGCGGGCTGGCTGGCTGGCTGCTCTGCCA
gbM81702.1              GCAGCTGGATATTTGGCTGGCT----------------GCTCTGTCTGGCTGCTCTGCCA
                        ** *   *    ** *  **                   *** ***************

pCbAOD1-amplified       TCTGCTGTGGC-CACCCCGCATCTCTGGATGCACGCCGTGCAGCTGGACGTGCGTCTACC
gbM81702.1              TCTGCTGTGGCCACCCCCGCATCTCTGGATGCACGCCGTGCAGCTGGACTTGCGTCTACC
                        ***********   ********************************** **********

pCbAOD1-amplified       CTGCAGCCGTGTGCCTTATTTCCCAATCTCCCAATCTCTCAATCTGCCAGTCAGCCAAAA
gbM81702.1              CTGCAGCCGTGTGCCTCATCTCCCAATCTCTCAATCAGCCAGTCAGCCAGCCAGCCAAAA
                        *************** ** ********** *****   ** ** ***** *********
```

```
pCbAOD1-amplified       CACCGGCCAGGCAG--------------------GCAGGCAGGCAGGCAGGCAGGCAGTG
gbM81702.1              TACGGGCCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGTG
                         ** **********                     **************************

pCbAOD1-amplified       AAGCCTTCCCACGCCCCACTCCGCATAAACATCCCCAGCAGTTTCCCCAGCAGTTTCCCC
gbM81702.1              ATGCCTTCCCACGCCCCACCCCGCATAAACATCCCCAGCAGTTTCCCCAGCAGTTTCCCC
                        * ***************** ****************************************

pCbAOD1-amplified       AGCTTTTCAATTTAATAAAATAGCCTGTTTCTGTTTCTGTTTTATATTATACAATTTTTT
gbM81702.1              AGCTTTTCAATTTAATAAAATAGCCTGTTTCTGTTTCTGTTTTATATTATACAATTTTTT
                        ************************************************************

pCbAOD1-amplified       ATCCTAATAATTACTCTTTCGGGAATTAAATAATAATTATATCATATACCCATATCACAT
gbM81702.1              ATCCTAATAATTACTCTTTTGGGAATTAAATAATAATTATATCATATACCCATATCACAT
                        ****************** ****************************************

pCbAOD1-amplified       TTTACTATATTTACTATCTATAAATAAATTCATATTATAATATTAATTTATATTCGCTTA
gbM81702.1              TTTACTATATTTACTATCTATAAATAATTTCATATTATAATATTAATTTATATTCGCTTA
                        ************************** ******************************

pCbAOD1-amplified       ATTAAAATGCTCTTTTCCATCATCATCATCATCATCATCACGAGTTTTCGGTTATCA
gbM81702.1              ATTAAAATGCTCTTTTCCATCATCATCA---TCATCATCATCACGAGTTTTCGGTTATCA
                        ***************************   ******************************

pCbAOD1-amplified       ATACTCTTTTCATTAATTTCTAGAATTTCATTATTTATTTTTTATTGACTGGAAATTTTC
gbM81702.1              ATACTCTTTTCATTAACTTCTAGAATTTCATTATTTATTTTTTATTGACTGGAAATTTTC
                        *************** ****************************************

pCbAOD1-amplified       AATCAATTTTATTTATTTTTATTTATTTATTTTCATATTCTTAGATTTAAACTTTTTAGA
gbM81702.1              AATCAATTTTATTTATTTTTATTTATTTATTTTCATATTCTTAGATTTAAACTTTTTAGA
                        ************************************************************

pCbAOD1-amplified       TGACCGCTATTTTACTTACTTACTTACTGTTGTTTTATATTATG----------------
gbM81702.1              TGACCGCTATTTTACTTACTTACTTACTTACTTACTTACTTACTTACATACCTACT
                        ***************************    *  *   ***

pCbAOD1-amplified       --------------ATAAGAATTAATTTTCATATTTATGATGATGATGATGTAAATTTAA
gbM81702.1              TACTGTGATTTTATAATATGATAAGAATTAATTTTCATATTTATGATGATGTAAATTTAA
                                     *  *  ** *   ** ** ** **  * *****************

pCbAOD1-amplified       CCTAGTATACTATTTTAAAGTTATCACTATCTTTTAGTGCTGGCATTTTTTATTCTATTT
gbM81702.1              CCTAGTATACTATTTTAAAGTTATCACTATCTTTTAGTGCTGGCATTTTTTATTCTATTT
                        ************************************************************

pCbAOD1-amplified       TCATATATGTATATACGTAAATTAAGTATCATCACGCTGCTTACTGTACGTTTAAAATGT
gbM81702.1              TCATATATGTATATAAGTAAATTAAGTATCATCACGCTGCTTACTGTACGTTTAAAATGT
                        *************** ****************************************
```

```
pCbAOD1-amplified       GGAGATGGAAATAGAGATGGGGATGAAGATGAAGATGATGAGAATTATAAACCATTCATT
gbM81702.1              GGAGATGGAAATAGAGATGGGGATGAAGATGAAGATGATGAGAATTATAAACCATTCATT
                        ************************************************************

pCbAOD1-amplified       CATTAATCAATCAATATAACTTATAAAAAAATTTATATTTAAATGAATTAATTTCCTTTA
gbM81702.1              CATTAATCAATCAATATAACTTATAAAAAAATTTATATTTAAATGAATTAATTTCCTTTA
                        ************************************************************

pCbAOD1-amplified       TTTTAATAATATCGTTAATTCTTTTAAATTCTATTTTATTTTAATTCTTTCTTTATCATA
gbM81702.1              TTTTAATAATATCGTTAATTCTTTTAAATTCTATTTTATTTTAATTCTTTCTTTATCATA
                        ************************************************************

pCbAOD1-amplified       GTTATCATATAACAATTATATAACATAGATACACAATTATTATTTCATTATCATATTATT
gbM81702.1              GTTATCATATAACAATTATATAACATAGATACACAATTATTATTTTATTATCATATTATT
                        ******************************************** **************

pCbAOD1-amplified       TTTTAAAATATTGATTATTTTTAAAATAATATCTTAATTAATTAATTTTTACGAATATAC
gbM81702.1              TTTTAAAATATTGATTATTTTTAAAATAATATCTTAATTAATTAATTTTTACGAATATAC
                        ************************************************************

pCbAOD1-amplified       AAATTTTAACGACTTACTTTTTTTAACGAATTTTAACGAACTTTTAAAAAAACAAAAAAA
gbM81702.1              AAATTTTAACGACTTTCTTTTTTTAACGAATTTTAACGAACTTTTAAAAAAACAAAAAAA
                        ************** ****************************************

pCbAOD1-amplified       AAAAAACAAAATTATTTTTCAATA
gbM81702.1              AAAAAACAAAATTATTTTTCAATA
                        ************************
```

The alignment shows the full length *CbAOD1* promoter. The different colors (shown below) represent the fragments 1-5 and the underlined part represents fragment 6.

$P_{CbAOD1\text{-}F1}$ *(CbARS)*, $P_{CbAOD1\text{-}F2}$, $P_{CbAOD1\text{-}F3}$, $P_{CbAOD1\text{-}F4}$, $P_{CbAOD1\text{-}F5}$

7.2 *CalB* expression with methanol induction

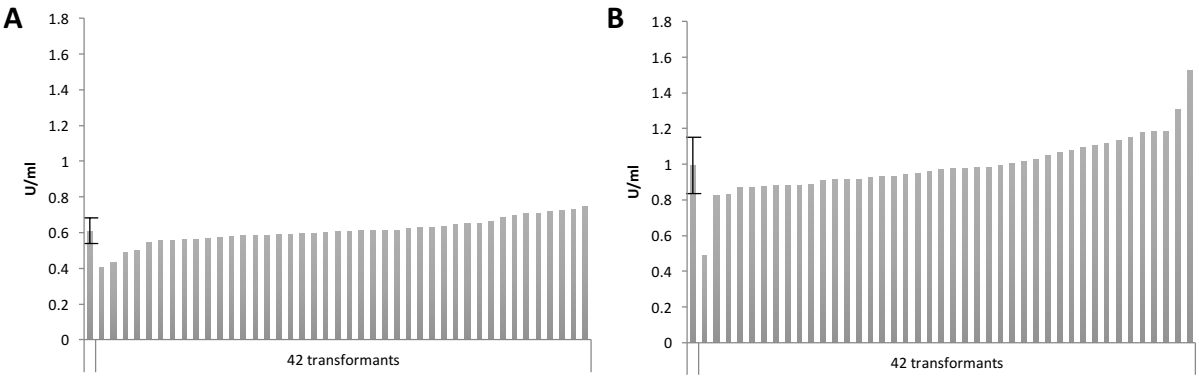

Figure 7.1 *CalB* expression with different ARSs. **A**, *CAT1-ARS*; **B**, *CbARS*. P_{CAT1} driven expression values after 108 h (60 h + 48 h induction) cultivation in selective media from 42 transformants are shown. The first bar on the left indicates the mean value and the error bar represents the standard deviation.

7.3 gBlocks and synthetic genes

gBlock1-Zeo

CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCT
ACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAA
ATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATC
CTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCA
GCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG
AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGC
ACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTT
GTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTT
TTGCTCACATGTATTTAAATTAATCGAACTCCGAATGCGGTTCTCCTGTAACCTTAATTGTAGCATAGATCACTTAAATAAAC
TCATGGCCTGACATCTGTACACGTTCTTATTGGTCTTTTAGCAATCTTGAAGTCTTTCTATTGTTCCGGTCGGCATTACCTAA
TAAATTCGAATCGAGATTGCTAGTACCTGATATCATATGAAGTAATCATCACATGCAAGTTCCATGATACCCTCTACTAATGG
AATTGAACAAAGTTTAAGCTTCTCGCACGAGACCGAATCCATACTATGCACCCCTCAAAGTTGGGATTAGTCAGGAAAGCTGA
GCAATTAACTTCCCTCGATTGGCCTGGACTTTTCGCTTAGCCTGCCGCAATCGGTAAGTTTCATTATCCCAGCGGGGTGATAG
CCTCTGTTGCTCATCAGGCCAAAATCATATATAAGCTGTAGACCCAGCACTTCAATTACTTGAAATTCACCATAACACTTGCT
CTAGTCAAGACTTACAATTAAAATGAGAAGAGCGAATTCGGCGCGCCGGTAAGATCCAAATCGATGAATTGACCAAGCACTAC
GGTATGAAGCCAGAAGACTACACTGCTGTCAGATGTGGTATGAATGTCGCCAAGTACATCATCGAAGATAAGATTGATGCTGG
TATTGGTATCGAATGTATGCAACAAGTCGAATTGGAAGAGTACTTGGCCAAGCAAGGCAGACCAGCTTCTGATGCTAAAATGT
TGAGAATTGACAAGTTGGCTTGCTTGGGTTGCTGTTGTTTCTGTACCGTTCTTTACATCTGCAACGATGAATTTTTGAAGAAA
AACCCTGAAAAGGTCAGAAAGTTCTTGAAAGCCATCAAGAAGGCAACCGACTACGTTCTAGCCGACCCTGTGAAGGCTTGGAA
AGAATACATCGACTTCAAGCCTCAATTGAACAGCTCTTCATAAGCGGCCGCTCAAGAGGATGTCAGAATGCCATTTGCCTGAG
AGATGCAGGCTTCATTTTTGATACTTTTTTATTTGTAACCTATATAGTATAGGATTTTTTTTGTCATTTTGTTTCTTCTCGTA
CGAGCTTGCTCCTGATCAGCCTATCTCGCAGCAGATGAATATCTTGTGGTAGGGGTTTGGGAAAATCATTCGAGTTTGATGTT
TTTCTTGGTATTTCCCACTCCTCTTCAGAGTACAGAAGATTAAGTGAGACCTTCGTTTGTGC

gBlock2-Zeo

GAAGATTAAGTGAGACCTTCGTTTGTGCGGATCCTTCAGTAATGTCTTGTTTCTTTTGTTGCAGTGGTGAGCCATTTTGACTT
CGTGAAAGTTTCTTTAGAATAGTTGTTTCCAGAGGCCAAACATTCCACCCGTAGTAAAGTGCAAGCGTAGGAAGACCAAGACT
GGCATAAATCAGGTATAAGTGTCGAGCACTGGCAGGTGATCTTCTGAAAGTTTCTACTAGCAGATAAGATCCAGTAGTCATGC
ATATGGCAACAATGTACCGTGTGGATCTAAGAACGCGTCCTACTAACCTTCGCATTCGTTGGTCCAGTTTGTTGTTATCGATC
AACGTGACAAGGTTGTCGATTCCGCGTAAGCATGCATACCCAAGGACGCCTGTTGCAATTCCAAGTGAGCCAGTTCCAACAAT
CTTTGTAATATTAGAGCACTTCATTGTGTTGCGCTTGAAAGTAAAATGCGAACAAATTAAGAGATAATCTCGAAACCGCGACT
TCAAACGCCAATATGATGTGCGGCACACAATAAGCGTTCATATCCGCTGGGTGACTTTCTCGCTTTAAAAAATTATCCGAAAA
AATTTTCTAGAGTGTTGTTACTTTATACTTCCGGCTCGTATAATACGACAAGGTGTAAGGAGGACTAAACCATGGCTAAACTC
ACCTCTGCTGTTCCAGTCCTGACTGCTCGTGATGTTGCTGGTGCTGTTGAGTTCTGGACTGATAGACTCGGTTTCTCCCGTGA
CTTCGTAGAGGACGACTTTGCCGGTGTTGTACGTGACGACGTTACCCTGTTCATCTCCGCAGTTCAGGACCAGGTTGTGCCAG
ACAACACTCTGGCATGGGTATGGGTTCGTGGTCTGGACGAACTGTACGCTGAGTGGTCTGAGGTCGTGTCTACCAACTTCCGT
GATGCATCTGGTCCAGCTATGACCGAGATCGGTGAACAGCCCTGGGGTCGTGAGTTTGCACTGCGTGATCCAGCTGGTAACTG
CGTGCATTTCGTCGCAGAGGAACAGGACTAA

gBlock1-GUT

GAAGATTAAGTGAGACCTTCGTTTGTGCGAGCTCATACCGAAAGGTAAACAACTTCGGGGAATTGCCCTGGTTAAACATTTAT
TAAGCGAGATAAATAGGGGATAGCGAGATAGGGGGCGGAGAAGAAGAAGGGTGTTAAATTGCTGAAATCTCTCAATCTGGAAG
AAACGGAATAAATTAACTCCTTCCTGAGATAATAAGATCCGACTCTGCTATGACCCCACACGGTACTGACCTCGGCATACCCC
ATTGGATCTGGTGCGAAGCAACAGGTCCTGAAACCTTTATCACGTGTAGTAGATTGACCTTCCAGCAAAAAAAGGCATTATAT
ATTTTGTTGTTGAAGGGGTGAGGGGAGGTGCAGGTGGTTCTTTTATTCGTCTTGTAGTTAATTTTCCCGGGGTTGCGGAGCGT
CAAAAGTTTGCCCGATCTGATAGCTTGCAAGATGCCACCGCTTATCCAACGCACTTCAGAGAGCTTGCCGTAGAAAGAACGTT
TTCCTCGTAGTATTCCAGCACTTCATGGTGAAGTCGCTATTTCACCGAAGGGGGGGTATTAAGGTTGCGCACCCCCTCCCCAC
ACCCCAGAATCGTTTATTGGCTGGGTTCAATGGCGTTTGAGTTAGCACATTTTTTCCTTAAACACCCTCCAAACACGGATAAA
AATGCATGTGCATCCTGAAACTGGTAGAGATGCGTACTCCGTGCTCCGATAATAACAGTGGTGTTGGGGTTGCTGTTAGCTCA
CGCACTCCGTTTTTTTTCAACCAGCAAAATTCGATGGGGAGAAACTTGGGGTACTTTGCCGACTCCTCCACCATACTGGTATA
TAAATAATACTCGCCCACTTTTCGTTTGCTGCTTTTATATTTCAAGGACTGAAAAAGACTCTTCTTCTACTTTTTCACACTAT
ACCACAGATATATCTACTATAATGGGAAAGGACTACACCCCTCTTGTCG

gBlock2-GUT

ATGGGAAAGGACTACACCCCTCTTGTCGCTACCATTGATATCGGTACTACCTCCACTCGTGCTATCCTGTTCGACTACCACGG
TCAAGAGGTGGCTAAGCACCAAATCGAGTATTCTACCTCCGCTCAAGATGACATCAAGAGAAAGAGATCCCAGATCATCTCTT
CCGAAGGTATCTCCCTTACCGTTTCCGACGACCTTGAGGTCGAATCTGTTGACAACAAGGCAGGTCCAACTCTTCAATTCCCA
CAACCTGGTTGGGTTGAATGTAGACCTTCTCACATTCTTGCTAACGCTGTTCAGTGTTTGGCTGCTTGTTTGGTTACTATGGA
GAACAAGAACTTGGATCGTGACGAGAAGAACAAGTACAAGTTGATCTCTATCGGTGTTGCTAACATGAGAGAAACTACCGTCG
TTTGGTCTAAGAAAACTGGAAAGCCATTGTACAACGGTATTGTTTGGAACGACACCAGAAACAATGATATCGTTGACGAGTAC
ACTGCTAAGTACTCTGAGAAAGAAAGAGAGGAGATGAGAACTCTTTGTGGTTGCCCAATCTCTACCTACTTCTCCGCTACCAA
GTTCAGATGGTTGCTTAAGCACGTTCCAGAAGTGAAACAGGCTTACGATAACGCAGACGGTGACCTGATGTTCGGAACTATTG
ACTCCTGGCTTATCTATCACCTGACTAACGAGAAGTCTCACGTTACCGATGTCACTAACGCATCAAGAACCAACTTCATGAAC
ATTGAAACCAACAAGTACGACGATAGACTGCTGAAGTTCTGGGACGTGGATACTTCTAAGGTCATCTTACCAGAGATCAGATC
CTCTGCTGAGGTGTACGGTCACTTCAAAGTTCCTCACCTTGAGTCTATCGGTTACGTTGAGTCATACTTAACTGACGACGCTT
TAGCTTTGTTGGAGACTATCGAAGGAGCACCATTGGCTGGTTGTTTGGGTGACCAATCCGCTTCCTTGGTTGGACAGTTGGCT
GTCAGAAAAGGTGACGCCAAGTGTACTTACGGTACTGGTGCATTTCTGTTGTACAACACCGGTGATCAGACTTTGATTTCAGA
ACATGGTGCATTGACTACCGTGGGTTACTGGTTCCCTGGTTTAGACGAGTCTGAGGACGGTAAGCATTCATCAAAGCCACAAT
ACGCATTGGAGGGATCAATCGCAGTTGCTGGTTCTGTGGTCCAATGGCTGCGTGATAACCTGAGATTGATTTCTAAGGCTCAG
GACGTTGGTCCTTTGGCTTCACAAGTTGACAACTCCGGTGGTGTTGTGTTCGTTCCTGCATTCTCAGGTCTTTTCGCTCCTTA

TTGGGACTCTAACTCTCGTGGAACTATCTTTGGTTTGACTCAATACACTTCAGCTTCCCACATTGCTAGAGCTGCCTTGGAAG
GTGTGTGCTTTCAAACCAGAGCTATCCTTAAGGCTATGATTTCTGATGCAGGAGCATCTGCCGATTTCTTGGAGGAGTCCTCT
AAGGCAACTGGTCACAACCCACTGTCTGTTTTGGCTGTCGATGGTGGTATGTCAAAGTCCGACGAAATGATGCAGATTCAAGC
TGACATCTTGGGTCCTTGTGTGACCGTTCGTAGATCAATCAACCCTGAGTGTACTGCTCTTGGTGCTGCAATTGCTGCTGGTT
TCGGAGTTCCAAAGGAAGACAGAATCTGGGGATCTTTGAAAGAATGCACCGAGGCTATCCTTGAAGGTAACAAGATGTACTTG
GCTGCTGGTAACACCTCTTTAGACTTCAAGGCCACTTTGTCTGACGAAGTTAGAAGAAAGGAATGGAGATTGTGGGAAAACGC
TATCGCTAAGGCAAAGGGTTGGTTGAAAGATACTGCTTAA

gBlock3-GUT (ordered as gene)

CTCGAGTAAAAAAAATCCTTAGCTTTCGCTAAGGATTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTAT
TTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCTAGTGCTGCA
ATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGG
TCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGC
GCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGA
TCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTT
GGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGA
CTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAAT
ACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCT
GTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAG
CAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATGTTTAGTCCTCCTTAC
ACCTTGTCGTATTATACGAGCCGGAAGTATAAAGTAACAACACTCTAGACGCGTCCCGTAGAAAAGATCAAAGGATCTTCTTG

pHTA1

GGGGGGAGGACTCTCGTTTCCTATGATACCTTATGGAATGGGACAGTACGCGCTAATTTGTCTCGTTCTGGATAGACTGCTCT
GTTTTTCGTTACATACGAGTGGGAAGTGCGAACTTTTGTACCAGAGTGCGAAAAAAGTGGTGGCAATCCTGACGTTTGTACCA
AGCGCGACGGCACATTATTGGTCAGTTTGTGCATTTTAGTTTTTCTTGAAAACATCTCGAAACATAACAACTAGGACACCAAT
TCTACAGTGAAAACCATTCAGTGGAATGTGAACCCACATAAATCTTGCAGCAATAATACAACAATTTATGGCTCCTGCATTCT
GATTGGCTGAGCGTTTATTGCCATCGTGACTGTCCAATTTAGGACACACCCTTTCCTCAGAATATATAAACTGGTGATAATTC
CTTCGTTCTGAGTTCCATCTCATACTCAAACTATATTAAAACTACAACA

pAOX1

AACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCCGACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGC
AACAGGAGGGGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTTCTCCTCAACACCCACTTTTGCCATC
GAAAAACCAGCCCAGTTATTGGGCTTGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCATGACTTTAT
TAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTGTTTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACAT
CACTCCAGATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGGCCCAAAACTGACAGTTTAAACGCTG
TCTTGGAACCTAATATGACAAAAGCGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACGGCCAGTTGGT
CAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTCTTGTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAATG
CTTAGCGCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGCAAATGGGGAAACACCCGCTTTTTGGATGA
TTATGCATTGTCTCCACATTGTATGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCATGATCAAAATTTAA
CTGTTCTAACCCCTACTTGACAGCAATATATAAACAGAAGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTTATCATCATTATTA
GCTTACTTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACTTTTAACGACAACTTGAGAAGATCAAAA
AACAACTAATTATTCGAAACG

Dalpha (mutated alpha mating factor signal sequence)

ATGAGATTCCCATCTATTTTCACCGCTGTCTTGTTCGCTGCCTCCTCTGCATTGGCTGCCCCTGTTAACACTACCACTGAAGA
CGAGACTGCTCAAATTCCAGCTGAAGCAGTTATCGGTTACTCTGACCTTGAGGGTGATTTCGACGTCGCTGTTTTGCCTTTCT
CTGCTTCCATTGCTGCTAAGGAAGAGGGTGTCTCTCTCGAGAAGAGAGAGGCCGAAGCT

7.4 Codon optimization

Predicted *GUT1* 5'UTR (using P_{GUT1})

CTTTGCCGACTCCTCCACCATACTGGTATATAAATAATACTCGCCCACTTTTCGTTTGCTGCTTTTATATTTCAAGGACTGAA
AAAGACTCTTCTTCTACTTTTTCACACTATACCACAGATATATCTACTATA

Predicted *GUT1* 3'UTR (using *CbARS* as transcription terminator)

GGAGTATACGTAAATATATAATTATATATAATCATATATATGAATACAATGCAATGAAAGTGAATATGATAAGATTGAAATAA
TAACAAACAGCGATAAATATATCTCAAAATGGAGTTACACAACAAATAATAATAAAATATAAATTATAAATTATAAATTATAA
AAGAATAAAAAATAAACCCCACTAATTTATTTTATTAAAAGATAGATTGGTATCTTTACTTAATAACAATTCTGAAACTTTAT
TCACTTAATTTTATTTAACTTATTTAATTTATTTT

7.5 Knockout generation

7.5.1 Homologous parts *GUT1*

5' homologous region

GCAGCCTGCATTCTCTTGATTTTATGGGGGAAACTAACAATAGTGTTGCCTTGATTTTAAGTGGCATTGTTCTTTGAAATCGA
AATTGGGGATAACGTCATACCGAAAGGTAAACAACTTCGGGGAATTGCCCTGGTTAAACATTTATTAAGCGAGATAAATAGGG
GATAGCGAGATAGGGGGCGGAGAAGAAGAAGGGTGTTAAATTGCTGAAATCTCTCAATCTGGAAGAAACGGAATAAATTAACT
CCTTCCTGAGATAATAAGATCCGACTCTGCTATGACCCCACACGGTACTGACCTCGGCATACCCCATTGGATCTGGTGCGAAG
CAACAGGTCCTGAAACCTTTATCACGTGTAGTAGATTGACCTTCCAGCAAAAAAAAGGCATTATATATTTTGTTGTTGAAGGG
GTGAGGGGAGGTGCAGGTGGTTCTTTTATTCGTCTTGTAGTTAATTTTCCCGGGGTTGCGGAGCGTCAAAAGTTTGCCCGATC
TGATAGCTTGCAAGATGCCACCGCTTATCCAACGCACTTCAGAGAGCTTGCCGTAGAAAGAACGTTTTCCTCGTAGTATTCCA
GCACTTCATGGTGAAGTCGCTATTTCACCGAAGGGGGGGTATTAAGGTTGCGCACCCCCTCCCCACACCCCAGAATCGTTTAT
TGGCTGGGTTCAATGGCGTTTGAGTTAGCACATTTTTTCCTTAAACACCCTCCAAACACGGATAAAAATGCATGTGCATCCTG
AAACTGGTAGAGATGCGTACTCCGTGCTCCGATAATAACAGTGGTGTTGGGGTTGCTGTTAGCTCACGCACTCCGTTTTTTTT
CAACCAGCAAAATTCGATGGGGAGAAACTTGGGGTACTTTGCCGACTCCTCCACCATACTGGTATATAAATAATACTCGCCCA
CTTTTCGTTTGCTGCTTTTATATTTCAAGGACTGAAAAAGACTCTTCTTCTACTTTTTCACACTATACCACAGATATATCTAC
TATA

3’ homologous region

GAAGAGCAGCTGTAATTATATTATCATGTTAGGTCATTAGACAATGTATTGTAGTTATATATGTAATGATAATGAAGAATGGA
AATCGATGGATATCGCGGATTTAATCACTTCCTCATCTCTTTCTTTCAGATTGCACTCAACGTAGGCAGCAGCAGTCCATGAC
CAGTGTGAAAGACTCCATCACTATTGACAGAGGGCTGGCACAGAATGTTGAAAAGATCCATGGAGTTAACCCTGTGTTTTTGA
TAGACAAAATTCTCAGGGAAAAAATCTTAGATAGTCTGTTTTGGAAAAAACACTGCTTTCTACTAGATATTCTGGAACTGCAA
GATCTTGCAAGTGAAGTGGAGATTATTGGCACATACGACACCAATCAAAACTCTAGAGCAACCGACTATATTTGTCTGTTACT
GAAAATGCTACAGTTACAGCCGAAAGACGAGATAGTGATATACATGTTGACCCAACCCTATTTCAAGTATGTGACAGCTCTGG
CAGCGTTGTATATTCGGTTGGTTTTCCCTTCTGTGAGAGTTTATCAGTTGTTAGAGCCGTTATTGAATGACTATAGAAAACTG
CGAATAAGAAAGCACGGCAACGCTGACTTGACCTACGTGGATCAGCTGGTAGATGAGTTATTAACTGAGGAGAAGTTTTGTGG
GTTAACACTACCCAGATTGGTCAACCGATTGCGACTGGAGGATGACGGTCTACTAGACCCTAGAGAAAGTGAACTACAGAGCG
AGTTAGAAAATGAGGAAGATGATATTGGGGGTTGAAATTGGGCTAATTTATTATGACCATTACAATTATATTAGATACATGTA
CATAGGCAGGACTTGATCGTTACATCCTTAGGATCCTTTTCTCTTCTACTTATATGCTTAGAAAGCCCTAGCAGATGGTCTGG
GTTTATCTCTGCCAACTCTTGGTCCACCTCTGATTGGATCTTTCTTGGCCACTTTATTCTTAGGCTTGACCAATGAGTCTGTC
AAAA

7.5.2 Homologous parts *KU70*

5’ homologous region

GATCCAGAAGATCAGCGGTTTCACTATTGTTACCAGGTTCAGCTATAGCTAAGCTGCATACCTCTTCCTCTTCGTTCATCAGG
ATCCTGGGGCCTCGTAGTTTGGAGTTGGACCTTTCAAGTCCAGAATTAGTGGTGGCCATCGACTTTGTTTCGTCATCAATATC
CTGTAGTAGCTTCTCCATAGAATCGTTCCTAGTTTCGTAAAGATTGGAAGAGCTGCTTCCTCTTGAAGGTTTCAAGAGTGGGA
TGGCCCCTTCCATCTCAAGAGAGTTCATTTCTTGTCCGCTTGACTAGTACTTCTGTCGGAAGAAAAAAAAAAATACTCAAGTGC
CTGATAGTACTCGCTTTGTTTGTTTTGGCTAAACTGAAATCTCACGGGTGATTACTTGTTTACATATTTCCCAAAACAGAGAG
TCTCCCCAGCTCGAGCGAGTAGGGAGGCATGCATCTTGAGATGTGCTTAGATAAGCAATGTGCATTGTGGCTCTTCAGAAGCC
TAACTTTAATGGGGTAAAATATCAGCAGTAGGCTGAATATAGAGCATGGCAACGTTTACGTCAACAGAGTTTAATAGTGAAAA
CTACGATAGGTTCAGGCCTGTATACCCAGACGAGTTGTATCAGCAATTAGTGGATTATCATGTTGGTGCTAAAGGATTGTGTG
TCGATGTTGGATGTGGCTCTGGCCAGGCAACGTTCACCCTGAAAAAGTACTTTGACAAGGTCATAGGATCTGATATCTCAGAA
AATCAGTTGGCAGTGGCTAGGAAAAGACAGCCTGCTGGAATTGAATTTAGGTTAGGAACGGGCGAGGATTTTTCTTGGCTAAC
GGAGACGCCAGATGTAATCACCGCCGCTGAATGCCTGCACTGGGTAGACCCACAGAAATTCGTTGCAAATGTCGCCAATTCTC
TAAGGGATCATGGCACCTTGAGTTACTGGCTTTATACAGAACCAATTTTTCAGAACGAAAGGGCCAATCAGGTTTACAACAAA
TTCACCTACGGGTCTGACTACCTGGGCCCATATTGGGATCCTGGCCGTACACATTTCAGAAATCATTTAAAGGAGTTGAATCA
CATCTTACTGGATAGCGAGCTTTTTGACGAAGTGAAAATTTCTAATTTTAAACAAGAGGAAGGGGTCAAAAACGGAGATATCT
TATACTTGGAAAAAGAGATGACAATCAGTGATTTCATCAATTTTGTATCTAGTTGGCCTTCTGTGTTTTCGTGGAAGCAGCAA
CGAGGAAAGGAGGGTATCCTAGATGATTTTTACAACGAACTGAACGACTGCTTTGAGGGGGGTAACATGAAAGTAATATGGAA
CTCCGTCCTAGTATTTGCCAGGAGGAAGCAAAGGGTTGTATAGGCTTTAGTACTTATAGAGGAAACGGGGTTACGTGCAAGCG
CGCATGCCTGAGCTTTGAGGGGGGGGACTTTCACATCTCTTCTTCTCACACTTAGCCCTAACACAGAGAATAATAAAAAGCAT
TGCAAG

3’ homologous region

TTCGTGTTCCTTACTTTTTCCTCGCAACGTGTTTTTTTCCCACCACATTGCCTATGTTGTAATGCAATGCAGATGCTGGCCCA
GTTTTTGACGATTCTCGAAAATTGGCATTTTCGTCGATGCCATTGGCCAAACTGAAAATTCAAGACAAAATAGATTGGATTTT
ATCTGCAACGTCTTCCACCTACACAACCACTCTACAAACTTCAGACAAACATGTTTATAAAAGCAGCTACTAGATCCAAAATG
ACAAGTTCGTTATTCTCTACTACGTTTGTTGTGGCATTTGGATTGGTGGCTAGCAACAACCTCTTGCCATGTCCTGTTGACCA
CTCTATGAATAACGAGACTCCGCAAGAATTGAAACCATTGCAGGCTGAATCTTCTACTAGAAAGTTGAACTCTTCCGCTTAAG
TCAAATAAAACTACTGACACAGATGATGCACAGAAACAACGGATCACGCTCTTGACTGATTAGTCCCGTCATTTTGGTTCTCA
TTTTCTTCACAGTCACCTATCAATGTATGATCACCTGGAAGGATTTCCCTACGATACTTCAAATCTTTTACTTGATAATATTA
CTCATTATGGCTCAGGAATGCAGACTGCCTGATTCAAGACGCTGCTCTTCTTATTTAACACTTGTACACTAACCCCATGGAAG
CCAGGGAAGGGAATAACCATCTCTCTGGTAATAAATCGGTCTTTATTTATGCATAGAAAAGGAATCTATTATATTTCGTTCAT
TTGGCACTCTGCTAACTGTAGATTAACGGGTCTCGTAAATTCAAAATCTTCTTCCGATCAAACCGGGGTGAAATATTACTTCT
CGTGCATAGCTAATTTTCAAATAACCGTCCTAAAATGAACGGTCATTTACCTGGACTCTCTTGCCAAATGGGCAACAAAACAT
AAAGCTGATCAGAACGTAACTAGTCTCTCGGAATCCATTTGGTTCCCCAAAAAGGGCTGGTCGACGTCGTCCTGGGACGATGA
CCTTCTATTCGGAAGCAGCTCTTTAAATACAGTGGATACAGATCTTTGAAACAGTTGATAAGTGCTAAAGAATACTCTAATAA
TACAGTTGACCACTTTATCGGTATTGTTCTCTTCAATTGGTCTTGAAGAACTCTCGACATTTTGTCCTGCAAAATTGTTGAGT
CTGGAAAATCCACCATTTCTTTTGATTCCCCTATTATAGACAAACCAAGCCCCTCCCAGGATGGATACACTTAACAACGTGAA
TAACGTGCCATGTTTGTCGTTATATTGTACACCACCAAAGACGGTCTTTGAATTCCAGAATTCAGAATCGTCATCGCCATAAC
TGAAACCTATCGTGTAGAGATTGCGAATATTTCTTTCTCTATCAGGTACTGAATATTCATTTTCAAGGACAATGAACTTCCGA
GATGCTCCAGACGACGAGTAGACAAGGTCATACACTTTTATCTTTTTGTCCGTAAAAACTTGTAATCCAGCCAAGTTTCTCCT
TCATCC

8 Appendix

8.1 Strains generated during the thesis

#	Name of the plasmid/strain	Organism	*E. coli* #	Description
RW001	pPpT4mutZeoMlyI-intArg4-pCAT1-500-eGFP-PARS1	*E. coli* TOP10 F'	E	ARS test
RW002	pPpT4mutZeoMlyI-intArg4-pCAT1-500-eGFP-CAT1ARS	*E. coli* TOP10 F'	E	
RW003	pPpT4mutZeoMlyI-intArg4-pCAT1-500-eGFP-CbAOD1-F1	*E. coli* TOP10 F'	E	
RW004	pPpT4mutZeoMlyI-intArg4-pCAT1-500-eGFP-CbAOD1-F2	*E. coli* TOP10 F'	E	
RW005	pPpT4mutZeoMlyI-intArg4-pCAT1-500-eGFP-CbAOD1-F3	*E. coli* TOP10 F'	E	
RW006	pPpT4mutZeoMlyI-intArg4-pCAT1-500-eGFP-CbAOD1-F4	*E. coli* TOP10 F'	E	
RW007	pPpT4mutZeoMlyI-intArg4-pCAT1-500-eGFP-CbAOD1-F5	*E. coli* TOP10 F'	E	
RW008	pPpT4mutZeoMlyI-intArg4-pCAT1-500-eGFP-CbAOD1-F6	*E. coli* TOP10 F'	E	
RW009*	pPpT4mutZeoMlyI-intArg4-pCAT1-500-eGFP-CbAOD1-Full	*E. coli* TOP10 F'	E	
RW012	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII-PARS1-TT test	*E. coli* TOP10 F'	E	TT test
RW013	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII-AOD_TT-TT test	*E. coli* TOP10 F'	E	
RW014	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII-GUT1_TT-TT test	*E. coli* TOP10 F'	E	
RW015	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII-CbAOD1-TT test	*E. coli* TOP10 F'	E	
RW016	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII-PARS1-TT test Clone 2A	*P. pastoris* BSYBG10	RW012	
RW017	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII- CbAOD1-F1-TT test Clone 10F	*P. pastoris* BSYBG10	RW015	
RW018	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII- AOD_TT-TT test Clone 5E	*P. pastoris* BSYBG10	RW013	
RW019	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII-GUT1_TT-TT test Clone 11C	*P. pastoris* BSYBG10	RW014	
RW020	pPpT4mutZeoSapI-PARS1-MCS	*E. coli* TOP10 F'	E	Final ARS vector test
RW021	pPpT4mutZeoSapI-CbAOD1-F1-MCS	*E. coli* TOP10 F'	E	
RW022	pPpT4mutZeoSapI-PARS1-SapI cloning	*E. coli* TOP10 F'	E	
RW023	pPpT4mutZeoSapI-CbAOD1-F1-SapI cloning	*E. coli* TOP10 F'	E	
RW024	pPpT4mutZeoSapI-CbAOD1-F1-pCAT1-500-eGFP	*E. coli* TOP10 F'	E	
RW025	pPpT4mutZeoSapI-PARS1-pCAT1-500-eGFP	*E. coli* TOP10 F'	E	
RW026	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII-pCAT1-TT test	*E. coli* TOP10 F'	E	TT test
RW027	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII-pCAT1noCore-TT test	*E. coli* TOP10 F'	E	
RW028	pPpT4mutZeoSapI-CbAOD1-F1-pGAP239A-eGFP	*E. coli* TOP10 F'	E	pGAP mutation
RW029	pPpT4mutZeoSapI-CbAOD1-F1-pGAP239T-eGFP	*E. coli* TOP10 F'	E	
RW030	pPpT4mutZeoSapI-CbAOD1-F1-pGAP239G-eGFP	*E. coli* TOP10 F'	E	
RW031	pPpT4mutZeoSapI-CbAOD1-F1-pGAPwt-eGFP	*E. coli* TOP10 F'	E	
RW033	CbARS-ZEO-pCAT1-SapI	*E. coli* TOP10 F'	E	Final Zeo vectors
RW034	CbARS-ZEO-pGAP-SapI	*E. coli* TOP10 F'	E	
RW035	CbARS-ZEO-pHTA1-SapI	*E. coli* TOP10 F'	E	
RW036	CbARS-ZEO-pAOX1-SapI	*E. coli* TOP10 F'	E	
RW037	CbARS-ZEO-pCAT1-SS-SapI	*E. coli* TOP10 F'	E	
RW038	CbARS-ZEO-pGAP-SS-SapI	*E. coli* TOP10 F'	E	
RW039	CbARS-ZEO-pHTA1-SS-SapI	*E. coli* TOP10 F'	E	

RW040	CbARS-ZEO-pAOX1-SS-SapI	*E. coli* TOP10 F'	E	
RW041	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII-pCAT1-TTtest Clone 2B	*P. pastoris* BSYBG10	RW026	TT test
RW042	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII-pCAT1noCore-TTtest Clone 11E	*P. pastoris* BSYBG10	RW027	
RW043	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII-SPG5-TTtest Clone 3D	*P. pastoris* BSYBG10	TV0756	
RW044	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII-NotIAOX*-TTtest Clone 12C	*P. pastoris* BSYBG10	TV0020	
RW045	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII-NoTT-TTtest Clone 1E	*P. pastoris* BSYBG10	TV0758	
RW046	pPpT4mutZeoMlyI-intArg4-eGFP-AOXIBglII-CAT1ARS-TTtest Clone 9A	*P. pastoris* BSYBG10	TV1254	
RW047	CbARS-GUT1-pCAT1-SapI	*E. coli* TOP10 F'	E	Final *GUT1* vectors
RW048	CbARS-GUT1-pHTA1-SapI	*E. coli* TOP10 F'	E	
RW049	CbARS-GUT1-pCAT1-SS-SapI	*E. coli* TOP10 F'	E	
RW050	CbARS-GUT1-pHTA1-SS-SapI	*E. coli* TOP10 F'	E	
RW051	CbARS-ZEO-pCAT1-SapI-noNotI	*E. coli* TOP10 F'	E	Final vectors without *Not*I site
RW052	CbARS-ZEO-pGAP-SapI-noNotI	*E. coli* TOP10 F'	E	
RW053	CbARS-ZEO-pHTA1-SapI-noNotI	*E. coli* TOP10 F'	E	
RW054	CbARS-ZEO-pAOX1-SapI-noNotI	*E. coli* TOP10 F'	E	
RW055	CbARS-ZEO-pCAT1-SS-SapI-noNotI	*E. coli* TOP10 F'	E	
RW056	CbARS-ZEO-pGAP-SS-SapI-noNotI	*E. coli* TOP10 F'	E	
RW057	CbARS-ZEO-pHTA1-SS-SapI-noNotI	*E. coli* TOP10 F'	E	
RW058	CbARS-ZEO-pAOX1-SS-SapI-noNotI	*E. coli* TOP10 F'	E	
RW059	CbARS-GUT1-pCAT1-SapI-noNotI	*E. coli* TOP10 F'	E	
RW060	CbARS-GUT1-pHTA1-SapI-noNotI	*E. coli* TOP10 F'	E	
RW061	CbARS-GUT1-pCAT1-SS-SapI-noNotI	*E. coli* TOP10 F'	E	
RW062	CbARS-GUT1-pHTA1-SS-SapI-noNotI	*E. coli* TOP10 F'	E	
RW063	5prGUT1-flipper cassette-pJET1.2	*E. coli* TOP10 F'	E	Knockout cassettes
RW064	5prKU70-flipper cassette-pJET1.2	*E. coli* TOP10 F'	E	
RW065	3prGUT1-flipper cassette-pJET1.2	*E. coli* TOP10 F'	E	
RW066	3prKU70-flipper cassette-pJET1.2	*E. coli* TOP10 F'	E	
RW067	*P.pastoris* BSYBG11 ΔGUT1 ZeoR	*P. pastoris* BSYBG11	---	*ΔGUT1* host strains
RW068	*P.pastoris* BSYBG10 ΔGUT1 ZeoR	*P. pastoris* BSYBG10	---	
RW069* (BT7423)	*P.pastoris* BSYBG10 GUT1::FRT	*P. pastoris* BSYBG10	---	
RW070* (BT7424)	*P.pastoris* BSYBG11 GUT1::FRT	*P. pastoris* BSYBG11	---	
RW071	pCAT1-692-FLP-GEN	*E. coli* TOP10 F'	E	
RW072	CbARS-ZEO-pCAT1-eGFP	*E. coli* TOP10 F'	E	eGFP/dTomato expression plasmids
RW073	CbARS-ZEO-pGAP-eGFP	*E. coli* TOP10 F'	E	
RW074	CbARS-ZEO-pGAP-sTOM	*E. coli* TOP10 F'	E	
RW075	CbARS-ZEO-pHTA1-eGFP	*E. coli* TOP10 F'	E	
RW076	CbARS-ZEO-pAOX1-eGFP	*E. coli* TOP10 F'	E	
RW077	CbARS-GUT1-pCAT1-eGFP	*E. coli* TOP10 F'	E	
RW078	CbARS-GUT1-pHTA1-eGFP	*E. coli* TOP10 F'	E	
RW079	pPpT4mutZeoMlyI-CAT1ARS-pCAT1-CBH2v04	*E. coli* TOP10 F'	E	Secretory CalB/CBH2
RW080	pPpT4mutZeoMlyI-CAT1ARS-pCAT1-CalB	*E. coli* TOP10 F'	E	

RW081	pPpT4mutZeoMlyI-CbARS-pCAT1-CBH2v04	*E. coli* TOP10 F'	E	expression plasmids
RW082	pPpT4mutZeoMlyI-CbARS-pCAT1-CalB	*E. coli* TOP10 F'	E	
RW083	CbARS-ZEO-pGAP-EM72TinselPurple	*E. coli* TOP10 F'	E	Positive control
RW084	CbARS-ZEO-pGAP-TinselPurple	*E. coli* TOP10 F'	E	
RW085	CbARS-ZEO-pCAT1-eGFP integrated	*P. pastoris* BSYBG10	RW072	Genomic eGFP expression strains
RW086	CbARS-ZEO-pGAP-eGFP integrated	*P. pastoris* BSYBG10	RW073	
RW087	CbARS-ZEO-pHTA1-eGFP integrated	*P. pastoris* BSYBG10	RW075	
RW088	CbARS-ZEO-pAOX1-eGFP integrated	*P. pastoris* BSYBG10	RW076	
RW089	CbARS-GUT1-pCAT1-eGFP integrated	*P. pastoris* BSYBG10	RW077	
RW090	CbARS-GUT1-pHTA1-eGFP integrated	*P. pastoris* BSYBG10	RW078	

RW indicates the internal strain collection and BT the strain collection of the Institute of Molecular Biotechnology. * These strains are in both strain collections.