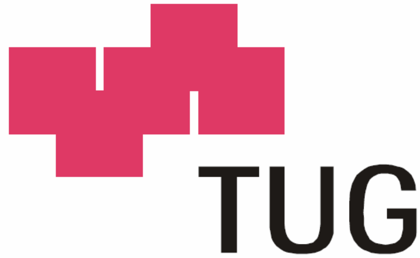

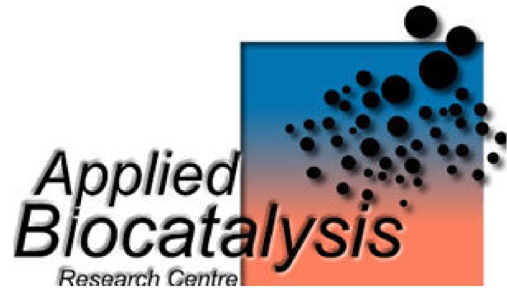

New strategies and tools for protein expression and engineering using *Pichia pastoris*

Sandra Ximena Abad Verdesoto

February 2010

Supervisor:
Ao.Univ.-Prof. Mag.rer.nat. Dr.rer.nat. Anton Glieder
Research Centre Applied Biocatalysis Institute of Molecular Biotechnology
Graz University of Technology

Deutsche Fassung:

Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008

Genehmigung des Senates am 1.12.2008

EIDESSTATTLICHE ERKLÄRUNG

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

Graz, am …………………………… ………………………………………………..

(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

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

…………………………… ………………………………………………..

date (signature)

Acknowledgement

Most importantly, I want to thank my supervisor Toni Glieder for the opportunity to work and learn in a stimulating and interesting environment.

I am grateful to the colleagues of the “Glieder group”. Especially the people from the third floor: Ulrike Schreiner, Aleksandra Mitrovich, Thorsten Bachler, Margit Winkler, Brigitte Höller, Karlheinz Flicker, Barbara Petschacher and Elena Loncar for their support in discussions, reading and correcting manuscripts, plating, etc…

Franz Hartner is kindly acknowledged for his support during the first year of my thesis and also Andrea Mellitzer for her help resolving problems with certain and often autonomous equipment on the first floor (=the fermenter) I want to thank “my” diploma students Kerstin Kitz, Astrid Hörmann, Christian Gruber and Gabriele Bergler for the time we worked together, their contribution for publications and for being my “test persons” during my unofficial supervision challenge.

Furthermore, I would like to thank Prof. Bernd Nidetzky for his support and the collaboration during the last year and for all the time he spent for manuscript preparation and correction.

Finally I am grateful to my family for their support during the last years as well as to my friends Sigrid Egger and Christiane Gödl for the girl talks, their advice and the nice time in Graz.

And finally, my deepest gratitude is expressed to Jochen Gerlach for all the profitable and occasionally less profitable discussions, for several manuscript proof-readings, for a lot of driven kilometres per year, and for his unconditional commitment for lactose-free diet.

Abstract

The focus of this thesis is set on new expression strategies and tools for efficient recombinant protein expression and engineering in *Pichia pastoris*. This methylotrophic yeast is a popular organism which has gained much attention during the last decades; nevertheless some research fields still remained passive like for example directed evolution.

Previously, a first tool like the PCR-based expression constructs has been successfully developed in our group and is being routinely used for ongoing scientific projects. In line with the PCR based expression constructs a *P. pastoris* platform strain has been developed, which boosts the transformation efficiency and facilitates library screening for directed evolution. Furthermore, the assembling of the expression cassettes became simpler due to a smaller expression construct size required with this strain. This new concept was also evaluated with two model enzymes namely *Candida antartica* lipase B and *Hevea brasiliensis* hydoxynitrile lyase.

Another fast and systematic engineering concept was applied for the “stepwise engineering of a *Pichia pastoris* D-amino acid oxidase whole cell catalyst”. Herein, the expression of the industrially relevant catalyst D-amino acid from *Trigonopsis variabilis* (*Tv*DAO) was significantly increased and an innovative whole cell biocatalyst was generated.

In addition an efficient analytical procedure for strain characterization was established. Strains containing comparable numbers of integrated expression constructs were identified and confirmed using quantitative real time PCR which was adapted for *Pichia pastoris*. This new protocol was systematically evaluated using different reporter systems and these studies also revealed the importance of molecular strain analysis for studies about promoter and chaperon effects on protein expression.

Kurzfassung

Der Schwerpunkt dieser Arbeit ist die Entwicklung neuer Werkzeuge und Strategien zur effizienten rekombinanten Protein Expression und für das Protein Engineering mit Hilfe von *Pichia pastoris*. In den letzten Jahrzehnten ist diese methylotrophe Hefe ein sehr populärer Organismus geworden, welcher viel Aufmerksamkeit erlangt hat. Trotzdem wurde die Verwendung in manchen Forschungsbereichen wie z. B. Directed Evolution nicht vorangetrieben.

Erste Werkzeuge wie PCR basierte Expressionskonstrukte wurden erfolgreich in unserer Arbeitsgruppe etabliert und werden in laufenden wissenschaftlichen Projekten verwendet. Passend zu diesen Konstrukten wurde ein *P. pastoris* platform Stamm hergestellt, welcher die Transformationseffizienz erhöht und das Screening von Mutanten Genbanken erleichtert. Weiters wurde der Zusammenbau von Expressionskassetten vereinfacht, da dieser platform Stamm die Verwendung kürzerer Konstrukte ermöglicht. Dieses neue Konzept wurde mit zwei Modelenzymen, der *Candida antartica* Lipase B und der *Hevea brasiliensis* Hydoxynitrile Lyase evaluiert.

Ein weiteres schnelles und systematisches Engineering Konzept wurde in der schrittweise Entwicklung von P*ichia pastoris* D-Aminosäure Oxidase Ganzzell Katalysator” angewandt. Dabei wurde die Expression eines industriell relevanten Katalysators nämlich der D-Aminosäure Oxidase von *Trigonopsis variabilis* (*Tv*DAO) signifikant verbessert, wodurch die Entwicklung eines innovativen *P. pastoris* Ganzzellkatalysators ermöglicht wurde.

Zusätzlich wurde eine analytische Methode zur Stamm Charakterisierung etabliert. Stämme mit vergleichbarer Anzahl an integrierten Expressionskassetten wurden identifiziert und mittels einer, speziell für *P. pastoris* adaptierten quantitative real time PCR Methode bestätigt. Das Protokoll wurde systematisch mit verschiedenen Reportersystemen evaluiert, wodurch die Bedeutung von Stamm Analysen auf molekularer Ebene für Promoter Studien und Chaperon Effekte unterstrichen wurde.

Content

1. Introduction ... 1
   1.1 Short introduction to *Pichia pastoris* ... 1
2. Concepts for intracellular targeted expression of enzymes in *Pichia pastoris* ... 2
   2.1 Nuclear and vacuolar targeting ... 2
   2.2 Mitochondrial targeting ... 2
   2.3 Endoplasmic reticulum and golgi complex targeting ... 3
   2.4 Membrane targeting ... 4
   2.5 Periplasmic targeting ... 4
   2.6 Peroxisomal targeting ... 5
      2.6.1 Peroxisomes and peroxisomal targeting in *Pichia pastoris* ... 5
      2.6.2 Engineering of D-amino acid oxidase for expression and use in *Pichia pastoris* ... 7
      2.6.3 Stepwise engineering of a *Pichia pastoris* D-amino acid oxidase whole cell catalyst ... 9
         2.6.3.1 Abstract ... 10
         2.6.3.2 Background ... 11
         2.6.3.3 Results and Discussion ... 14
         2.6.3.4 Conclusions ... 21
         2.6.3.5 Methods ... 22
3. Innovative high throughput technologies ... 29
   3.1 Construction of a ura$^-$ platform strain for library screening ... 30
   3.2 *Pichia pastoris* library platform strain ... 34
      3.2.1 Abstract ... 35
      3.2.2 Introduction ... 35
      3.2.3 Results and discussion ... 37
      3.2.4 Conclusions ... 49
      3.2.5 Materials and methods ... 50
   3.3 Gene dosage determination in *Pichia pastoris* ... 52
      3.3.1 Real time PCR based determination of gene copy numbers in *Pichia pastoris* ... 54
         3.3.1.1 Abstract ... 55
         3.3.1.2 Introduction ... 55
         3.3.1.3 Materials and methods ... 57
         3.3.1.4 Results and discussion ... 62
         3.3.1.5 Concluding remarks ... 67
4. Conclusions and future perspectives ... 69
5. References ... 71
6. Appendix ... 79
   6.1 Strain collection and VNTI files ... 79

1. Introduction

1.1 Short introduction to *Pichia pastoris*

Pichia pastoris is a methylotrophic yeast which has become very popular during the last decades. It has gained its popularity mainly due to the extraordinary ability in heterologous protein production, which includes examples yielding up to 20 g/L [1-6]. More recently, several helper proteins (e.g. calnexin, protein disulfide isomerase, binding protein, Sec4p, etc.) [7-10] have been described to enhance the yield of secreted proteins, and definitely boosted the success of this yeast. Nevertheless, the cellular understanding of *P. pastoris*, especially in comparison to e.g. *Saccharomyces cerevisiae* still remains at its beginning. This situation is changing since a deeper understanding due to an available genome sequence and consequently facilitates transcriptomics analysis [7, 11].

However, in addition to available genome omics data and their interpretation there are at least two additional bottlenecks for an efficient use of *P. pastoris* as expression host which were the starting point of this PhD thesis.

1. I want to emphasize on alternative expression strategies that differ from the standard intra- and extra-cellular expression strategies. Analysing the target protein environment in its native host and looking for similar conditions in the expression host can help to increase the success rate and yields in heterologous protein production. Therefore, one goal of this thesis was to demonstrate that a systematic analysis of the amino acid sequence of the target enzyme and an adaptation of the expression strategy based on this sequence analysis can contribute significantly to efficient expression.

2. Current technologies and tools for heterologous enzyme expression in *P. pastoris* are still limiting the efficiency of protein engineering efforts employing this yeast as a host system. Therefore a second part of this thesis was dedicated to alternative strains and molecular analytic tools which facilitate high throughput experimentation with this interesting commercially important host system.

2. Concepts for intracellular targeted expression of enzymes in *Pichia pastoris*

Extra-cellular expression in *P. pastoris* has developed to a standard recombinant expression approach, but following this main stream is not always the most efficient expression strategy. Alternative approaches and advantages by intracellular expression are shortly reviewed here. An example from own research, which includes sequence analysis of an industrially relevant enzyme and its successful peroxisomal targeting in the production host demonstrates the importance of such adapted expression strategies.

Relevant characteristics for intracellular expression such as pH, reductive/oxidative environment, protease-free localization, space and shelter need to be systematically analysed in regard to high level recombinant expression. Important to mention is that in the same way as a universal expression host does not exist so far, a universal expression/targeting strategy is currently non-existent. Every target enzyme has to be analysed separately to design the most promising expression strategy. Clustering into similar cases such as enzymes where ER signal sequences are found or where localization in mitochondria, membranes or peroxisomes can be predicted facilitates decisions about the expression strategy.

2.1 Nuclear and vacuolar targeting

On a first view, due to space related issues, most of the organelles seem not to be attractive as localization targets for high level expression. Therefore targeting to the nucleus and vacuoles seem not to be generally interesting targeting strategies for highly expressed enzymes. The latter one also disqualifies because of the fact of being a protease rich space.

2.2 Mitochondrial targeting

Space limitations in mitochondria can also be an issue for high level expression into these sub-cellular compartments. Noteworthy is that yeast cells under certain growth conditions, e.g. if glycerol is used as sole carbon source are able to increase the number of mitochondria [12]. In *P. pastoris* the proliferation of mitochondria is also caused by other substrates such as glucose, sorbitol as well as methanol and oleic acid [13]. Interestingly, the choice of the C-source has an effect on membrane

composition like diverse lipids chain length and its degree of unsaturation, which has an impact on membrane fluidity and consequently can affect or promote expression of active membrane proteins [13].
MeOH is indispensable when the P_{AOX1} is used. Since this C-source induces the proliferation of mitochondria and peroxisomes, it should be suitable for mitochondrial target protein expression. Alternatively a promoter system (e.g. the constitutive P_{GAP}) which does not rely on MeOH induction and works well in presence of glycerol can be considered.

2.3 Endoplasmic reticulum and golgi complex targeting

Targeting of proteins into the endoplasmic reticulum (ER) or golgi complex can be initiated by the introduction of a yeast specific C-terminal consensus sequence "HDEL". This has been used for the construction of engineered strains displaying a human like N-glycosylation machinery (glyco-engineered strains) [14]. The topic is evidently of high interest as it can be recognized by the lucrative acquisition of Gycofi Inc. (company responsible for the development of the method) by Merck & Co. The method includes the deletion of the endogenous glycosyltransferase gene, and the introduction of eukaryotic glycosidase or glycosyltransferase activities into a *P. pastoris* strain. Similar efforts were reported at the same time by the group of R. Contreras and N. Callewaert in Ghent who developed and employed the "Glycoswitch technology" or vectors (for more information see [15]). Employing such engineered strains target proteins which pass through the secretory pathway can receive a human like glycosylation pattern in contrast to the typical hyper-mannose type of yeasts. Engineering of these strains requires the introduction of several different eukaryotic activities, which slowed down the cell growth. Although this suggests that the cell is suffering stress related issues and that an efficient production of target protein is unlikely, several examples demonstrated that high amounts of target protein with humane like glycosylation can be obtained.
Recently, such a glyco-engineered strain in combination with surface display has been reported to be successful for the expression of four heterologous proteins in *P. pastoris*. It has been claimed that these strains can be universally used for screening of a number of target molecules [15].

These examples showed that ER/ golgi complex targeting of the proteins responsible glycosylation was successful. If there is no ER retention signal, the target proteins are usually dedicated to secretion in the extracellular space. Since also protein secretion involves the ER and Golgi, similar improvement strategies including over-expression of chaperons and foldases can help.

2.4 Membrane targeting

Membrane proteins are starting to receive high attention in the past years [16]. This can also be observed at important conferences in the area of protein expression such as PepTalks and PEGS, where a majority of the presentations of the past few meetings were focused on membrane protein expression. Currently only few is known about these proteins, maybe because of the difficult access to protein structures of insoluble domains, which was also caused by the lack of efficient microbial expression strategies so far. For the pharmaceutical industry a better understanding of membrane proteins (e.g. class A G-protein coupled receptors (GPCR), transporters and channel proteins) seems to be the main driving force for this field. New tools and strategies, like the transcriptome analysis of a *P. pastoris* strain expressing heterologous membrane proteins offer new opportunities in this field.

Recently, analysis of a *P. pastoris* strain expressing a recombinant ER-membrane protein by transcriptomics in our lab revealed the importance of increased expression of the endogenous transcription factor HAC1 for higher expression. Unexpectedly HAC1 induces the expansion of the ER but not other players of the unfolded protein response (UPR), which led to higher expression rates of the recombinant ER-membrane protein, probably due to more space in the membrane itself (Freigassner M. unpublished). There was no positive effect on expression of other proteins leaving the ER such as secreted proteins and proteins which were bound to the cytoplasmic membrane (Freigassner, manuscript in preparation).

2.5 Periplasmic targeting

Many proteins that are engineered for secretion into the media find their way through the secretory pathway, receiving the yeast typical glycosylation pattern, but at the end they are often retained in the periplasmic space. One explanation for this is the

size of the native (sometimes multimeric) protein molecules which hinders cell-wall trespassing (e.g. pig liver esterase in *P. pastoris* >180 kDA [17]). Periplasmic localization offers less space than other organelles like the ER or the golgi complex [18], but this proteins display typical eukaryotic posttranslational modifications. Usually this unwanted retention affects the applicability of the strategy since expensive and laborious down stream processing (DSP) is necessary. On the other hand a correctly processed and fully active enzyme is protected by the cell wall and a gentle chemical treatment can be used to reduce limitations caused by mass transfer issues. This kind of catalyst stabilization due to cell wall protection can offer an advantage for instable proteins compared to standard molecular displaying techniques [19]. Alternatively, to avoid chemical, "permeabilization" has also been pursued at the molecular level. This might also be an advantage in order to achieve more reproducible results. However, the necessary deletion of genes relevant for the cell wall structure usually results in a decreased growth as observed e.g. for *GAS1* deletion strains [20].

2.6 Peroxisomal targeting

Intracellular expression into the peroxisomes is an alternative where *P. pastoris* offers many advantages. Together with *Hansenula polymorpha* these yeasts are the major model organisms for basic research on peroxisomes, offering a perfect platform for exiting applications regarding recombinant expression and targeting into this organelle. Due to the large fraction of peroxisomes in methylotrophic yeasts and the natural efficiency of expression of some native enzymes in peroxisomes this targeting strategy merits a more precise overview.

2.6.1 Peroxisomes and peroxisomal targeting in *Pichia pastoris*

Peroxisomes play a crucial role in the assimilatory methanol utilization pathway in methylotrophic yeasts mainly due to the localization of the first two reaction steps (methanol $\rightarrow$ formaldehyde $\rightarrow$ dihydroxyacetone). It is well known that in *P. pastoris* when methanol is present as a sole C-source, induction of the P_{AOX1} [21] and proliferation of peroxisomes up to 80% of the cytoplasmic space occurs [22]. This also offers opportunities for strong transcription of potential target enzymes by using

the above mentioned advantages [23]. In addition peroxisomal expression can also be advantageous in situations, where the space and composition of peroxisomal membranes, neutralization of toxic intermediates (e.g. H_2O_2) or a native-like environment (e.g. redox-environment, pH) is needed.

These organelles have received extensive interest during the last years. They have no DNA and ribosome, and are unable to perform protein biosynthesis, consequently for protein acquisition peroxisomes relay on transport systems. At the moment at least two are well described, the Pex5p and Pex7p, that recognize the peroxisomal targeting sequence PTS1 and PTS2, respectively. PTS2 is a nonapeptide (RLX5(H/Q)L) localized in the N-terminus of proteins whereas the PTS1 is localized at the extreme C-terminus. Pex5p mediated transport is more frequently observed among the major peroxisomal proteins (e.g. enzymes of methanol metabolism [23]). The consensus sequence of PTS1 is described as serine-lysine-leucine (–SKL) or similar sequences, which associate with the C-terminus called the tetratricopeptide (TRP) region of the receptor, Pex5p [24]. Lately, studies about a *H. polymorpha* ∆pex4 and a ∆pex14 mutant strain in combination with Pex5p over-expression in the van der Klei research group brought new insights in the peroxisomal protein import machinery of this yeast.

Pex4p is a protein involved in the recycling of Pex5p from the peroxisome to the cytosol, and the absence of it in line with Pex5p over-expression leads to an accumulation of the transporter protein in the inner surface of the peroxisomal protein [25]. These results show the role of Pex4p and underline that *H. polymorpha* cells can tolerate up to 100% more of Pex5p compared to the wildtype strain, which was accounted to four additional copies of the gene (Southern blot analysis).

In ∆pex14 deletion strains normal peroxisomes were absent when methanol was used as a carbon source. Instead peroxisomal membrane remnants could be identified. Herein, the co-expression of *Hp*Pex5p also resulted in up to 100% more receptor protein than in the *H. polymorpha* wild type strain (western blot anaylsis). The over-expression *Hp*Pex5p in ∆pex14 mutants stimulated AOX and dihydroxyacetone synthase (DAS) transport in the mutant strain, but interestingly not of catalase (CAT) [26].

Important to mention is that the exact mechanism of peroxisomal transport is not completely clear and several new insights during the last years propose a more versatile transport mechanism, which is only temporarily possible in smaller

peroxisomes, before they become simple “enzyme bags” [27]. Transport has also been observed as PTS1 independent Pex5p transport, suggesting an additional conformational recognition for *H. polymorpha* alcohol oxidase (*AOX*), which relies also on FAD cofactor binding and additional helper proteins [28]. In order to elucidate the exact mechanism of AOX import a screening of mutants defective for AOX assembly was performed. Surprisingly an enzyme involved in the citric acid cycle namely *H. polymorpha* pyruvate carboxylase (Pyc) was identified as a helper protein candidate for the AOX peroxisomal import. Introduction of specific mutations in *Hp*Pyc influenced the AOX assembly but not the Pyc enzyme activity, suggesting an exclusive peroxisomal transport involvement as a chaperone and not as an enzyme [29-31].

Therefore a focus on peroxisomal targeting, combined with gene optimization might be promising to increase expression levels for such proteins. In this thesis I followed a systematic combination of natural/typical PTS1 and native/codon optimized genes and has been evaluated in regard of high level recombinant protein expression. The industrially important enzyme D-amino acid oxidase was used as a model enzyme.

2.6.2 Engineering of D-amino acid oxidase for expression and use in *Pichia pastoris*

D-amino acid oxidases (DAO, E.C. 1.4.3.3) are well characterized peroxisomal enzymes which have found many applications such as the first step of enzymatic hydrolysis of cephalosporin C to 7-aminocephalosporanic acid [32] and the production of α-keto acids [33, 34]. Many new application have been reported more recently [35-37] like the elimination of traces of D-amino acids from unnatural amino acid preparations [38].

The physiological role of this enzyme in eukaryotic cells can be described as a regulatory protein in processes like arterial hypertension, neural signalling and hormone secretion, while in microorganisms DAOs enable the cells to metabolize D-amino acids as sole source of carbon, nitrogen and energy [37].

DAO isolated from pig kidney and *Rhodothorula gracilis* (*Rg*DAO) gave access to the first crystal structures and thereby provided a basis for the mechanistic understanding of these flavoenzymes [33, 39, 40]. Nevertheless, a third DAO sequence isolated from the yeast *Trigonopsis variabilis* (*Tv*DAO) [41] showed better performance in terms of stability at higher temperature and specific activity for

hydrolysis of cephalosporin C into 7-(5-oxoadipoamido)-cephalosporanic acid [42, 43]. Also higher expression rates in the native host *T. variabilis* compared to other DAO producing yeasts [35, 44] made this DAO variant more interesting. Therefore a multi-ton scale per year process has been implemented for the cephalosporin C conversion by *Tv*DAO [32].

Analysis of protein sequences from *Rg*DAO (GenBank: AAB51107) and *Tv*DAO (Genbank: AAR98816) in terms of cellular targeting showed interesting results. While *Rg*DAO carries a well recognized PTS1 (–SKL) sequence which shows high prediction scores using publicly available targeting prediction software for peroxisomal targeting (PTS1 [45, 46], TargetP [47], SignalP [47], PSORT II [48], etc.), *Tv*DAO localization is not certain. Therefore, *Tv*DAO due to its rare PTS1, prolin-asparigine-leucine (-PNL) is a perfect model enzyme for the evaluation of high expression rates related with a peroxisomal transport enhancement. Noteworthy is that both proteins have a predicted secretion signal peptide. This leader sequences is regarded as an evolutionary vestige, since peroxisomes were observed to be generated form domains of the ER [49]. For an overview and comparison of *Rg*DAO and *Tv*DAO characteristics, see Table 1.

Table 1: Overview of characteristics relevant for *Tv*DAO and *Rg*DAO recombinant expression. Prediction data was calculated using public available tools such as SignalP, TargetP and PTS1. The *Tv*DAO and *Rg*DAO sequences are also public available (NCBI) with the accession number AAR98816 and AAB51107, respectively.

	*Tv*DAO	*Rg*DAO
Signal peptide for secretion [47]	Yes	Yes
Predicted cleavage site (of the signal peptide) [47]	17/18 aa	20/21 aa
PTS1	-PNL	-SKL
Peroxisomal targeting (fungi specific prediction function) [45, 46]	not targeted	targeted
Predicted targeting [47]	peroxisome/ER	peroxisome/ER
Heterologous expression	*E. coli* [50] , *S. pombe* [51], *P. pastoris* [52], *S. cerevisiae* and *K. lactis* [53],	*E. coli* [54]

The results obtained from this research work are summarized in the next chapter. The "stepwise engineering of a *Pichia pastoris* D-amino acid oxidase whole cell

catalyst" highlights the positive impact in terms of TvDAO expression in combination with a frequently used PTS1 (-SKL) in contrast to its rare –PNL C-terminal tripeptide.

2.6.3 Stepwise engineering of a *Pichia pastoris* D-amino acid oxidase whole cell catalyst

The manuscript describing these studies was submitted to the Journal Microbial Cell Factories and was judged as acceptable after revision.

Sandra Abad[1], Jozef Nahalka[1,2,3], Gabriele Bergler[1], S. Alison Arnold[4], Robert Speight[4], Ian Fotheringham[4], Bernd Nidetzky[1,2] and Anton Glieder[1§]

[1]Applied Biocatalysis Research Centre, c/o Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria
[2]Institute of Biotechnology and Biochemical Engineering, Graz University of Technology, Petersgasse 12, 8010 Graz, Austria
[3]Current address: Institute of Chemistry, Center of Glycomics, Slovak Academy of Sciences, Dúbravska cesta 9, SK-84538 Bratislava, Slovak Republic
[4]Ingenza Ltd., Wallace Building, Roslin BioCentre, Roslin, UK, EH25 9PP
[§]Corresponding author

Email addresses:

SA: sandra.abad@a-b.at
JN: jozef.nahalka@savba.sk
GB: gabriele.bergler@a-b.at
RS: r.speight@ingenza.com
AA : alison.arnold@ingenza.com
IF: ian.fotheringham@ingenza.com
BN: bernd.nidetzky@tugraz.at
AG: anton.glieder@a-b.at

2.6.3.1 Abstract

Background

Trigonopsis variabilis D-amino acid oxidase (*Tv*DAO) is a well characterized enzyme used for cephalosporin C conversion on industrial scale. However, the demands on the enzyme with respect to activity, operational stability and costs also vary with the field of application. Processes that use the soluble enzyme suffer from fast inactivation of *Tv*DAO while immobilized oxidase preparations raise issues related to expensive carriers and catalyst efficiency. Therefore, oxidase preparations that are more robust and active than those currently available would enable a much broader range of economically viable applications of this enzyme in fine chemical syntheses. A multi-step engineering approach was chosen here to develop a robust and highly active *Pichia pastoris* *Tv*DAO whole-cell biocatalyst.

Results

As compared to the native *T. variabilis* host, a more than seven-fold enhancement of the intracellular level of oxidase activity was achieved in *P. pastoris* through expression optimization by codon redesign as well as efficient subcellular targeting of the enzyme to peroxisomes. Multi copy integration further doubled expression and the specific activity of the whole cell catalyst. From a multicopy production strain, about 1.3×10^3 U/g wet cell weight (wcw) were derived by standard induction conditions feeding pure methanol. A fed-batch cultivation protocol using a mixture of methanol and glycerol in the induction phase attenuated the apparent toxicity of the recombinant oxidase to yield final biomass concentrations in the bioreactor of ≥ 200 g/L compared to only 117 g/L using the standard methanol feed. Permeabilization of *P. pastoris* using 10% isopropanol yielded a whole-cell enzyme preparation that showed 49% of the total available intracellular oxidase activity and was notably stabilized (by three times compared to a widely used *Tv*DAO expressing *Escherichia coli* strain) under conditions of D-methionine conversion using vigorous aeration.

Conclusions

Stepwise optimization using a multi-level engineering approach has delivered a new *P. pastoris* whole cell *Tv*DAO biocatalyst showing substantially enhanced specific activity and stability under operational conditions as compared to previously reported preparations of the enzyme. The production of the oxidase through fed-batch

bioreactor culture and subsequent cell permeabilization is high-yielding and efficient. Therefore this *P. pastoris* catalyst has been evaluated for industrial purposes.

2.6.3.2 Background

D-Amino acid oxidases (DAO, E.C. 1.4.3.3) are well characterized flavoenzymes that have been considered for various applications during the last 20 years. Among these applications, the industrial use of DAO to catalyze the first step of chemo-enzymatic conversion of cephalosporin C to 7-aminocephalosporanic acid [32] is an outstanding example. It represents the only known case of a large-scale (> 1000 tons/year) biocatalytic process employing an oxidase as biocatalyst (Figure 1A). Other promising applications of DAO include the production of α-keto acids [33, 34] , analytical determination of D-amino acids in biological samples and foodstuffs [35], the development of DAO biosensors [36] for analytics, and the elimination of traces of D-amino acids from non-natural L-amino acid preparations [38]. More recently, the use of DAO for the chemo-enzymatic synthesis of physiologically active compounds (e.g. L-methionine, phenyl pyruvate, L-6-hydroxynorleucine, L-2-naphtylalanine) has reinforced the applied potential of this enzyme [37]. Figure 1B is a schematic representation of a deracemisation process employing DAO in its first step and scaled up for industrial application [38].

DAO from the yeast *Trigonopsis variabilis* (*Tv*DAO) [41] was the preferred choice of catalyst for cephalosporin C conversion, primarily because it was more efficient on the industrial substrate than other known DAOs, such as enzymes from *Rhodotorula gracilis* and pig kidney for example [42, 43]. *T. variabilis* also showed higher productivity (volumetric and specific) for biosynthesis of DAO than relevant other native producers of the oxidase [43, 55].

The operational stability of *Tv*DAO in biocatalytic conversions has always been a critical issue in process development and has therefore attracted considerable attention. Enzymatic conversion of cephalosporin C depends on O_2 as co-substrate and furthermore relies on the H_2O_2 produced in the enzymatic reaction (Figure 1A). While isolated *Tv*DAO is sensitive to the conditions applied in the process (bubble aeration, high concentration of oxidants), carrier-bound and entrapped immobilisates of the enzyme showed enhanced persistence [56, 57]. However, demands on the enzyme with respect to activity and stability but also with respect to the costs incurred

vary with the process considered. For simple oxidations in fine chemical production, the H_2O_2 usually has to be removed to avoid enzyme inactivation. To our knowledge, a preparation of *Tv*DAO fulfilling the requirements of a broad-scope biocatalyst for fine chemical synthesis was not previously available.

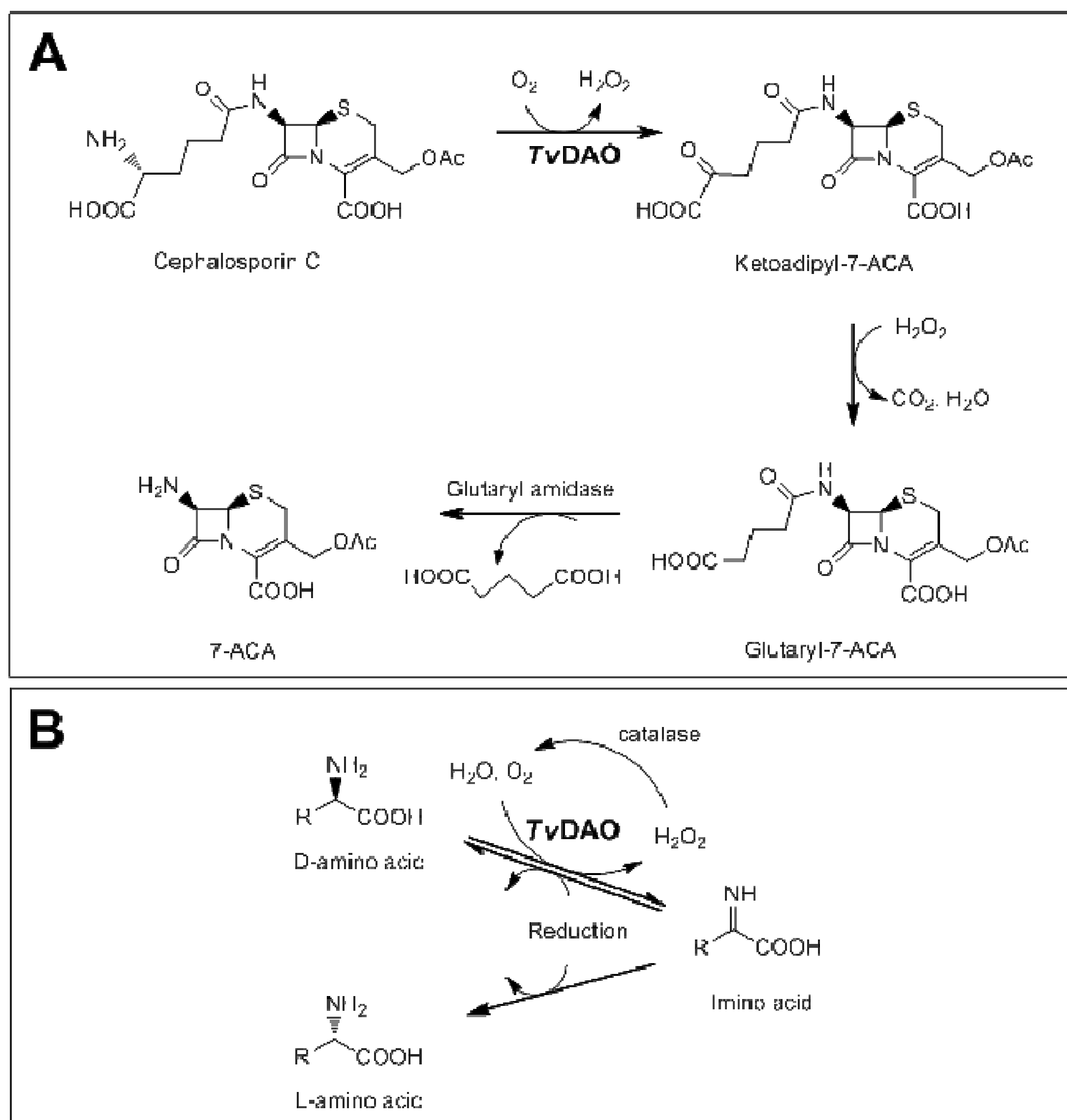

Figure 1 - Industrial applications of *Tv*DAO

A: Chemo-enzymatic conversion of cephalosporin C to 7-aminocephalosporanic acid. Spontaneous decarboxylation of ketoadipyl-7-ACA (aminocephalosporanic acid) is promoted by the H_2O_2 formed in the oxidase reaction of *Tv*DAO

B: Chemoenzymatic production of pure enantiomers of amino acids by amino acid oxidases in combination with imine reduction benefits from the additional presence of a catalase that destroys H_2O_2

While *Tv*DAO production in the native host presents industrial standard [55], enhancement of specific as well as volumetric productivity as compared to *T. variabilis* is also desirable (see later). Heterologous expression in *E. coli* has provided oxidase preparations having fine-tuned stability [57, 58], activity [59, 60] and even solubility [50]. A number of yeasts including *Kluveromyces lactis* [53], *Saccharomyces cerevisiae* [53], a catalase-deficient strain of *Schizosaccharomyces pombe* [51] and the methylotrophic yeast *P. pastoris* [52, 61] have also been employed for recombinant production of *Tv*DAO.
Contrary to cephalosporin C conversion where H_2O_2 formed in the reaction of *Tv*DAO is crucial to promote oxidative decarboxylation of the α-keto acid product (Figure 1A), processes for production of chiral amino acids (see Figure 1B) benefit from reduced by-product formation at low peroxide levels (see [62, 63]). A whole-cell biocatalyst producing both *Tv*DAO and catalase in large amounts would therefore be beneficial and the methylotrophic yeast *P. pastoris* seemed to be an ideal point of departure for its development. Heterologous gene expression in *P. pastoris* is often performed under control of the very strong *alcohol oxidase 1* promoter (P_{AOX1}) [21] which becomes fully induced when methanol is utilized as the sole carbon source. In addition to promoting recombinant protein production, these conditions also lead to an extensive proliferation of peroxisomes, which can cover up to 80% of the cytoplasmic space [22]. Efficient targeting of recombinant *Tv*DAO into peroxisomes of *P. pastoris* would allow to make efficient use of this intracellular space for the over-expressed protein and at the same time, co-localize the enzyme producing H_2O_2 with the one destroying it (i.e. the endogenous catalase [23]). Another plausible advantage of having *Tv*DAO and catalase in the same cellular compartment is that conversion of H_2O_2 by catalase recycles O_2 for the oxidase reaction. Previously described *P. pastoris* strains over-expressing the *Tv*DAO gene displayed only modest productivity [52, 61, 64]. Driven by a clear industrial demand (Figure 1B) and using a stepwise engineering approach, this paper reports on the successful design and generation of an innovative and notably improved *P. pastoris*-based DAO biocatalyst.

2.6.3.3 Results and Discussion

Expression engineering

Expression engineering to optimise oxidase production involved the re-design of both the gene and the amino acid sequence of *Tv*DAO. The coding gene was optimized for expression in *P. pastoris* according to the codon usage of genes that provide high levels of recombinant protein under methanol induction conditions for P_{AOX1} (see Material and methods). Like other DAOs, *Tv*DAO was assumed to be localized in the peroxisomes of the native host [53]. A targeting sequence recognized by either one of the common transport proteins Pex5p and Pex7p is a general requirement for peroxisomal protein import. A bioinformatic analysis of the *Tv*DAO primary structure using TargetP 1.1 [47] and PSORTII [48] revealed the absence of a Pex7p-selective motif PTS2. The occurrence of PTS1, the sequence recognized by Pex5p, was uncertain using publicly available predictor software [45, 46] but indicated the C-terminal peptide –Pro-Asn-Leu (PNL) as a putative PTS1. With the aim of enhancing peroxisomal transport of the enzyme under conditions of heterologous expression, we exchanged the native PNL motif by the well known peroxisomal targeting sequence –Ser-Lys-Leu (SKL) which has previously been successfully exploited to achieve peroxisomal localisation of recombinant green-fluorescent protein in *P. pastoris* [65].

Four different gene combinations were obtained by combining native and codon-optimized genes with the C-terminal sequence motifs SKL and PNL. This permutation was necessary because of the currently limited ability to predict the actual success of codon optimization. These DAO variants were combined with the necessary 5´and 3´ DNA sequences for expression cassettes delivering chromosomal integration in *P. pastoris* and expression under control of P_{AOX1}. *Pichia* transformants were cultivated in 96-well format [66] and screened for putative single copy strains using an enzyme-coupled assay for fast detection of DAO activity. This was done under the assumption that expression increases with increasing copy numbers of integrated expression cassettes. Only single copy strains allow the comparison of the effect of individual gene constructs. Quantitative real time-PCR was then used to analyze the selected strains, and the results confirmed that comparable number of expression cassettes (1-2 copies) had been integrated in each strain.

These "low-copy-number" strains were compared with respect to biomass formation and enzyme production in 1.5-liter bioreactor cultivations (Table 1). Among the four strains, the strain Tv1 harbouring the codon-optimized gene for production of *Tv*DAO-SKL showed the highest DAO activity. The exchange of the native putative peroxisomal targeting sequence -PNL to -SKL had no positive effect on enzyme production when the native gene was used (strains Tv5a and Tv1a). However, the same substitution brought a substantial increase in oxidase activity when comparing the codon-optimized genes (strains Tv5 and Tv1). The engineering strategies directed towards codon optimization, which is expected to facilitate translation and optimized subcellular targeting at the same time resulted in a synergistic effect on whole cell *Tv*DAO activity. Strain Tv1 showed an about 5-fold improvement in specific oxidase activity as compared to the reference strain Tv5a. To our knowledge, this is the first report describing a combination of PTS1 exchange with codon optimization of a gene to improve intracellular expression.

Table 1 - Enzyme and strain engineering for efficient expression of *Tv*DAO in *Pichia pastoris*

Enzyme activities are reported for *P. pastoris* cells obtained in 1.5-L bioreactor cultivation employing standard induction with 3mL/min methanol. Codon usage: native as in the original *Tv*DAO gene, or optimized for expression using the AOX1 promoter. *PTS1* – peroxisomal targeting using the native C-terminal tripeptide PNL or the engineered motif SKL. *CN*: Number of copies of the expression cassette integrated into the *P. pastoris* genome. wcw: wet cell weight.

Strain	Sequence		CN	Intracellular activity	Biomass
	Codon usage	PTS1		[U/g wcw]	[g wcw/batch]
Tv5a	native	PNL	1	137	220
Tv1a	native	SKL	1	90	274
Tv5	optimized	PNL	1-2	348	259
Tv1	optimized	SKL	1	767	144
Tv1a_mc	optimized	SKL	5	550	240
Tv1_mc	optimized	SKL	16-17	1283	117

All low copy strains except Tv1 which integrated a single copy of the expression cassette containing the codon optimized DAO gene with –SKL C-terminal end, allowed biomass concentrations well above 200 g/L using standard *P. pastoris* cultivation procedures including methanol feed in the production phase. A relatively lower biomass yield of 144 g wcw/batch for Tv1 employing the same cultivation

conditions might be explained by the stress, which is generated when large amount of recombinant DAO accumulate in the cell. The strain Tv1 grew normally in glycerol containing media where *Tv*DAO expression was repressed, but not during the induction phase using methanol. Similar observations were made with other yeast host organisms, where reduced growth during induction due to an apparent cell toxicity of recombinant DAO was reported [53, 61] .

Development of a high-yielding production strain

The codon-optimized gene for *Tv*DAO-SKL was used for further strain development. A new expression construct was made employing the expression vector pPpB1, which facilitates the generation and detection of *P. pastoris* multi-copy transformants. Using activity-based screening of *P. pastoris* transformants, the best active strain denoted Tv1_mc was selected. Characterization by real time-PCR revealed integration of about 16 – 17 copies of the expression cassette. Table 1 shows that the Tv1_mc strain gave a twofold enhancement of specific oxidase activity as compared to the single copy clone Tv1. This result implies that the number of gene copies did not translate linearly into an increasing titre of active *Tv*DAO. Notwithstanding, Tv1_mc is the most efficient host for *Tv*DAO production that has been reported so far. A comparison with earlier reports shows that this strain reached an expression level up to 1.5 and 1 order of magnitude above the expressions levels reported for *E. coli* and *P. pastoris*, respectively [50, 52]. Other yeasts such as *S. cerevisiae* and *K. lactis* showed significant lower *Tv*DAO production [53]. Unfortunately, the biomass yield for the multi copy transformant Tv1_mc was even lower than in case of the strain Tv1. As with Tv1, growth was impaired only during induction with methanol. For comparison reasons we constructed another multi-copy strain of *P. pastoris*, termed Tv1a_mc (native gene with -SKL). This strain produced a lower amount of *Tv*DAO activity (550 U/g wcw) than the strains Tv1 and Tv1_mc containing codon optimized genes and it grew normally in the methanol feeding phase. This finding further supports the notion that the recombinant oxidase, at least when it exceeds a certain level in the cell, seems to be inhibitory to *P. pastoris* growth. We therefore considered optimization of the cultivation and induction conditions for recombinant DAO protein production by the best strain Tv1_mc.

Using the methanol feed of 3 mL/h as a benchmark, alternative strategies for fed-batch cultivation of Tv1_mc were evaluated. It was known [67-70] that a mixed

substrate feed in the induction phase can support *P. pastoris* growth and productivity for the target protein at the same time. Figure 1 depicts the results obtained using feeds of glucose-methanol and glycerol-methanol (each at 3 mL/h) and a reduced pure methanol feed at 1.5 mL/h. The switch from pure methanol feed to mixed substrate feed caused an about 40% increase in biomass yield (~200 g wcw/batch) without compromising the specific activity of the yeast cells. Glycerol-methanol was superior to glucose-methanol at the evaluated feed rate. On the other hand a reduced flow rate (1.5 mL/h) of the methanol feed resulted in a 1.4-fold enhancement of specific activity as compared to the reference, while the biomass yield was still remained rather low (~140 g/batch).

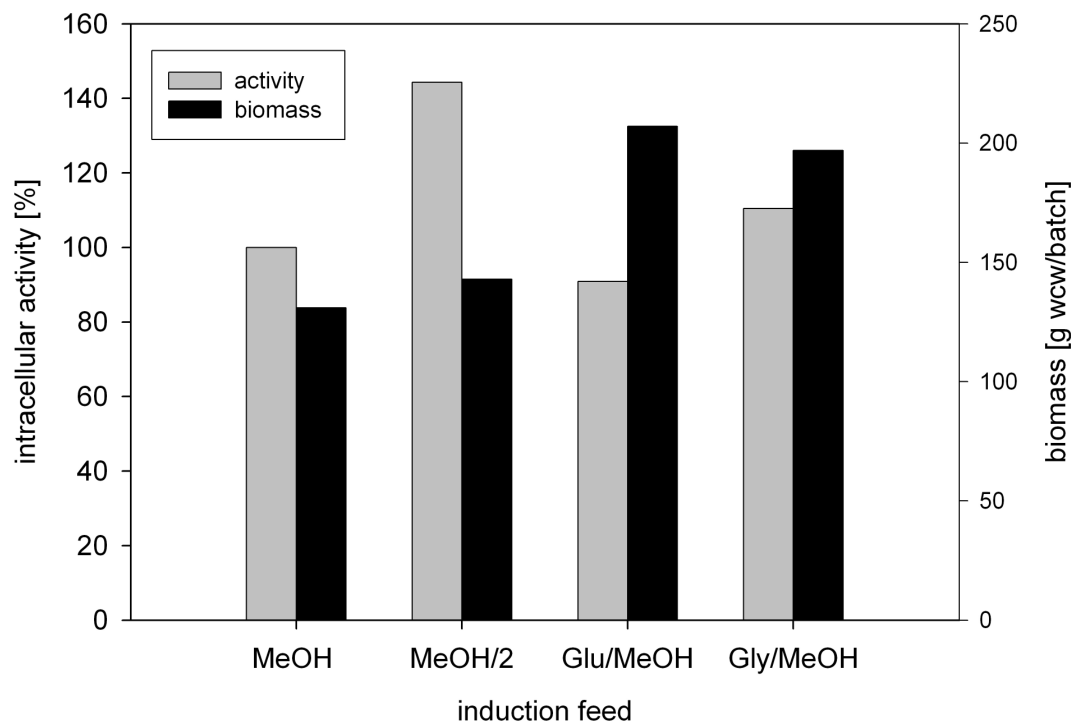

Figure 2 - High-cell-density bioreactor cultivation of *P. pastoris* strain Tv1_mc using induction with a mix-feed strategy.

Values are referenced to oxidase activity obtained using the standard methanol feed and set as 100%. The fed-batch induction phase lasted 90 h in each case. *MeOH*: 3 mL/h methanol. *MeOH/2*: 1.5 mL/h methanol, *Glu/MeOH*: 3mL/h glucose-methanol mixture, *Gly/MeOH*: 3 mL/h glycerol-methanol mixture.

However the volumetric productivity under these conditions was the highest. The Tv1_mc cells subjected to induction by a reduced methanol feed gave a volumetric oxidase yield of 218 kU/L and a productivity of 2.0 kU/(h L). In comparison the

Tv1_mc strain fed with glycerol-methanol resulted in lower yields of 167 kU/L (1.5 kU/hL), and 152 kU/L (1.4 kU/hL) for glucose-methanol induction. These values clearly exceeded those of other known *Tv*DAO expression strains as summarized in Table 2.

Table 2 - Comparison of relevant host strains used for production of *Tv*DAO

Differences in the conditions used for assaying the enzyme activity in different publications make direct and quantitative comparison of published data difficult. However, we measured a 2.9 fold lower specific activity of *Tv*DAO with D-alanine than with D-methionine.

Expression host	Activity	Reference
Trigonopsis variabilis [a]	4,620 U/L [d]	[55]
Pichia pastoris	7,596 U/L [e]	[61]
Pichia pastoris	23,000 U/L [d]	[52]
Pichia pastoris	12,532 U/L [d]	[64]
Pichia pastoris [b]	218,926 U/L [e]	this study
Saccharomyces cerevisiae	~110 U/g dcw [e]	[53]
Kluveromyces lactis	~150 U/g dcw [e]	[53]
Escherichia coli [c]	12,340 U/L [e]	[50]

[a] Induced with N-carbamoyl-D-alanine

[b] Induced with reduced methanol feed

[c] Additional D-methionine induction was reported to yield in a 4.5 fold expression improvement

[d] Activity measured with D-alanine as substrate

[e] Activity measured with D-methionine as substrate

[dcw] dryed cell weight

Whole cell biocatalyst optimization

Due to already described instability issues of the enzyme a cellular “encapsulation” of the oxidase seemed a promising option. Therefore intact cells obtained in a standard bioreactor cultivation of Tv1 (methanol feed of 3 mL/h) were subjected to stability examination. The determination was performed according to a previously reported assay that provides a first-order inactivation constant (k_{in}) as measure of the operational (in)stability of the oxidase activity [57]. The value of k_{in} is inversely proportional to stability, according to the relationship τ (half-life time) = $\ln 2/k_{in}$. Preliminary results showed an extraordinary half-life time of more than 10 h for the whole cell catalyst Tv1_mc. However, based on the same oxygen consumption measurements the specific oxidase activity of intact Tv1_mc cells (5 U/g wcw) was

less than 1% of the latent intracellular activity, presumably because of severe mass transfer limitations occurring in the intact cell system.

On the other hand the analysis of a lysate preparation after centrifugation was examined as the soluble fraction and the pellet fraction (probably containing partially lysed cells and peroxisomes) and showed the opposite results. The soluble fraction confirmed the high oxidase activity (~1200 U/g wcw) measured before but a high k_{in} value of 0.0208 min^{-1} (τ = 0.5 h). Because k_{in} for the partially lysed pellet fraction was only one-tenth of the value measured for the soluble fraction ($k_{in}^{pellet\ fraction}$ = 0.0026 min^{-1}; τ = 4.4 h) we still felt strongly encouraged to pursue the development of a whole-cell catalyst. *P. pastoris Tv*DAO-containing Tv1_mc cells served as the starting material. While cell permeabilization was a clear option to attenuate diffusional effects, it was a challenge to prevent a trade-off between enhancement of substrate availability and loss of enzyme stability in the permeabilized whole-cell catalyst. Among various protocols proposed for yeast cell permeabilization [71, 72], the treatment with 2-propanol appeared to be most suitable for our purpose because of its simplicity, efficiency and low costs.

Batches of Tv1_mc and Tv1a_mc cells (10 g each) containing about 1300 and 500 U/g wcw intracellular activity, respectively, were lyophilized prior to treatment with isopropanol considering that handling of dried biomass is easier and more reproducible than that of wet biomass. Key variables of the permeabilization (concentrations of isopropanol and biomass, incubation time) were examined systematically. Oxidase activity measured in the treated cells was related to the total intracellular activity (Table 3). Freeze-dried cells suspended in buffer were poorly active. However, their incubation in aqueous suspension caused the gradual “appearance” of enzyme activity, reaching ~20% of the maximum available activity after 20 min. Use of isopropanol further enhanced the apparent activity, yielding an effectiveness factor η (= apparent activity / total intracellular activity) of 0.48 (= 627 / 1300) under optimized conditions where 10% (by volume) of co-solvent was applied. A suitable incubation time was between 5.3 and 10.7 h. Concentrations of isopropanol greater than 10% caused a decrease in η, and incubation times longer than 5 h did not result in further improvement. Selected whole-cell preparations showing high η (> 0.40; Table 3) were analyzed for the stability of the oxidase activity under operational conditions. Values of k_{in} (hence, the stabilities) were not affected by the permeabilization as compared to the corresponding k_{in} ($k_{in}^{perm.\ cells}$= 0.0023

min^{-1}) reaching at least the stability of the Tv1_mc pellet fraction. Different biomass concentrations in the range 4 - 200 g/L were examined for permeabilization where the highest biomass level reflects the end concentration for standard fed-batch enzyme production in the bioreactor. The value of η obtained through permeabilization in the presence of 10% isopropanol was independent of the biomass concentration although the time required to achieve a maximum level of apparent activity increased with increasing biomass concentration (Figure 3).

Table 3 - Systematic optimization of process variables for the permeabilization of *P. pastoris* cells expressing *Tv*DAO

The accessible activity is the activity of the whole-cell system after permeabilization, as measured in U/g wcw and expressed as η, the effectiveness factor (= apparent activity / total intracellular activity), whereas the total intracellular activity was determined after cell disruption with a French Press. Strains Tv1a_mc and Tv1_mc contained 550 and 1283 U/g wcw of total intracellular activity, respectively. Lyophilized *Pichia* cells were used for permeabilization with a concentration being constant at 4 mg/mL.

Isopropanol concentration	Incubation time	Accessible activity after permeabilization			
		Tv1a_mc		Tv1_mc	
[%, by vol.]	[h]	[U/g wcw]	η	[U/g wcw]	η
-	-	6	~0.01	5	<0.01
0	0.3	100	0.20	321	0.23
4	0.3	118	0.23		
8	0.3	141	0.28		
10	0.0	100	0.20		
10	0.3	184	0.36	430	0.33
10	0.7	193	0.38		
10	1.5	206	0.41		
10	2.5			579	0.45
10	3.0	226	0.45		
10	5.3			627	0.49
10	10.7			621	0.48
12	0.3	176	0.35	424	0.33
15	0.3	170	0.34		
20	0.3	64	0.13		
40	0.3	14	0.03		

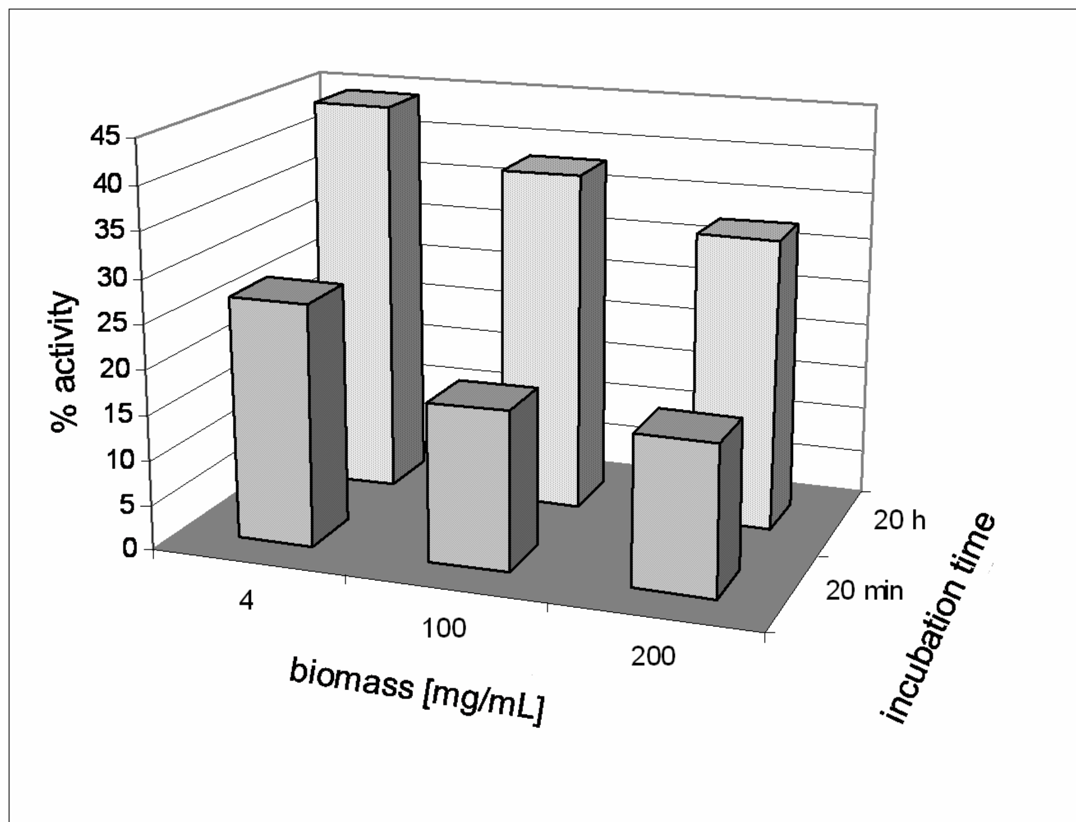

Figure 3 - Effect of biomass concentration on permeabilization of recombinant *Pichia pastoris* cells expressing *Tv*DAO.

Cells of Tv1_mc were treated with isopropanol for 20 min and 20 h. The activity of permeabilized cells is given as a percentage of the total intracellular activity (1283 U g/wcw) measured after cell lysis.

The stability of the whole-cell catalysts was not affected by the variation of biomass concentration during permeabilization. These results led to the conclusion that the permeabilization of Tv1_mc delivered a novel and promising whole-cell catalyst that was equally active (~500 U/g wcw) but three times more stable than the reported *E. coli* oxidase preparations [50]. The stability of *Tv*DAO in permeabilized lyophilized *P. pastoris* cells relative to the cell free *P. pastoris* preparation was enhanced by one order of magnitude.

2.6.3.4 Conclusions

A multi-level engineering approach was successfully applied to develop an innovative whole-cell enzyme preparation of *Tv*DAO that is suitable for applications in industrial deracemization processes (Scheme 1B). The outstanding capabilities of *P. pastoris* in recombinant protein production were harnessed to provide high specific activity

(1.3 kU/g wcw) and, due to high cell density cultivation of *P. pastoris,* also very high volumetric oxidase activities (150 kU/L) under standard induction conditions, which could be improved by a reduced methanol feed (218 kU/L). These values are well above the highest levels reported in literature (see Table 2). Mild permeabilization of the *Pichia* biomass allowed about half of the latent intracellular activity to be utilized, while at the same time retaining the good operational stability of the enzyme entrapped in the cell matrix of *P. pastoris* peroxisomes. Peroxisomal targeting might be a generally useful strategy of producing recombinant oxidases in *P. pastoris*, especially under circumstances where a whole-cell catalyst preparation is considered and immediate removal of the H_2O_2 generated in the enzymatic reaction is desired. In addition, when protein expression is induced by methanol induction, increased peroxisome formation provides increased capacity for the expressed and targeted oxidase. Due to these advantages a *P. pastoris*-based whole-cell *Tv*DAO preparation and deracemisation processes (Scheme 1B) are being scaled up and the new *Tv*DAO whole cell catalyst has been rapidly integrated into the existing process chemistry.

2.6.3.5 Methods

Chemicals and media

Oligonucleotides were purchased from IDT Integrated DNA Technologies BVBA (Leuven, Belgium) or Invitrogen Corp. (Carlsbad, CA, USA). Sterile water was from Fresenius Kabi Austria (Graz, Austria). All DNA-modifying enzymes were obtained from Fermentas GmbH (Burlington, Ontario, Canada). Unless otherwise stated, all chemicals were from Carl Roth GmbH (Karlsruhe, Germany), Becton, Dickinson and Company (Franklin Lakes, NJ, USA) or Sigma-Aldrich (St Louis, MO, USA).

Complex media contained 10 g/L yeast extract, 20 g/L peptone and 20 g/L glucose (YPD). Media for plates were solidified by addition of agar to 1.5% w/v. The minimal media used in this work contained 200 mM potassium phosphate buffer (pH 6.0), 13.4 g/L yeast nitrogen base and 0.0004 g/L D-biotin. They differed with respect to the carbon source concentration: 10 g/L glucose for BMD, 1 and 5% (by vol.) methanol for BMM2 and BMM10, respectively.

Cloning and engineering of the *Tv*DAO gene, and construction of *P. pastoris* expression cassettes

Escherichia coli TOP10F' (Invitrogen Corp.) was used as host for all cloning steps. The protein sequence [GenBank: AAR98816] of D-amino acid oxidase from *Trigonopsis variabilis* ATCC 10679 was employed for the design of a codon optimized gene for over-expression in *Pichia pastoris* under conditions of methanol induction, using the program Gene Designer (DNA2.0, Menlo Park, CA, USA). The optimized codon usage was calculated from three highly transcribed genes encoding enzymes of the *Pichia pastoris* methanol utilization pathway (alcohol oxidase 1 AOX1 [GenBank: U96967]; dihydoxyacetone synthase DAS1 [GenBank: FJ754551]; formaldehyde dehydrogenase FLD [Genbank: XP_002493270]) and from a plant gene (*Hevea brasiliensis* hydroxynitrile lyase *Hb*HNL [Genbank: U40402]) that is efficiently expressed in *P. pastoris* [6]. The resulting gene was synthesized by GenScript Corporation (Piscataway, NJ, USA). The native sequence for *Tv*DAO was kindly provided by Ingenza Ltd. (Roslin, UK). Exchange of the putative peroxisomal targeting sequence at the C-terminus of *Tv*DAO was carried out using a PCR employing a suitable modified reverse primer. Phusion^{TM} High-Fidelity DNA polymerase was used in this and all other PCR experiments applying a protocol provided by the supplier (Finnzymes Oy, Espoo, Finland).

Construction of *P. pastoris* expression cassettes

Products obtained from PCRs were purified with the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany), digested with *Spe*I and *Asc*I (native *Tv*DAO genes) or *EcoR*I and *Not*I (synthetic *Tv*DAO genes). The digested DNA fragments were ligated via *Spe*I/*Asc*I or *Eco*RI/*Not*I restriction sites into equally digested in-house *E. coli* - *P. pastoris* shuttle vectors pPpT2 or pPpB1 (TU Graz strain collection BT 5713 and 5709, respectively) that provide single or multi-copy chromosomal integration of target genes in *P. pastoris*, respectively. Relevant features and sequence elements of these shuttle vectors include: origin of replication of the *E. coli* plasmid pBR322 [73]; the *AOX1* promoter (P_{AOX1}) starting with a *Bgl*II site; a multiple cloning site with unique restriction sites for *Eco*RI, *Spe*I, *Asc*I and *Not*I; the *AOX1* transcription termination sequence; and an antibiotic resistance cassette consisting of a synthetic bacterial promoter called EM72 in tandem with a truncated version of the *P. pastoris* *ILV5* (*acetohydroxyacid reductoisomerase*) promoter, a synthetic gene coding for the

amino acid sequence of bleomycin (ble) conferring resistance against Zeocin from *Streptoalloteichus hindustanus*. The *ble* gene had been codon optimized for function in *E. coli* as well as *P. pastoris*. The pPpT2 shuttle vector further contained the *P. pastoris AOD* (*alternative oxidase*) transcription termination sequence [16] at the 3´ of the coding sequence of the synthetic *ble* gene. The pPpB1 vector had the same additional features except that the *Saccharomyces cerevisiae* ADH1 (alcohol dehydrogenase 1) promoter and terminator controlled the transcription of the *ble* gene.

Transformation of *P. pastoris*

A Mut^S strain derived from *P. pastoris* CBS 7435 (TU Graz strain collection number BT 3132) was used for transformation. A condensed protocol [74] was used in which the shuttle vectors were employed after linearization with *Bgl*II and purification with the QIAquick PCR Purification Kit (Qiagen). Transformations were performed in ice-cold electro-transformation cuvettes (0.2 cm, Cell Projects Ltd., Kent, UK) using pulse at 200 Ω, 25 µF and 1.5 kV. 0.5 mL of ice-cold sorbitol (1 M; in water) was added immediately after the electro shock, and the suspension was transferred to a sterile 12 mL polypropylene tube (Greiner, Frickenhausen, Germany). YPD medium (0.5 mL) was added to the tube which was then incubated for 2h at 30℃ using agitation at 60 rpm. After this regeneration step, aliquots (0.2 mL) were plated on solid selection medium containing 100 mg/L Zeocin.

Copy number determination by quantitative PCR

The number of copies of *Tv*DAO expression cassettes integrated into the *P. pastoris* genome was determined by quantitative real-time PCR as described in (Abad et al., submitted) employing the endogenous ARG4 gene as a reference. Power SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA) was used in an ABI PRISM 7300 Real Time PCR System (Applied Biosystems). The following oligonucleotide primers were applied in a concentration of 250 nM when using 2 ng genomic DNA of *P. pastoris* [75] as template: AOX1-fw-RT (gaagctgccctgtcttaaacctt)/AOX1-rv-RT (caaaagcttgtcaattggaacca) and ARG4-RT-fw (tcctccggtggcagttctt)/ARG4-RT-rv (tccattgactcccgttttgag). The temperature conditions were: 10min at 95℃; 40 cycles for 15 s at 95℃ and 60 s at 60℃ followed by a

dissociation step (15 s at 95℃, 30 s at 60℃, 15 s at 95℃) at the end of the last cycle.

Small-scale cultivation of *P. pastoris*

P. pastoris transformants were first cultivated in deep well plates (96-well format) using a modified procedure after Weis et al. [66] in a Multitron II stackable incubation system (Infors, Bottmingen, Switzerland). Briefly, following a 60 h long incubation in 250 µL BMD medium (320 rpm, 28℃, 80% air humidity), the cultures were induced by adding 250 µL BMM2. 50 µL BMM10 were additionally supplied after ~70 h as well as after ~84 h total cultivation time. After a 48 h-long induction phase (108 h of cultivation). 50 µL of each culture were transferred to V-bottom microtiter plates (Greiner Bio-One #651101) from which glycerol stocks were prepared. The remainder of the cell suspension was centrifuged (Eppendorf centrifuge 5810R: 3220 × g, 4℃, 10 min) and the pellets were collected for activity determinations. 300 µL Yeast Buster reagent (Novagen, Darmstadt, Germany) were added to the cell pellet which was resuspended and incubated for ~30 min at room temperature using agitation at 1400 rpm using a TITRAMAX 1000 shaker (Heidolph Instruments GmbH & Co.KG, Germany). Following removal of cell debris by centrifugation (Eppendorf centrifuge 5810R: 3220 × g, 4℃, 10 min), 10 µL of the supernatant were employed for measuring *Tv*DAO activity using the photometric assay as described below.

Bioreactor cultivation of *P. pastoris*

The inoculum (optical density OD_{600} 10 - 15) was prepared in two preculture steps in 250 mL baffled shake flasks using 50 mL BMGY (10 g/L yeast extract, 20 g/L peptone, 100 mM potassium phosphate buffer (pH 6.0), 13.4 g/L yeast nitrogen base and 10 g/L glycerol). Incubations were done for about 12 and 8 h at 28 ℃ using agitation at 120 rpm (Certomat BS-1). A 1.5-L fed batch-pro® bioreactor system (DASGIP AG, Juelich, Germany) equipped with a six-bladed Rushton turbine impeller and suitable controllers for pH and dissolved O_2 was used for biomass and enzyme production. The basal medium of Invitrogen Corp. was modified according to Hellwig et al. [76, 77] and contained 40 g/L glycerol, 0.17 g/L calcium sulphate dihydrate, 2.32 g/L magnesium sulphate heptahydrate, 2.86 g/L potassium sulphate, 7.18 g/L aqueous phosphoric acid (85%), 0.64 g/L potassium hydroxide, 0.22 g/L sodium chloride, 0.6 g/L EDTA disodium dihydrate, 0.20 mL/L antifoam (ACEPOL 83 E;

Lubrizol Additives, Wickliffe, Ohio, USA). PTM_1 trace element solution was added (4.35 mL/L medium) after in situ sterilisation of the basal medium. Its composition was as suggested by Invitrogen, namely 0.2 g/L biotin, 6 g/L $CuSO_4 \times 5\ H_2O$, 80 mg/L NaI, 3.067 g/L $MnSO_4 \times H_2O$, 0.2 g/L $Na_2MoO_4 \times 2\ H_2O$, 20 mg/L H_3BO_3, 0.916 g/L $CoCl_2 \times 6\ H_2O$, 20 g/L $ZnCl_2$, 65 g/L $FeSO_4 \times 7\ H_2O$, and 5 mL/L H_2SO_4. The pH of the final medium was adjusted to 6.0 using an ammonia solution (25% by volume; technical quality).

The cultivation started from an initial liquid volume of 0.65 L (in total) with 50 mL of the 2nd preculture in a batch phase (28℃, aeration at 0.7 L/min) in which glycerol served as the sole source of carbon. The stirrer speed varied between 500 and 1200 rpm as required to maintain a level of dissolved O_2 greater or equal 30% air saturation. The ammonia solution was used for pH control and likewise as nitrogen source. Following depletion of glycerol in the batch phase, typically after ~16 – 18 h, glycerol was fed from a 700 g/L substrate solution containing 12 mL/L PTM_1 solution. The feed flow rate increased exponentially over 6 h from 4.77 mL/h to 11.72 mL/h, afterwards gradually decreased to zero within 2 h. Decreasing glycerol supply was accompanied by the start of methanol feeding (supplemented with 12 mL/L PTM_1 solution). The flow rate was increased linearly to 3 mL/min within the 2 h where the glycerol feed was reduced and remained constant at this high level for another 90 h. Feeding of mix-substrates (methanol/glucose and methanol/glycerol) occurred in the same manner. Both mix fed contained 452 g/L methanol and 346 g/L glucose or 321 g/L glycerol, respectively and were supplemented with 12 mL/L PTM_1 solution.

Biomass was harvested using an Avanti J-20 XP centrifuge (Beckman Coulter, Krefeld, Germany) employing a JA-10 rotor at 2831 × g (10 min, 4℃). The pellet was washed once with a 100 mM potassium phosphate buffer (pH 6.0) and stored frozen (–20℃) until further use.

For cell lysis the thawed cell mass (~5 g) was suspended in 10 mL Tris buffer (100 mM, pH 7.5) and passed 2 times through an Aminco French press using an FA-030 cell (SLM Instruments, Rochester, NY, USA) at approx. 150 bar. Insoluble material was separated from the supernatant by centrifugation (20 min, 14000 rpm, 4℃) and further washed three times with 10 mL Tris buffer as above. Both the supernatant and the insoluble fraction were used for the measurement of *Tv*DAO activity.

Assays for oxidase activity and stability

An enzyme-coupled colorimetric assay for oxidase activity was used as described by Alexeeva et al. [78] with slight modifications. Briefly, 10 µL of appropriately diluted Yeast Buster cell lysate were transferred into a microtiter plate well. Assay solution (190 µL) was added and colour development (measured at 510 nm) monitored for 5 min at room temperature. The composition of the assay solution was 100 mM potassium phosphate buffer (pH 7.8), 0.5 mM 2,4,6-tribromo-3-hydroxybenzoic acid, 0.75 mM 4-aminoantipyrine, 10 mM D-methionine, and 0.025 mg/ml horseradish peroxidase (type VIa; Sigma-Aldrich catalogue number P6782).

A direct assay of *Tv*DAO activity used measurement of O_2 consumption at 30°C. A previously described glass mini-reactor with a working volume of 30 mL was employed [57]. The reactor was equipped with a fibre-optic oxygen micro-optode (PreSens GmbH, Regensburg, Germany), a temperature sensor, and a teflon sparging tube (1mm internal diameter) through which air O_2 was supplied. Mixing was achieved with a magnetic stirrer operated at 300 rpm.

Initial rates of O_2 conversion were recorded using 10 mM D-methionine as the substrate dissolved in 10 mM Tris buffer (pH 7.5). Reactions were started by adding 20 µL of appropriately diluted *P. pastoris* cell extract, obtained through either lysis or French press disruption of biomass, to 30 mL of reaction mixture. Alternatively, a suitable amount of untreated or permeabilized yeast cells was added. Note the standard protocol for cell lysis which involved mixing of 100 mg wet biomass with 400 µL Yeast Buster reagent followed by incubation of the suspension for 10 min at 4ºC. One unit of oxidase activity refers to 1µmol O_2 consumed/min under the conditions used.

A previously reported procedure was applied for the determination of "operational" stability of the different oxidase preparations [57]. Air was sparged into the mini-reactor at a flow rate of 30 L/h. The substrate solution contained 100 mM D-methionine dissolved in 100 mM Tris buffer (pH 7.5). The reaction was started by addition of a suitable amount of oxidase (cell extract, whole-cell preparation) such that the level of dissolved O_2 initially dropped to about 50 µM (~25% air saturation). The time course for $[O_2]$ was then recorded until the concentration of oxygen returned ~100% air saturation. Data were fitted with an exponential decay function to obtain an estimate for the first-order inactivation constant k_{in} [57].

Permeabilization

About 10 g wet *P. pastoris* biomass was suspended in 10 mL distilled water and freeze-dried over night using a Christ Alpha 1-4 LSC freeze dryer (Martin Christ Gefriertrocknungsanlagen, Osterode, Germany) operated at 0℃ and 0.570 mbar. Permeabilization by isopropanol was carried out in a total volume of 1 mL, using 4 mg dried cells unless otherwise stated. The alcohol concentration was varied in the range 2 – 40% (by volume, in water). Pure water served as reference. Incubation of the cells was done at 4℃ without agitation; the time was variable between 20 min and 50 h as indicated. 100 µL cell suspension was used to measure activity and stability.

Competing interests: The authors SA, JN, GB, BN and AG declare that they have no competing interests. RS is employed by Ingenza Ltd. who is interested in the commercialization of an industrial chemo-enzymatic process for amino acid deracemization.

Authors' contributions: SA and GB constructed the various *Tv*DAO expression cassettes and carried out *Pichia pastoris* strain construction. They performed the screening work and did the bioreactor cultivations. The screening assay was provided by Ingenza. SA also supervised the RT-PCR experiments and interpreted data concerning protein expression. JN performed cell permeabilization experiments and determined the activities and stabilities of whole-cell catalysts. IF, RS, BN and AG made substantial contributions to the conception, design and discussion of the overall project strategy and the experiments. SA, AG and BN wrote the paper. All authors have read and approved the final version of the manuscript.

Acknowledgements: The authors thank Thorsten Bachler for excellent technical support, and Claudia Ruth for providing *Pichia* expression vectors. The FFG, the Province of Styria, SFG and Ingenza Ltd. are acknowledged for financial support.

3. Innovative high throughput technologies

Nowadays, colony picking devices, high trough-put cultivation protocols [79] established spotting devices [80] or simply the available molecular tools for producing huge numbers of variants push the establishment of even more high throughput technologies. Although, high throughput methods have even lead to strain engineering up to the construction of a minimal yeast genome [81], still some research areas remained rather unaffected, as this is the case for directed evolution in higher eukaryotes, like *P. pastoris*.

The high production rates of certain proteins (e.g. *Tv*DAO) in *P. pastoris* reinforce the establishment of this expression host for the engineering of proteins that need correct organelle localization and/or posttranslational modifications. Following the main rule of directed evolution "you get what you screen for" [82], further protein engineering (e.g. directed evolution, site directed mutagenesis) should be performed in the same organism used for production, because using the so far established engineering host *Escherichia coli* as standard procedure can result in effects which are not reproducible in the final production host. As an alternative the generation of libraries can be performed in the bacterial host and screened on *P. pastoris*, but reduction of the diversity of libraries because of sub-cloning into shuttle plasmids and amplification of libraries in *E. coli* are some drawbacks that complicate the success of protein engineering.

Therefore we have designed, constructed and evaluated *P. pastoris* library platform strains. One main part of this chapter resumes the engineering of these strains for efficient mutant library screening, with main focus on high transformation rates, integration of shorter expression cassettes and reduction of false interpretation of measured activity improvements (see chapters 3.1/3.2).

Analysis of transformants or mutants libraries in terms of higher expression rate or selectivity can be done in a variety of assay formats. The most common ones are solid and liquid phase assays. One main drawback for the last one is that proper equipment (e.g. HT- pipetting roboter and HT-plate reader) is necessary. A review of enzyme assay used for high through-put screening is reviewed in [83].

Additionally, a number of established methods used to characterize proteins have been adapted in a high through-put format and are commercially available like Lab

Chip Separation microfluid separations (Caliper Life Sciences, Hopkinton, MA, USA) or enzyme-linked immunosorbent assay (ELISA) micro-array technologies.

More recently flow cytometry facilities like fluorescence-activated cell sorting (FACS), which allow library screening of up to >10^{10} mutants have been reported. At the moment it has been classified in at least four detection categories, which include cell surface display, GFP as activity reporter, product entrapment within the cells and in vitro compartmentalization (for details see [84]).

The characterization of the strain is not only limited to protein or enzyme analytics, also quantitative molecular characterization of the manipulated genome is an important issue, and therefore the second main focus of this chapter. The knowledge about the changes in the genome is essential for correct evaluation of experimental results and for the detailed characterization of production strains. These analyses are also needed in a medium to high throughput way. Because in *P. pastoris* no stable plasmids systems are available, integration of expression construct is a well established technique. But, after a transformation event, colonies harbouring different numbers of integrated expression cassettes can be isolated, which consequently deliver different activities/expression rates. In this manner caution should be taken when comparisons of different strains are done, since the observed results can be a copy number effect and not because of the performed engineering. Therefore RT-PCR was implemented for fast copy number characterization in *P. pastoris* to facilitate interpretation of experimental results (see chapters 3.3/3.3.1).

3.1 Construction of a ura$^-$ platform strain for library screening

Previous works made the demand of a new methodology for efficient library screening in *P. pastoris* evident. In our research group we have established efficient PCR-based expression cassettes for direct *Pichia pastoris* transformation [72, 85] that results in high transformation efficiencies (up to 10^4 transformants/µg DNA). These cassettes are assembled via overlap extension PCR (oePCR) with optimised universal linker sequences at each side of the target gene. The left arm (e.g. $P_{GAP,}$ or P_{AOX1}) and the right arm (e.g. *AOX1* TT and selection marker) can easily be assembled due to an established oePCR protocol. But the assembling of large oePCR (>4 kb) lowers the PCR yield and also the *P. pastoris* transformation efficiencies.

Truncation of the expression cassettes can solve this bottleneck. But, a 5´-end truncation (promoter region) would affect the expression, while a 3´-end truncation would result in an incomplete resistance gene, and growing under selection conditions (e.g. with Zeocin or auxotrophy) would not be possible.

Therefore a *Pichia pastoris* screening platform strain that facilitates the integration of truncated PCR-based linear expression cassettes was pursued. Also, important to mention is that integration cassettes usually deliver an undesired multi copy effect, which is a further drawback of the integration systems. In this case it is not clear if an expression improvement can be caused by several integrated cassettes or by a desired mutation and only a copy number determination or a retransformation of the recovered construct can elucidate the results. Therefore, this effect should also be attenuated by the construction of this platform strain.

In a first construction approach realized by Krammer *et al.* (Figure 1A) a knock out of the orotidine-5´-phosphate decarboxylase gene (*URA3*) from the wild type phenotype *P. pastoris* strain X-33 was pursued. In order to select a strain displaying an ura- phenotype a Geneticin (G418) selection marker was successfully used. Then the replacement of the *AOX1* gene was aimed with the insertion of an integration cassette carrying the P_{AOX1} and 3´AOX1 region at the 5´- and 3´-end, respectively (Figure 1B). The integration cassette also carried a *URA3* gene, which is non functional due to a deleted "Start" codon and two frameshift (located within the first 15 nucleotides of the latent coding sequence). Additionally for selection issues a functional Zeocin selection marker between the promoter region and the non functional *URA3* gene was inserted. This integration step in combination with the *AOX1* knock out is important since more reliable expression has been observed in deep well plate cultivations with strains displaying the resulting slow methanol utilization phenotype (Mut^S), and because the defect *URA3* genomic background was placed in the *AOX1* locus.

This engineered genomic constellation in the *AOX1* locus library screening platform strain enables the integration of 3´- end (and 5´-end) truncated expression cassettes. A wild type strain carrying the truncated expression cassette is not viable under uracil selecting conditions, as it is the case with the platform strain containing the non functional *URA3* gene. But, the combination of both due to homologous recombination would result in ura+ strains (Figure 1 C).

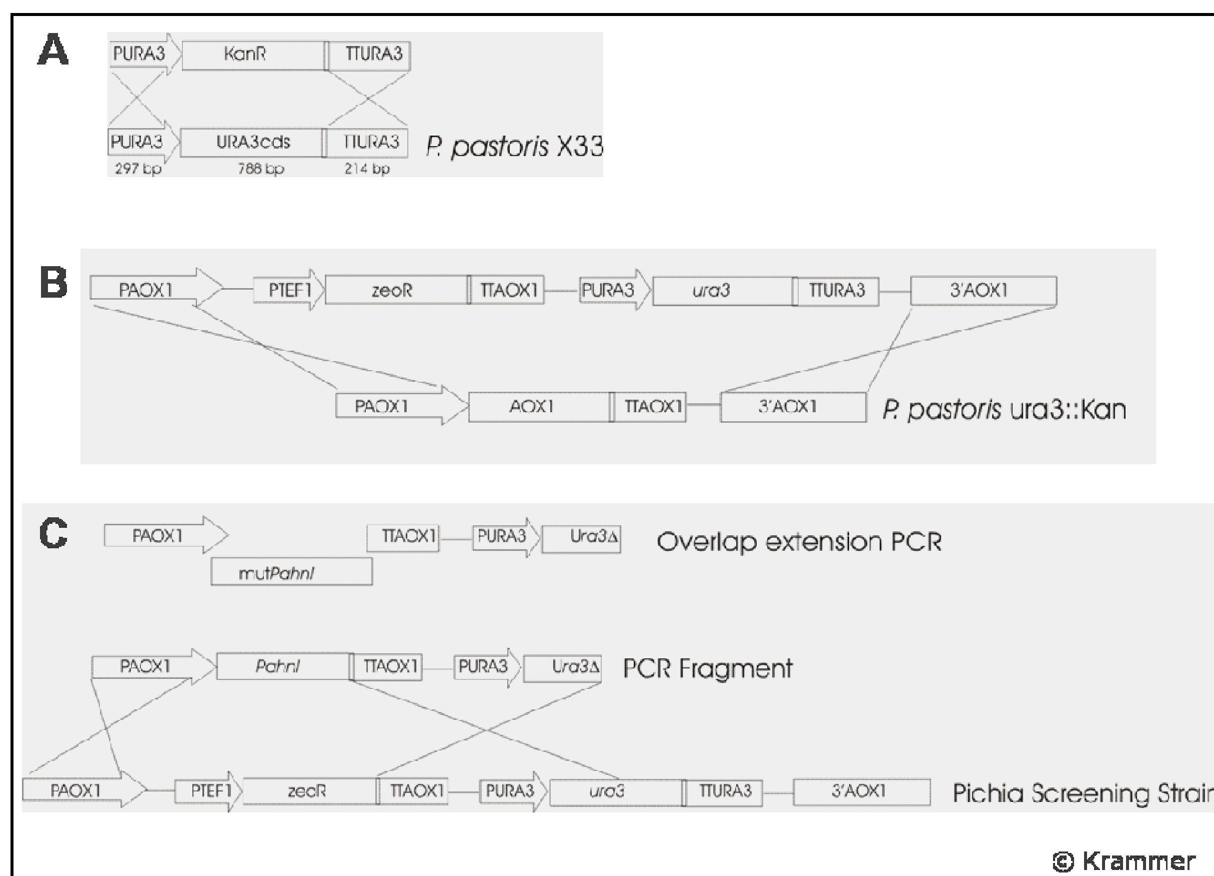

Figure 1: Strategy for the construction of a *P. pastoris* library screening platform strain. A: Knock-out of the *URA3* gene with a deletion cassettes containing a Kanamycin resistance. B: Replacement of the *AOX1* gene. And, insertion of a non- functional uracil cds with Zeocin as functional selection marker. C: Integration via double knock-out of a PCR created by error prone PCR and assembled via oePCR.

Noteworthy is that high transformation rates are a requisite for library screening and should be considered for the systematic truncation of the 5´- and 3´-end. After the identification of a correct 3´-end truncation, where selectivity and high transformation efficiency is feasible, the P_{AOX1} (0.9 kb) will be truncated without affecting the promoter activity. This is only possible if the expression cassette integrates into the correct *AOX1* locus via double knock-out. In this manner the P_{AOX1} as well as the ura selection marker would be fully functional again.
The uracil auxotrophy was chosen as genomic background of the platform strain due to its size (1.7 kb). This is in line with the main goal to reduce the size of the expression cassette for more efficient oePCR and better transformation rates. The alternative to work with an antibiotic selection marker like Zeocin (1.2 kb) due to its rather unspecific mutagenic effect seemed not as attractive as working with an

auxotroph marker. Unfortunately, the resulting uracil auxotroph strain displayed a slow cell growth, even when using complex media supplemented with uracil. This complicated the preparation of competent cells for the second transformation step. Nevertheless few transformants grew under Zeocin selecting conditions. But colony PCR demonstrated that the *AOX1* gene was not deleted. Due to this fact and the growth related issues this strategy was no longer pursued. But, these preliminary results were the starting point for another strategy in order to build an alternative library screening platform strain based on a very similar strategy but using a non functional Zeocin resistance instead of the *URA3* gene and are resumed in the next chapter.

3.2 *Pichia pastoris* library platform strain

This manuscript has been was prepared for publication in Journal of Microbiological Methods.

Sandra Abad, Franz S. Hartner, Anton Glieder [§].

Research Centre Applied Biocatalysis

c/o Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria

[§]Corresponding author

Correspondence: Dr. Anton Glieder

Petersgasse 14/II, 8010 Graz, Austria

email: anton.glieder@a-b.at

tel: +43 316 8734074

fax: +43 316 8739302

Keywords: *Ca*lB, *Hb*HNL, *Pichia pastoris* expression libraries.

3.2.1 Abstract

This manuscript describes the construction of a *P. pastoris* platform strain facilitating directed evolution employing this methylotrophic yeast. The described strain and method for protein engineering provides an alternative to *E. coli* based concepts in cases were eukaryotic microorganisms are the preferred host system for expression
The *P. pastoris alcohol oxidase I* (*AOX1*) locus was manipulated in order to carry a non functional (due to a defect translation start) but full length Zeocin resistance marker. This genomic background allows the integration of minimal expression cassettes, resulting in functional selection properties. The evaluation of the developed platform strain was performed with two reporter constructs, coding for two model enzymes, the secreted *Candida antartica* lipase B (*Ca*lB) and the intracellularly expressed *Hevea brasiliensis* hydroxynitrile lyase (*Hb*HNL), respectively. The method was evaluated by systematic integration of truncated versions of each reporter construct.

3.2.2 Introduction

During the past two decades *Pichia pastoris* has emerged as one of the most important expression hosts. Many examples of high yield expression by this methylotrophic yeast have been reported [2, 4]. Product yields between 10 and 30 g/L in bioreactor experiments have directed special attention to this organism. A small but well characterized and efficient molecular toolbox, which includes protein expression under the control of the strong *alcohol oxidase 1* promoter (P_{AOX1}) [21], the possibility for high cell density cultivations, posttranslational modifications and the broad and simple availability as an expression kit were mainly responsible for its success during the past years. Especially for proteins that demand typical eukaryotic posttranslational processing to obtain correctly folded, soluble and biologically active proteins *Pichia pastoris* is an attractive option. However, to avoid a possible bias from the expression system also the engineering of target proteins (e.g. directed evolution, site directed mutagenesis) has to be performed in the same organism. Otherwise engineering primarily leads to adaptations to the expression host which is used for engineering and consequently the obtained fitness improvements are not reproducible in the final

production host. Time consuming steps like sub-cloning into bacterial plasmids and amplification of libraries in *Escherichia coli* which also reduce the diversity of libraries were main drawbacks for protein engineering of more complex eukaryotic enzymes.

Recently, we reported on PCR-based expression constructs for *P. pastoris*, which enabled successful engineering of *Prunus amygdalus* hydroxynitrile lyase for the conversion of non natural substrates [72, 85]. In these studies, the linear expression cassettes including genes which were mutated by error prone PCR, saturation mutagenesis or recombined single gene mutants were assembled via overlap extension PCR (oePCR). Due to the introduction of universal linker regions (Linker 1 and Linker 2) between promoter and target gene as well as between target gene and terminator (TT)/selection marker the construction of the expression cassette was facilitated (Figure 1). Previous experiments have shown that these linker regions in the final construct do not affect expression (B. Pscheidt, unpublished results).

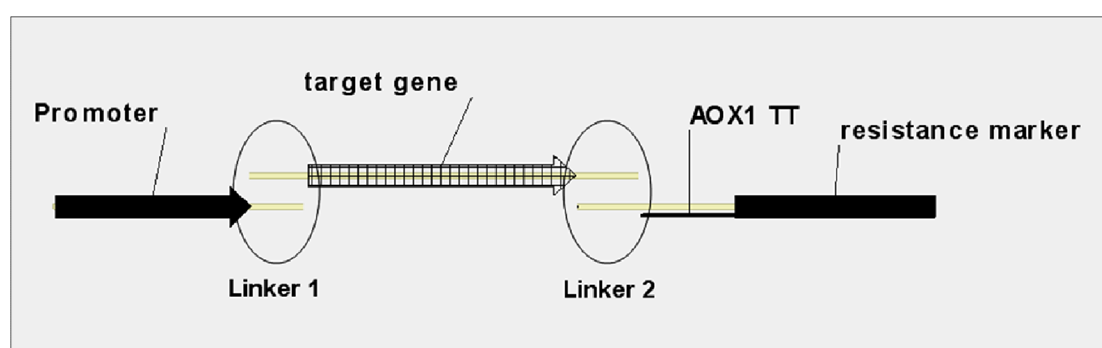

Figure 1: Expression cassette assembled via oePCR. Linker regions 1 and 2 facilitate the construction of the linear expression cassettes without affecting the expression of the target gene.

The obtained PCR product can be directly used for *Pichia pastoris* transformation, resulting in transformation efficiencies of 10^3 – 10^4 transformants/µg DNA. However, *P. pastoris* transformation efficiencies and the efficiency of oePCR assembling depend on the length of the final expression cassette. Therefore in this study we developed a *Pichia pastoris* platform strain, which was engineered for effective integration of shorter or even minimal expression cassettes, and shorter promoter regions (5´-end) and selection markers (3´-end) can facilitate the assembling of PCR-based linear expression cassettes. Herein, the *P. pastoris AOX1* locus was manipulated by the integration of a non functional Zeocin resistance marker. Consequently just a truncated version of 3´ -end of the PCR based expression cassette with an incomplete

but correct *Streptoalloteichus hindustanus* bleomycin (*ble*) gene, which codes for Zeocin resistance, is necessary for effective *Pichia pastoris* selection. Thus homologous recombination at the manipulated *AOX1* locus restores a functional gene for Zeocin resistance. In the same way the 5´-end (promoter region) truncation can be restored by the genomic background of this platform strain. For a first examination of this platform strain especially designed for screening two reporter constructs were tested. The secreted *Candida antartica* lipase B (*Ca*lB) and the intracellularly expressed *Hevea brasiliensis* hydroxynitrile lyase (*Hb*HNL) were used for the validation of the method. Systematic truncation of the 5´- and 3´- end of the expression cassettes was evaluated in terms of transformation efficiency and expression.

3.2.3 Results and discussion

Platform strain construction

Platform strain construction involved in a first step the construction of an integration cassette carrying a non functional Zeocin selection marker. Therefore the translation initiation site of the *ble* resistance gene from *Streptoalloteichus hindustanus* (*Sh*ble) was exchanged from ATG $\rightarrow$ TTG. Furthermore the deletion of one nucleotide (position: +10) lead to an additional “Stop” codon right after the original translation start. This non translatable *ble* gene was assembled between P_{AOX1}/*AOX1* TT and 3´AOX sequences as well as a functional KanMX6 cassette [86] for selection of *P. pastoris* transformants on Geneticin (G418) containing media. For the modification of the *AOX1* locus two alternative integration constructs were assembled. These constructs constitute the genomic background of the platform strain, which should later on facilitate the integration of linear expression cassettes. The first construct a KanMX6 cassette was placed between the terminator *Saccharomyces cerevisiae CyC1* TT (Zeocin terminator region of the pPICZ-plasmids of Invitrogen Corp.) and the 3´AOX1 region (Figure 2A, construct 1). For the left fragment of the construct, P_{AOX1} and *AOX1* TT were amplified in one step from pPICZ-B with 136 nucleotides (nt) spacer in between, which is the multiple cloning side (mcs) of the plasmid. In contrast to construct 1, in the second variant (Figure 2B, construct 2) the Kanamycin selection

marker was placed between the P_{AOX1} and the terminator *AOX1* TT. All integration cassettes were assembled by oePCR.

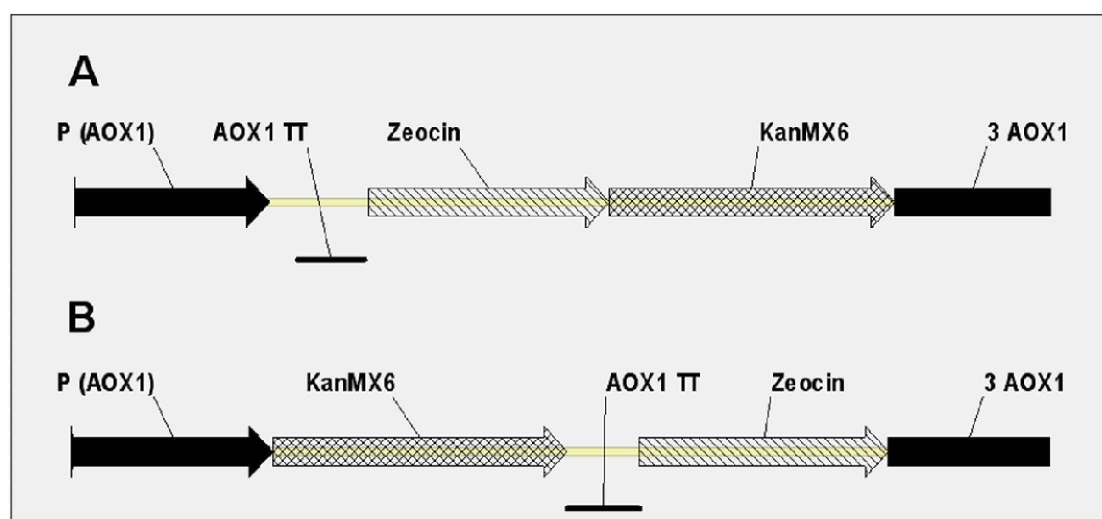

Figure 2: Integration constructs carrying the genomic background of the platform strain. Both constructs harbour a P_{AOX1}, AOX1 TT, 3´AOX region, a Kanamycin and a Zeocin resistance gene. The latter one has a frameshift at position +10 as well as no "Start" codon. A: Kanamycin resistance gene is placed between the non- functional Zeocin resistance and the 3´AOX region. B: Kanamycin resistance gene is placed between the P_{AOX1} and the *AOX1* TT.

Pichia pastoris X-33 transformation was successful for construct 1, but the same selection on standard YPD-G418 agar plates delivered no transformants for construct 2. We speculate that the KanMX6 *Ashbya gossypii* translation elongation factor 1α promoter (P_{TEF}) was co-regulated by P_{AOX1} and therefore repressed due to the constellation of construct 2. The inhibition was probably caused due to the proximity of the P_{AOX1} and its repression under the glucose conditions on the selection plates. This could be bypassed using MSM-G418 agar plates (1.5% agar, 1.34% yeast nitrogen base w/o amino acids, 4×10^{-5}% d-biotin, 2% sorbitol, 1% methanol, 300 mg/L G418). Under P_{AOX1} induced conditions the Kanamycin resistance gene appeared to be up-regulated since extraordinary high transformation efficiency could be observed.

The platform strain candidates were cultivated in a deep-well plate for characterization of the phenotype. Replica stamping on minimal dextrose and minimal methanol agar plates allows the identification of the methanol utilization phenotype. The deletion of the *alcohol oxidase 1* gene (or its destroyed functionality) results in a slow methanol utilization (Mut^S) phenotype. Previous experiences have shown that screening with Mut^S strains are more robust during deep well plate as well as bioreactor cultivations

than strains having a functional *AOX1* gene. Correct integration of the designed cassettes should result in a AOX1 knock-out and therefore a Mut^S phenotype. Analysis of the growth phenotype on methanol as sole carbon source plates revealed that only ~2% of examined transformants displayed Mut^S behaviour. Replica stamping on YPD-Zeocin showed that all analyzed transformants were unable to grow in presence of 100 mg/L Zeocin, and therefore recognized as Zeocin sensitive (Zeo^S).
Single copy integration of the non functional Zeocin selection cassette into the *AOX1* locus was desired, because several copies thereof in the genomic background of the platform strain can result in several undesired integrations of the expression cassettes containing whole or truncates *Sh*ble genes. In order to characterize the putative single copy platform strain candidates a Southern blot analysis was pursued. Two *Pichia pastoris* transformants from construct 1 or 2 denoted B2 and G3, respectively, and identified as Mut^S, Kan^R and Zeo^S phenotype were chosen for the analysis. *Nde*I which does not cut within the expression cassettes was used to digest genomic DNA from *P. pastoris* X-33, the platform strains B2 and G3 and two negative controls G4 and H4, which displayed a Mut^+ phenotype and were derived from construct 1 and 2, respectively. In combination with a digoxigenin (dig) labelled P_{AOX1} region as hybridization probe, the Southern blot (Figure 3) showed a single band around 9 kb for *P. pastoris* X-33, which is in agreement with the size prediction for the P_{AOX1} containing fragment from the now available *P. pastoris* genome sequence of *P. pastoris* GS115 [11]. Integration of our constructs in the *AOX1* locus would result in a ~4.7 kb larger band. Only one band for B2 and G3, respectively at ~ 12 kb confirmed the presence of a single copy of these constructs in the *AOX1* locus and at the same time the replacement of the *AOX1* gene and terminator (-2.3 kb).

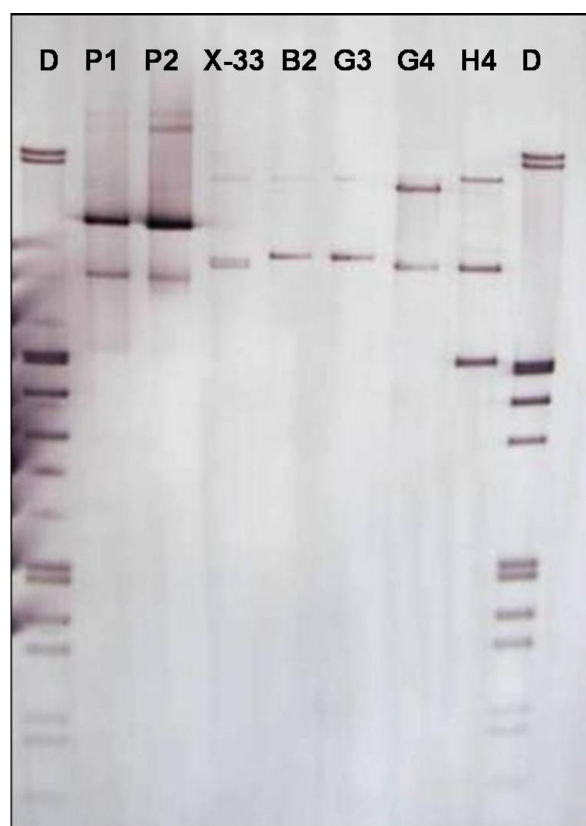

Figure 3: Southern blot analysis from the platform strains B2 and G3. A dig-labelled P_{AOX1} region was used for hybridization of the *Nde*I- digested genomic DNA. **D:** DNA Molecular Weight Marker III, digoxigenin-labelled (Roche Diagnostics GmbH). Fragments in marker 125, 564, 831, 947, 1375, 1584,1904, 2027, 3530, 4268, 4973, 5148, 21226 bp. **P1 and P2:** 1 and 10 ng of plasmid DNA carrying the P_{AOX1} (positive controls). **X-33:** *P. pastoris* starting strain for platform strain construction. **B2:** *P. pastoris* platform strain harbouring construct 1, and displaying Mut^S phenotype. **G3:** *P. pastoris* platform strain harbouring construct 2, and displaying Mut^S phenotype. **G4:** *P. pastoris* strain harbouring construct 1 and displaying Mut^+ phenotype. **H4:** *P. pastoris* strain harbouring construct 2 and displaying Mut^+ phenotype.

For further evaluation, platform strains, B2, G3 and X-33 were transformed with reporter constructs harbouring genes coding for the model enzymes, *Cal*B and *Hb*HNL, respectively. The main objective was to identify a truncation position of the expression cassettes that guarantees high transformation ratios and at the same time single copy expression for efficient mutant library screening. Therefore reporter expression cassettes with a correct but incomplete Zeocin resistance gene were amplified. If these cassettes are correctly integrated into the manipulated *AOX1* locus, which already harbours a non functional Zeocin gene, then the resulting strain should be able to grow on Zeocin selecting conditions. As a control, *P. pastoris* strain X-33

was transformed with the same truncated reporter cassettes. Growth of X-33 on Zeocin selecting media would show that the truncated *Sh*ble version still enables a functional Zeocin resistance under these conditions. In this case the undesired multi-copy effect can not be prevented by the newly designed genomic background of the platform strains. A combination of low DNA amounts (~0.5 µg) and low Zeocin concentrations (50 mg/L) also forces a reduced number of transformants with multi copy integration [85].

In total three approaches were followed. First, the two reporter constructs, *Ca*lB and *Hb*HNL under the control of the P_{AOX1} were amplified in three truncation versions, and a standard full length expression cassette (P_{AOX1}-*Cyc* TT) as control. The three truncations were selected in order to roughly cover the whole Zeocin coding sequence (cds) for narrowing down the search for an optimal truncation position. Second, the Zeocin truncation region was refined (~50 nucleotides/step) in order to provide selectivity (no growth for X-33, but growth for B2 and G3) but still deliver high transformation efficiencies. Third, 5´-end truncation is combined with the “best” 3´-end reduced cassette, resulting in a minimal expression cassette. The 5´-end truncation involves the promoter region in our case P_{AOX1}, which can be chosen in a way that no or limited expression is possible. A correct integration of this minimal expression cassette in the platform strain can result in expression under the promoter activity of the P_{AOX1} present in the manipulated *AOX1* locus and growth under Zeocin selecting conditions.

Rough truncation of the Zeocin selection marker

Transformation efficiencies are depicted in Table 1. Up to 2×10^3 cfu/µg DNA could be reached for the control transformation with the standard expression cassette length.

Position 374 truncation

The shortest 3´ truncation involves the removal of only the cytochrom C terminator region (*CYC* TT) of the Zeocin selection marker, leaving an intact *Sh*ble coding sequence and called 374 truncation (due to the Zeocin cds position). In this case all strains including X-33 resulted in lower transformants. The same reduction ratio in transformation efficiency in strains with different genomic background (e.g. X-33 vs. B2 or X-33 vs. G3), suggest that the first reduction of the homologous region (of about

300 nt) affect the integration efficiency due to the truncated and no longer effective expression of the Zeocin resistance.

Table 1: Transformation efficiencies reached with truncated expression constructs. Zeocin cds was amplified using primers designed to bind at position 38, 107 and 378.

		P. pastoris transformation efficiency [cfu/µg DNA]		
		X-33	B2	G3
*Hb*HNL	P_{AOX1}-Cyc TT	469	1699	681
	P_{AOX1}-374	88	1174	103
	P_{AOX1}-107	0	239	3
	P_{AOX1}-38	0	64	3
*Cal*B	P_{AOX1}-Cyc TT	570	2226	572
	P_{AOX1}-374	255	1209	250
	P_{AOX1}-107	2	187	5
	P_{AOX1}-38	1	23	2

Position 107 truncation

The following truncation of about 600 bp interrupts the Zeocin cds; therefore *P. pastoris* X-33 transformation with the minimized cassette should not confer Zeocin resistance under selective conditions. Surprisingly, this was not only the case for X-33, but also for G3. In contrast, positive B2 transformants with evident reduction in transformation efficiency could be observed.

Position 38 truncation

The shortest reporter cassette version leaves only 38 bp of the Zeocin cassette (truncation ~670 nt), enough for repairing the damaged Zeocin sequence in the platform strain, but a severe reduction of the transformation efficiency is evident for B2. Again no transformants were observed in the case of X-33 and G3.

In general every truncation resulted in a reduction of the transformation efficiency. Contrary to our previous observations in this study the length reduction of the expression cassette does not improve the transformation ratio in the *P. pastoris* strain tested. Probably, this is the case because the new strains only allow efficient selection for correctly integrated truncated expression cassettes into the manipulated *AOX1* locus. We assume that homologous recombination (or non-homologous recombination) in *P. pastoris* constitutes the major method's bottleneck and lead to these results. Probably the positive effect of longer homologous regions (due to the

genomic background) for efficient homologous recombination is outperformed by the selection of expression cassettes into the manipulated locus. Consequently a refined search for an efficient truncation region between position 107 and 374 was pursued in order to reach the highest possible transformation efficiency for our purpose.

Resulting transformants derived from all truncations were cultivated in deep well plates. *Ca*lB and *Hb*HNL activity measurements displayed varying expression levels, indicating a still existent multi-copy effect (data not shown). These results suggest that the platform strain genomic background is not sufficient for the elimination this effect.

Refined truncation of the Zeocin selection marker

The dramatic reduction of the transformation efficiencies from $10^3 \rightarrow 10$ transformants/µg DNA highlighted the demand for a readjustment of the 3´- end truncation. The main objective here was to obtain high transformation efficiencies in order to guaranty efficient mutant library screening. On the other hand the incomplete *Sh*ble should not be able to confer selectivity against Zeocin in X-33, but in B2 and G3.

Expression cassettes with truncations between position 107 and 374 in 50 nt steps have been amplified. For the fine tuned search for an optimal truncation position *Sh*ble cassettes harbouring *Hb*HNL and *Ca*lB were used as templates, and platform strains B2 and G3 as well as X-33 were transformed in duplicates.

Resulting transformation efficiencies are depicted in Table 2. The search for a compromise between transformation efficiency (high number of transformants for B2 and G3) still retained selectivity under selecting conditions (low number of transformants for X-33) resulted to be challenging. Consequently, two 3´-end truncations 300 and 358, which refer to the length of the *Sh*ble in the amplified construct, were selected as the best choice.

The measured *Ca*lB activity of the chosen constructs (P_{AOX1}-300/358) showed again the undesired multi-copy effect (Figure 4A). In this special case it is known that the expression of this lipase is interrupted by folding related issues in the endoplasmic reticulum. Therefore a multi-copy integrant displays lower expression levels than single or double integration strains (S. Abad, unpublished data). Unexpectedly, only 50% of the transformants harbouring the P_{AOX1}-300 expression construct were active.

Table 2: Systematic optimization of the 3´-end truncation. Amplification of the expression cassette via PCR was performed with primers that bind at different regions (50 bp/step) of the Zeocin cds. The transformations were performed in duplicate (I/II). *Cal*B and *Hb*HNL were used as model enzymes.

		P. pastoris transformation efficiency [cfu/µg DNA]					
		X-33		B2		G3	
		I	II	I	II	I	II
*Cal*B	P_{AOX1}-38	2	0	108	30	6	0
	P_{AOX1}-107	32	20	614	790	60	52
	P_{AOX1}-197	4	18	314	582	6	2
	P_{AOX1}-250	0	0	276	32	0	8
	P_{AOX1}-300	12	46	782	200	52	80
	P_{AOX1}-358	1508	1368	3336	1430	1786	678
	P_{AOX1}-374	5288	1076	2546	980	2126	994
	P_{AOX1}-Cyc TT	3616	3106	5288	2040	2470	1680
*Hb*HNL	P_{AOX1}-38	0	0	132	128	2	0
	P_{AOX1}-107	16	44	882	620	70	68
	P_{AOX1}-197	0	6	586	398	0	2
	P_{AOX1}-250	0	0	476	598	0	2
	P_{AOX1}-300	300	40	1270	304	26	6
	P_{AOX1}-358	32	970	1966	1926	528	322
	P_{AOX1}-374	1762	1794	3828	1180	2104	802
	P_{AOX1}-Cyc TT	7664	6248	*	6008	5226	7044

* high transformation efficiency.

5´-end truncation

In order to reduce the observed multi-copy effect expression cassettes with 5´-end truncations were designed. In this manner the expression cassette has no regulatory expression element (e.g. P_{AOX1}) that controls the target/reporter gene. Consequently a multi-copy integration would not result in a "multi-copy" expression, because only in case of correct integration into the manipulated *AOX1* locus a functional P_{AOX1} is reconstituted for expression in the genomic background of the platform strain. Important to elucidate is which 5´-end region still guarantees a correct integration and expression of the expression cassette.

Three different 5´-ends were chosen for this issue. The first truncation leaves 90 nt of the 3´-end of the P_{AOX1} intact, which are localized after the Δ9 deletion described by Hartner *et al.* [87]. This 90 bp long region is called Δ9* and lacks a transcription start, therefore expression of ~5% compared to the wild type promoter is expected. Two

more 5´-end reduced variants were also amplified. In these cases only the linker region that is routinely used for assembling the P_{AOX1} with the gene of interest via oePCR was appended to the 5´-end of the target gene (*Ca*lB and *Hb*HNL). The whole (28 nt) and the last 10 nt of the linker were amplified at the 5´-end of the expression construct, and called L1 and 10L1, respectively.

Systematic combination of the mentioned 5´- end and the two selected 3´-end truncations (position 300 and 358) were pursued. Only *Ca*lB constructs were amplified in combination with 358-end, and *P. pastoris* transformation was performed in duplicates. In every case a full length expression cassette, as well as the single truncated constructs were used as controls.

Results are depicted in Table 3. The control experiments displayed high efficiencies of 10^3-10^4 transformants/µg DNA. In line with results mentioned before the transformation efficiencies were reduced when incomplete expression construct were used, especially constructs with 300-end. Additionally, in this case selectivity for integration into the *AOX1* locus is evident, since X-33 shows no grow on Zeocin selecting conditions. Again, platform strain G3 shows similar behaviour as observed for control strain X-33. Results observed for B2, G3 and X-33 in combination with the 358-end expression constructs; suggest that this *Sh*ble length is still able to confer a Zeocin resistance since even X-33 transformants carrying this truncated expression cassette version are viable under selecting conditions. Therefore enhanced recombination into another locus than *AOX1* as well as multi copy integration should be taken into account. For all cases the truncation of the 5´-end did not affect the transformation efficiency.

Around 90 B2, G3 and X-33-transformants per construct were cultivated in a deep well plate as described before. B2 activity landscapes are depicted in Figure 4B/C. All cultivated transformants were replica stamped on YPD-Zeocin and YPD-G418 plates for characterization of the phenotype; especially for G3 the identification of single integrations by KanMX6 deletion (double cross-over) was possible. But less than 1% of all analyzed G3 strains were Kanamycin sensitive, suggesting that in general a deletion of the region between P_{AOX1} and *AOX1* TT by integration via homologous recombination of the 5´- and 3´-end region of the reporter expression cassette is unlikely. Integration mainly driven by homologous recombination of the 3´-region of the expression cassette would also explain the lack of effect of the 5´-end truncation on transformation efficiencies.

Table 3: Transformation efficiency reached with 5´- and 3´- end truncated expression cassettes. The transformations were performed in duplicate (I/II). *Cal*B and *Hb*HNL were used as model enzymes. Three different 5´- ends, namely Δ9*, L1 and 10L1 were combined with the 3´end truncations at position 300 and 358. As controls the whole length cassettes (P_{AOX1}-Cyc TT) as well as the single truncated versions (P_{AOX1}-300, P_{AOX1}-358, Δ9*-Cyc TT and L1-Cyc TT) were tested in parallel.

		P. pastoris transformation efficiency [cfu/µg DNA]					
		X-33		B2		G3	
		I	II	I	II	I	II
*Cal*B	Δ9*-300	6	8	290	632	0	0
	L1-300	2	0	206	384	0	0
	10L1-300	2	0	264	186	0	0
	P_{AOX1}-300	8	0	572	334	4	0
	P_{AOX1}-Cyc TT	1666	1834	912	1448	278	210
*Hb*HNL	Δ9*-300	0	0	184	510	0	0
	L1-300	0	0	236	150	2	0
	10L1-300	0	4	416	140	0	0
	P_{AOX1}-300	20	8	516	538	0	2
	P_{AOX1}-Cyc TT	2386	2078	2568	7140	2686	654
*Cal*B	Δ9*-358	778	800	1042	1298	1352	712
	L1-358	1536	784	1794	1210	1126	1168
	P_{AOX1}-358	2072	616	2704	1638	3324	3002
	Δ9*-Cyc TT	7498	5866	8356	11568	6606	5202
	L1-Cyc TT	4870	5942	4958	5996	7882	6524
	P_{AOX1}-Cyc TT	10652	7414	7776	7256	8722	7338

L1 and 10L1 truncations

Activity measurement showed that for L1/10L1 5´-end truncations in combination with any 3´-end no detectable expression took place. This result is unexpected since also for L1/10L1-300 cassettes, where at least one integration into the manipulated *AOX1* locus is prerequisite for selection no activity was detected. Apparently, a P_{AOX1}-mcs-L1 reporter gene constellation acts as a spacer (mcs+L1=164 nt) between P_{AOX1} and the reporter gene interrupting the transcription start and no expression of *Cal*B is possible.

Δ9* truncation

The Δ9*-truncation showed a reduced number of functional expressing transformants. The activity landscapes showed that only ~30% of transformants harbouring the minimal expression cassette were positive in terms of expression.

Although putative multi copy effect is possible in combination with *Cyc*-TT and 358 3´-end expression cassettes, only for the first construct resulted in several *Cal*B expression level, while from Δ9*-358 activity landscape only one or two rather expression levels could be identified. Suggesting, that several expression cassettes are present in B2, resulting in the already mentioned bottlenecks during the secretion pathway. This was forced by the chosen selecting conditions in combination with a non-efficient expression of the Zeocin resistance caused by the truncation of the gene, and the non homologous recombination into the genome.

In agreement with our assumption that the 300-end selects the integration into the manipulated *AOX1* locus and acquires one fully functional Zeocin reistance, also only one expression level (plateau) can be recognized in transformants harbouring the Δ9*-300 minimal expression cassette. Interestingly, the expression level is not as high as expected with one copy of the construct under the control of the wild type P_{AOX1} promoter. Suggesting that the constellation P_{AOX1}-mcs-Δ9*-reporter gene results in a reduced expression especially compared with previously observed results (e.g. P_{AOX1}-*Cyc* TT).

Further interpretations are not possible since no real-time PCR analysis for copy number determination of integrated expression cassettes was performed. The rather low transformation efficiencies reached with the minimal expression cassettes (e.g. Δ9*-300) in combination with only ~30% functional expressing transformants did not meat the main goal of the platform strain.

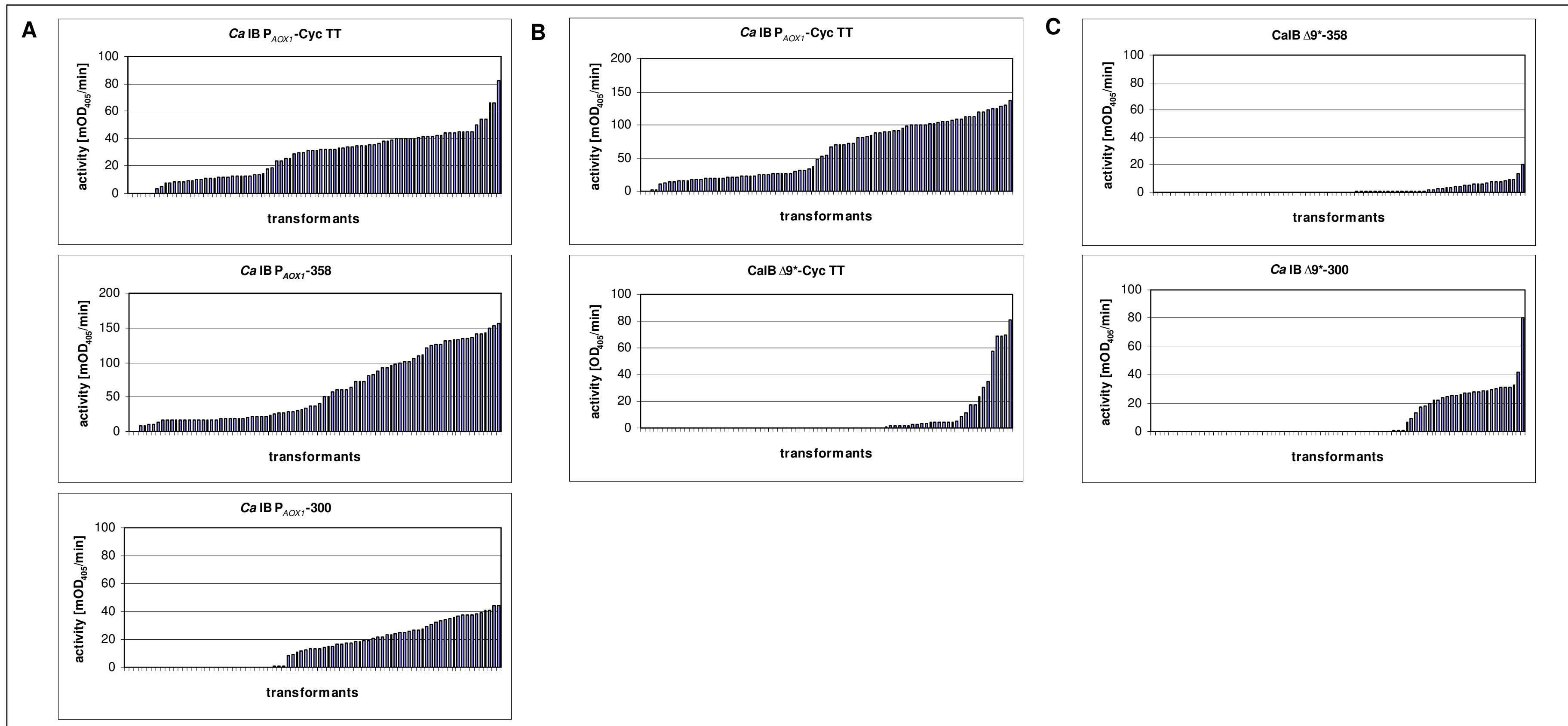

Figure 4: Activity landscapes from B2 platform strain expressing *Ca*lB. Comparison of effect caused by minimal expression cassettes (and full length constructs as controls). **A:** 3´-end truncations of the Zeocin cds at position 300 and 358. **B:** 5´-end truncations of linear expression cassettes. Δ9*-and L1-truncations (as well as P_{AOX1} as control) combined with a complete and functional Zeocin resistance gene. **C:** Minimal expression cassettes (5´- and 3´- end truncations). Δ9*-and L1 combines with 358 and Δ9*-300 3´- end truncation. L1-300 (~90 transformants) resulted in inactive transformants.

3.2.4 Conclusions

A *P. pastoris* platform strain, designed for efficient mutant library screening has been constructed. Especial focus for the design was set on a facilitated interpretation of the experimental results by generation of single copy integrations and minimal expression cassettes (5´ and 3´-end truncations). Therefore the *AOX1* locus has been successfully manipulated by insertion of a genomic background containing a non functional Zeocin coding sequence, which in combination of shorter expression cassettes provides a functional Zeocin resistance in the platform strain.

In general all truncations delivered reduced transformation efficiencies, because only correctly integration of expression constructs into the manipulated *AOX1* locus is viable under selecting conditions, and suggests that the non-homologous recombination machinery plays a major role during integrations event in *P. pastoris*.

Nevertheless integration of the minimal expression cassettes into the new platform strain was possible. Expression cassette 3´-end truncation of the *Sh*ble gene at position 358 results in reduced Zeocin resistance, whereas truncation at position 300 is no longer functional and therefore both lengths were chosen for further 5´end truncation. L1 and 10L1 5´-ends in combination with 300/358 and *Cyc1* TT resulted in no activity, suggesting that the 160 nt spacer between P_{AOX1} and reporter gene destroyed promoter's activity. Whereas P_{AOX1}-mcs-Δ9*-reporter gene constellation showed a reduced expression of *Ca*lB in comparison to a standard P_{AOX1}- reporter gene constellation, but resulted in one expression level for the construct Δ9*-300.

The application of the new platform strain also constitutes a possibility for more efficient oePCR, since smaller constructs have to be assembled, at the same time also avoiding the particularly difficult terminator region during PCR amplification of such long oePCR products. A minimal expression cassette, displaying a total size reduction of 1250 bp (e.g. Δ9*-300), which can be constructed in a two fragments/one step oePCR resulted in reduced transformation efficiencies (10^2 transformants/µg DNA) and only one third of the screened transformant were active.

During this study, we observed that the deletion of genes via double knock out is only possible for 1-2% of the analyzed transformants, although homologous region up to 1kb (left arm) and 1.5 kb (right arm) were used.

3.2.5 Materials and methods

Chemicals and materials

Sterile water was purchased from Fresenius Kabi Austria (Graz, Austria). Unless otherwise stated, all chemicals were purchased from Carl Roth GmbH (Karlsruhe, Germany), Becton, Dickinson and Company (Franklin Lakes, NJ, USA) and Sigma-Aldrich (St Louis, MO, USA). DNA modifying enzymes were supplied by Fermentas GmbH (Burlington, Ontario, Canada). All PCRs were performed with Phusion HF according to the producer's manual (Finnzymes, Finland).

KanMX6/ZeoD integration cassette

P_{AOX1}, AOX1 TT as well as the Zeocin cassette were amplified from vector pPICZ-B (Invitrogen, Carlsbad, USA), 3´AOX region was amplified form pHILDII (Invitrogen) and KanMX6 from PFAa-kanMX6 [86]. All PCR products were purified by agarose gel electrophoresis prior to overlap extension PCR [85].

The obtained PCR products displayed 4.6 and 4.7 kb and were purified by agarose gel electrophoresis and gel extraction using Wizard® SV Gel and PCR Clean-Up system (Promega, Germany) and ligated into pTOPO-blunt II (Invitrogen, Carlsbad, USA) and named KanMX6/ZeoD and ZeoD/KanMX6, respectively. The correct sequence, especially the Zeocin replacement of the start codon replacement and the introduction of a frame shift was verified by sequencing.

Ca**lB and *Hb*****HNL reporter expression cassette**

Codon optimized *Ca*lB (GenBank: ACIO6118) fused with *Saccharomyces cerevisiae* α-mating factor leader sequence was assembled via oePCR as described before. The resulting gene was *Eco*RI/*Not*I digested, gel purified and cloned into an equally digested pPICZ-B plasmid. The resulting construct was analyzed by sequencing, and amplified via PCR for *P. pastoris* transformation. pSHNL-AOX1 described in [6] and modified in [88] was used as template for PCR amplification. Purified PCR product was subjected for *Pichia pastoris* transformation.

After transformation the cell suspensions were plated on YPD-Zeocin agar plates (Zeocin: 50 and 100 mg/L) and incubated at 28℃ for 2 days. Transformants were cultivated in deep-well plates as described in [79], and screened for *Ca*lB activity.

Table 4: Primers for amplification of reporter expression cassette

Name	Sequence
AOX-F	5´-GATCTAACATCCAAAGACGAAAGG-3´
d9-P(AOX1)-F:	5´-ACTGGTTCCAATTGACAAGC-3
L1-CalB-F:	5´-CTAGGTACTTCGAAACGAGGACTTCACGATGAGATTTCCTTCAATTTTTAC-3´
10L1-CalB-F:	5´-AGGACTTCACG-ATGAGATTTCCTTCAATTTTTACTGCTGT-3´
Zeo38-R	5´-CGCGGTGAGCACCGGAAC-3´
Zeo107-R	5´-GTCCTCCACGAAGTCCCGG-3´
Zeo197-R	5´-ACCCAGGCCAGGGTGTTGT-3´
Zeo250-R	5´-TGGACACGACCTCCGACCAC-3´
Zeo300-R	5´-CTGCTCGCCGATCTCGGTC-3´
Zeo358-R	5´-CCACGAAGTGCACGCAGTTG-3´
Zeo374-R	5´-TCAGTCCTGCTCCTCGGC-3´
CYC1 TT-R	5´-GATCTTGAGATAAATTTCACGTTTAAAATC-3´

P. pastoris transformation

The integration cassettes were amplified by PCR and purified by agarose gel electrophoresis. P. pastoris transformation was performed as described in [74]. After a 2h regeneration phase, aliquots were spread on YPD-G418 (300 mg/L) agar plates and incubated for 2-3 days at 30℃.

Analysis of methanol utilization phenotype

For analysis of the methanol utilization phenotype, *Pichia pastoris* 1 day microscale cultures [79] were transferred on MM and MD agar plates using a 96-pin replicator and incubated at 30℃ for 2 days. Colonies which appear in both plates were depicted as Mut^+ while strains growing only on MD as Mut^S.

Southern blot analysis

Genomic DNA from *Pichia pastoris* X-33, D2, G3, G4, and H4 was isolated employing the Easy-DNA™-Kit from Invitrogen. 10 µg from the gDNA was digested with *Nde*I over 12 hours and separated in 0.8% TAE agarose gel. In a sandwich construction the DNA was transferred through capillary action to a positively charged nylon membrane (Roche Diagnostics GmbH, Vienna Austria) following the method by Southern [89]. DIG labelled P_{AOX1} was used in the hybridization solution to estimate KanMX6/ZeoD integration cassettes, and to confirm methanol utilization phenotype. Colorimetric detection was performed using anti-DIG-alkaline-phosphatase and the substrate NBT/BCIP (Roche Diagnostics GmbH).

***Ca*lB activity assay**

*Ca*lB/esterase activity was performed according to [90] with minor modifications. Briefly, 20 µl of the supernatant or a dilution of it was pipetted into a microtiter plate well. The *Ca*lB activity was measured as colour development at 405 nm for 5min at room temperature after addition of 180 µL of the assay solution. The assay solution consisted of 300 mM Tris/HCl pH7, 1% DMSO and 4 mM of *p*-nitrophenyl butyrate.

***Hb*HNL activity assay**

The HNL activity was determined essentially as described earlier [79], with minor modification. Briefly, 100 µL culture were centrifuged and the pellet was lysed with 50 µL Y-PER reagent (Thermo Fisher Scientific Inc., Rockford, Illinois, USA) for 30 minutes at 25℃. After centrifugation, 20 µL of supernatant were diluted 1:5 with 5 mM potassium phosphate buffer, pH 6.5. 50 µL of the diluted sample were taken for the activity assay, the reaction was followed after addition of the substrate solution.

Acknowledgements: The authors acknowledge Dr. Barbara Petschacher for carefully revision of the manuscript, and to FFG, the Province of Styria and SFG for financial support.

Conflict of Interest statement: The authors have declared no conflict of interest.

3.3 Gene dosage determination in *Pichia pastoris*

For this characterization Southern blot analysis is a routine method along the *Pichia* community. But, although it can deliver the necessary information it still is a time-consuming and laborious method and requires large amounts of genomic DNA. Lately, Southern blot analysis has been used to validate transcriptomics based strain engineering [7]. Herein, Gasser *et al.* identified a number of *P. pastoris* genes that have a positive effect during heterologous protein production at certain cell stress situations (e.g. recombinant expression). For the characterization of the effect the construction of strains harbouring one expression cassette with the gene responsible

for secretion enhancement was pursued. The copy number characterization was performed via Southern analysis. In this case valuable information obtained via transcriptomics analysis is slowed down by the validation of results, which depends on the analysis of single copy strains. This is one example that reflects the necessity of a faster method for copy number determination of the increasing number of strains. Recognizing this lack of a fast method we have optimized a gene dosage method based on real time PCR for *P. pastoris*. Although some reports summarizing the method for plants and cell cultures were available, a detailed protocol for yeasts was not established yet. A procedure for copy number determination in *P. pastoris* has been summarized in the manuscript "Real time PCR based determination of gene copy numbers in *Pichia pastoris*", which can be found in the next chapter. The already achieved impact of this method during the past two years is shortly overviewed.

Discovery of new promoters and their analysis is one of the application fields were a fast determination of integrated copy numbers is needed. Recently, Hartner *et al.* described the results of a P_{AOX1} deletion library combined with GFP as reporter protein [88]. One deletion variant called $P_{AOX1}\Delta 6^*$ in comparison with the wild type promoter revealed the influence of this deleted region as a transcription element responsible for the inhibition of the promoter under glucose de-repressing conditions. Suggesting that expression under the strong P_{AOX1} without methanol can be possible. This interpretation depends on comparison of strains harbouring the same number of expression cassettes (if possible only one integration).

Also for the metabolic engineering research field the knowledge of single copy integration variants is useful. Schroer *et al.* identified bottlenecks during metabolic flux by over-expression of genes involved in the pathway of interest, in this case the *P. pastoris* methanol utilization pathway [91]. Herein the systematically over-expression of genes responsible for methanol dissimilation led to the identification of formaldehyde dehydrogenase as the rate limiting step in this pathway. The interpretation of these results was possible due to comparison of single copy strains, which again was facilitated by a fast determination of integrated expression cassettes.

A third application showing the importance of this method in the field of strain engineering is also described in the next chapter.

3.3.1 Real time PCR based determination of gene copy numbers in *Pichia pastoris*

This manuscript has been accepted for publication in Biotechnology Journal (accepted January 26th 2010, DOI: 10.1002/biot.200900233).

Sandra Abad [1,*], Kerstin Kitz [1,*], Astrid Hörmann [2], Ulrike Schreiner[1], Franz S. Hartner [1], Anton Glieder [1,2 §].

[1]Applied Biocatalysis Research Centre, Petersgasse 14, 8010 Graz, Austria
[2]Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria
[*]equally contributing authors
[§]Corresponding author

Keywords: real-time PCR, *Pichia pastoris*, *Ca*lB

Correspondence: Dr. Anton Glieder
Petersgasse 14/II, 8010 Graz, Austria
email: anton.glieder@a-b.at
tel: +43 316 8734074
fax: +43 316 8739302

3.3.1.1 Abstract

Pichia pastoris is a preferred host for heterologous protein production. Expression cassettes are usually integrated into the genome of this methylotrophic yeast. This manuscript describes a method for fast and reliable gene copy number determinations for *P. pastoris* expression strains. We believe that gene copy number determinations are important for all researchers working with *P. pastoris* and also many other research groups using similar gene integration techniques for the transformation of other yeasts. The described method uses Real Time PCR to quantify the integrated expression cassettes. Similar methods were employed previously for other host systems such as animal and plant cells but no such method comparing different detection methods and describing details for yeast analysis by quantitative PCR is known to us especially not for methylotrophic yeasts such as *P. pastoris*. Neglecting gene copy numbers can easily lead to false interpretations of experimental results from codon optimization or promoter studies and coexpression of helper proteins as demonstrated in an application example, which is also described in this manuscript.

3.3.1.2 Introduction

The methylotrophic yeast *Pichia pastoris* has emerged to a frequently used protein expression host. High product titers have been obtained with the standard expression system [4, 92] employing the methanol inducible *AOX1* promoter (P_{AOX1}). Nevertheless, recent studies reported even higher expression rates when the target gene was under the control of deletion variants of the *AOX1* promoter [88] or in some cases also with the constitutive glyceraldehyde-3-phosphate dehydrogenase promoter (P_{GAP}) [61, 88]. Depending on the target gene, high copy number expression strains do not necessarily lead to optimal expression rates. For example, for some secreted proteins rising numbers of expression cassettes resulted in a decreased expression due to other bottlenecks than transcript levels [93]. Other approaches such as co-expression of helper proteins e.g. chaperones or proteins which initiate the unfolded protein response have shown reasonable expression improvements [7-9, 94]

Although many studies describe improved expression due to different promoters variants, codon optimization or co-expression of helper proteins, hardly any information of copy numbers of the expression cassettes is given. Consequently, false

interpretation of experimental results is possible, leaving the question whether improvements either correlated with a certain promoter, co-expressed chaperone or an accompanying copy number effect. Inan *et al.* pointed out the importance of the copy number determination for *P. pastoris* expression strains, reporting the correlation between increased protein secretion and the copy number of the co-expressed chaperone protein disulfide isomerase from *P. pastoris*, *Pp*PDI. Copy numbers were determined by southern blot analysis, which was so far the method of choice for this question. However, it is a time-consuming and laborious method, and requires large amounts of genomic DNA. In addition, restriction site loss during integration is possible, complicating the molecular analysis. To overcome these drawbacks and also in order to deal with increasing numbers of expression strains from high throughput experiments, an additional reliable and faster method for copy number determination of integrated expression cassettes was needed.

Quantitative PCR emerged as an important and widely used analytical tool (reviewed in [9]). But, there are only few studies reporting on real time PCR based copy number determination in human-, animal- and plants cells [95, 96]. No detailed method for yeast could be found, probably also since for many yeasts plasmid systems are still quite common. In general, assay design, template preparation and analytical methods are essential for accurate quantitative gene amplification [97, 98] and were especially adapted for *P. pastoris* in this study as a fast and simple alternative to southern blotting.

Two settled detection methods, namely TaqMan and SYBR Green were compared, as well as approaches for absolute and relative quantification. For a simple visualization of our results and for a correlation between copy numbers and expression, green fluorescent protein (GFP) fusion with a Zeocin resistance marker has been chosen as a simple intracellular reporter system [99]. Furthermore the importance of expression cassettes quantitation in *P. pastoris* was emphasized in an industrially relevant example. Real time PCR was already used for copy number determination in *P. pastoris* in our group and independently also by others before [20, 88, 91]. A demonstration of the importance of copy number determination as a routinely performed experiment for molecular strain characterization and a first detailed protocol especially adapted for *P. pastoris* and including the comparison of two different detection methods, were the main goals of this work.

3.3.1.3 Materials and methods

Chemicals and materials

Sterile water was purchased from Fresenius Kabi Austria (Graz, Austria). Unless otherwise stated, all chemicals were purchased from Carl Roth GmbH (Karlsruhe, Germany), Becton, Dickinson and Company (Franklin Lakes, NJ, USA) and Sigma-Aldrich (St Louis, MO, USA). DNA modifying enzymes were supplied by Fermentas GmbH (Burlington, Ontario, Canada). All PCRs were performed with Phusion HF according to the producer's manual (Finnzymes Oy, Espoo, Finland).

GFP reporter expression cassette

P_{AOX1} and the *AOX1* terminator were amplified from vector pPICZ-B (Invitrogen, Carlsbad, USA), and cycle-3-GFP-Zeocin from pTracer™-CMV2 [99, 100]. All PCR products were purified by agarose gel electrophoresis prior to overlap extension PCR [85].

The obtained PCR product displayed 2.4 kb and was purified by agarose gel electrophoresis and gel extraction using Wizard® SV Gel and PCR Clean-Up system (Promega, Madison, WI, USA). pAOX was constructed by digesting the P_{AOX1}-GFP-Zeo-*AOX*1TT fragment with *Kpn*I and *Not*I and ligation with *Kpn*I/*Not*I-digested pBlueScript® SK$^-$ vector (Stratagene, La Jolla, CA, USA).

The *P. pastoris* P_{AOX1}d6* promoter variant was excised from vector pPICZ-GFP-d6* [88] employing *Bgl*II/*Eco*RI restriction sites into the vector pAOX thereby generating vector pAOXd6*. The correct sequence was verified by sequencing (AGOWA Genomics, Berlin, Germany).

***P. pastoris* transformation**

*Bgl*II and *Not*I linearized pAOX or pAOXΔ6* was transformed into electro-competent *P. pastoris* cells. Electro-transformation was performed as described in [101]. After a regeneration phase, aliquots were spread on MSM-Zeo agar plates (1.5% Bacto™ Agar, 1.34% Difco™ Yeast Nitrogen Base w/o Amino Acids, $4 \cdot 10^{-5}$% d-Biotin, 2% Sorbitol, 1% methanol, 100 µg/mL Zeocin™) and incubated for 3-4 days at 30°C.

***Pichia pastoris* microscale cultivation and GFP reporter expression**

P. pastoris cultivation for GFP expressing strains was performed as described in [66]. GFP intensity was detected with SPECTRA MAX Gemini XS plate reader (Molecular Devices Corp., Sunnyvale, CA, USA) as described in [88].

Generation of a *Pichia pastoris* genomic DNA-library

Genomic DNA (gDNA) from *P. pastoris* X-33 was isolated employing the Easy-DNA™-Kit (Invitrogen). 10µg from the DNA was partially digested with *Sau*3AI. Gel purified fragments (6-9 kb) were ligated into *Xho*I digested pGAPZ plasmid (Invitrogen) which was partially filled with Klenow fragment. Ligation was transformed into *E.coli* Top 10 F' (Invitrogen). Transformation was plated into LB-Zeocin (25 mg/mL). Transformants were pooled and cultivated for plasmid isolation with Wizard® Plus SV (Promega) according to manufacturer's protocol. The library contained 10,500 clones and the average insert size of 14 independent transformants was estimated to 4.5 kb. The probability to find any given unique sequence in this library is 99%.

***Ca*lB reporter strain construction**

Zeocin resistance cassette from pPICZ-B (Invitrogen) was replaced with the KanMX6 cassette from PFA6a-kanMX6 [86], which confers resistance against Kanamycin in *E. coli* and Geneticin (G418) in yeasts. The resulting plasmid was named pPICK. Codon optimized *Candida antartica* lipase B (*Ca*lB, GenBank: ACIO6118) was assembled via oePCR. The resulting gene was *Eco*RI/*Not*I digested, gel purified and cloned into an equally digested pPICK-plasmid. *E.coli* Top 10 F' (Invitrogen) was transformed with this construct and plated onto LB-Kanamycin (50 mg/mL). The construct was analyzed by sequencing, and linearized with *Bgl*II for *P. pastoris* KM71H (Invitrogen) transformation as described before. After transformation the cell suspensions were plated on YPD-G418-plates (300 mg/L) and incubated at 28℃ for 2 days. Transformants were cultivated in deep-well plates as described before, and screened for *Ca*lB activity.

Transformation of X-33 gDNA library into *Ca*lB reporter strain

The *P. pastoris* *Ca*lB-H5 (multicopy *Ca*lB strain) was transformed with 3 µg of *Bgl*II linearized genomic DNA (gDNA) library. Cells were plated in YPD-Zeocin-plates (50 mg/L) and incubated at 28℃ for 2 days. Using a Genetix QPixII robot (Genetix Limited, Hampshire, UK) the transformants were picked into 384-plates filled with 50 µL BMD1%. Incubation was carried out at 28℃ for 5 days. Glycerol-stocks were

prepared by addition of 30 µL of glycerol/BMD1% to a final concentration of 15% glycerol. The 384-plates were covered with foil seal (Silverseal Aluminium-tape, Greiner, Frickenhausen, Germany) and stored at -80℃. Transformants were replicated into deep-well plates and cultivated as described before with one minimal modification, cell harvest after 60 h of methanol induction.

*Ca*lB activity assay

P. pastoris strains were cultivated in deep-well plates as described before. Esterase activity determination of *Ca*lB was performed according to [68] with minimal modifications. Briefly, 20 µL of the supernatant or a dilution of it was pipetted into a micro-titer plate well. The *Ca*lB esterase activity was measured following color development (due to p-nitro phenolate formation) at 405 nm for 5min at RT after addition of 180 µL of the assay solution. The assay solution consisted of 300 mM Tris/HCl pH7, 1% ethanol and 4 mM of *p*-nitrophenyl butyrate. One unit was defined as the formation of 1 µmol *para* nitrophenol per minute (ε= 17,700 $M^{-1}cm^{-1}$)

Recovery of the library-insert from the *Ca*lB-clones

gDNA from selected strains was isolated employing the Easy-DNA™-Kit from Invitrogen. 500 ng gDNA were digested with *Bgl*II, after purification with Wizard® SV Gel and PCR Clean-Up system (Promega), self-ligation with T4-ligase was incubated at 16℃ over night. 2 µL of the ligation mixture were transformed into chemical competent One Shot® Top10 (Invitrogen). Resulting transformants were sent for sequencing (AGOWA genomics).

Isolation of genomic DNA for RT-PCR

The isolation procedure was performed as described by Hoffman and Winston [75]. The DNA concentration was measured at 260 nm. DNA quality was checked by gel electrophoresis and photometrically (SPECTRAmax PLUS plate reader, Molecular Devices Corp.). Samples displaying 260 nm/280 nm ratios higher than 1.8 were used for further analysis. Alternatively gDNA was isolated employing the E.N.Z.A.® Yeast DNA Kit (Omega Bio-Tek, Norcross, Georgia, USA) and the Easy-DNA™-Kit (Invitrogen) as described by the suppliers protocols. gDNA isolation was also performed with Y-PER Yeast Protein Extraction Reagent (Thermo Fisher Scientific Inc.,Rockford, Illinois, USA). Briefly, the cells of a 10-15 mL an overnight culture were lysed with 300 µL Y-PER (30 min, room temperature, mixed on an Eppendorf shaker

with 600 rpm). Lysed cells were spun down at 13,000 rpm, 10 min at room temperature. The supernatant then was subjected to a chloroform/phenol/isopropanol extraction followed by ethanol precipitation. Finally, the DNA pellet was resuspended with 50 µL H_2O, dest.

Primer sets for real time PCR

The primer design was performed using *Primer Express v2.0* software (Applied Biosystems, Foster City, CA, USA). The primers had similar melting temperatures (T_m) (58 – 60 °C) and the T_m of the probe was 10 °C higher. The amplicon size settings were defined between 80 and 150 bp. The nucleotide sequences are shown in Table 1. Primers were purchased from Invitrogen and probes were purchased from MWG-Biotech (Ebersberg, Germany).

The real time PCR assays were optimized with respect to the primer concentration by titration experiments. Thereby, combinations of different concentrations (100 nM – 300 nM) of forward and reverse primers for *GFP* and *ARG4* genes, respectively, were tested, using the same template amount per real time PCR reaction (see below). The primer combinations with lowest Ct values and no occurrence of by-products in the non-template control (NTC) reactions were considered as optimal.

Real time PCR using SYBR Green and TaqMan probes

Real time PCR amplification was performed using an ABI 7300 instrument with Sequence Detection Software SDS v1.2 (Applied Biosystems). Real-time PCR mixtures were prepared using the TaqMan® Universal PCR Master Mix or Power SYBR® Green Master Mix (both Applied Biosystems), respectively.

For TaqMan probe based determinations, a single reaction contained 1x TaqMan® Universal PCR Master Mix, 200 nM of each primer and 125 nM of the corresponding FAM/TAMRA labeled probe. For SYBR Green examinations, 200 nM of each primer were added to the 1x Power SYBR® Green Master Mix. In both cases, reactions were performed in triplicate for the method establishment and in duplicate for faster routine strain analysis. 2 ng of genomic DNA were added in a reaction volume of 20 µL.

For each unknown strain, reactions with GFP and ARG4 primer sets were performed in separate tubes. Standard curves and NTCs for both genes were recorded in every plate. The reactions were done in MicroAmp® Optical 96-Well Reaction Plates sealed with MicroAmp® Optical Adhesive Covers (Applied Biosystems).

The thermal profile initiates with a 10 min step at 95°C followed by 40 cycles of 15 sec at 95°C and 1 min at 60°C. The data collection of the fluorescence signal was performed at the end of the elongation step. These settings were the same for both detection strategies. In case of the SYBR Green samples, the amplification period was followed by a melting curve analysis with a temperature gradient of 0.1°C/s from 70 to 95°C to exclude amplification of unspecific products.

Establishment of standard curves for copy number determination

The GFP-Zeo strain X-33 pAOX D2 was first identified as single copy strain via two colony PCRs. The first primer pair served as control and binds in the coding region of the GFP gene, the second pair was set on the origin of replication region (forward) of the integrated linearized plasmid and in the P_{AOX1} (reverse). One or multiple tandem integration of the expression cassette delivers a defined band with both primer combinations, whereas the single copy integration only displays a band with the control primer setting. The strain was confirmed by previously measured intracellular fluorescence level as a single copy strain due to its low value in the fluorescent/transformant landscape and selected as calibrator for the establishment of the *GFP* and *ARG4* standard curves. In addition the integration of further expression cassettes in other orientations or in other loci was excluded by southern blot analysis (data not shown). The standard curves cover a copy quantity range from 1.2×10^4 to 3.1×10^6 and 1.1×10^5 to 3.1×10^6 copies per reaction for SYBR Green and TaqMan detection, respectively.

For calculations with the Sequence Detection Software SDS v1.2 (Applied Biosystems), the copy quantity equation as described in [22] was used, considering a *P. pastoris* genome size of 9.7 Mbp [102] since at the time when this work was performed the genome sequence was not public available. Resulting in 94,000 copies of the genome present in 1 ng of haploid *P. pastoris* genomic DNA. Meanwhile the published sequence (GenBank: FN392319.1, FN392320.1, FN392321.1, FN392322.1, FN392323.1, FN392324.1 and FN392325.1) which still contains a few gaps shows a similar size of about 9.4 Mbp [11].

The mean Ct values were plotted against the $\log^{10}$ of their initial template copy quantity and standard curves were generated by a linear regression of the plotted points. The PCR efficiency was calculated from the slope of each standard curve by the following equation [22]:

$E = 10^{-1/slope} - 1$

Samples displaying Ct values lower than 26 and standard deviations lower than 0.3 were used for copy number determinations according to absolute and relative calculations.

Absolute and relative quantification

For correct determination of the starting copy quantity regardless of precise amounts and qualities of input genomic DNAs, the reference gene *ARG4* was also quantified in parallel. The normalized copy number calculated by absolute quantification is given by the following equation [96]:

$$Copy\ number_{TARGET\ GENE} = \frac{Copy\ quantity_{TARGET\ GENE}}{Copy\ quantity_{ARG4GENE}}$$

Relative quantification of the copy number was performed according to the $2^{-\Delta\Delta Ct}$ method [103], where $\Delta\Delta$Ct = ΔCt of target – ΔCt of calibrator; ΔCt = Ct of target or calibrator – Ct of reference (*ARG4*); E = PCR efficiency.

$$Copy\ number_{TARGET\ GENE} = (1+E)^{-\Delta\Delta Ct}$$

A strain with known copy number (in our case *P. pastoris* X-33 pAOX D2), ideally with one copy of the target gene served as a calibrator strain.

3.3.1.4 Results and discussion

The quality of the DNA sample is one key step for successful quantification via RT-PCR [98]. Therefore the search of a gDNA preparation method which delivers suitable gDNA with a fast isolation protocol was an important step. *P. pastoris* standard fast isolation protocols using commercially available chemicals delivered high yields of gDNA, but also caused DNA degradation (Figure 1). Excellent results were obtained using glass beads and phenol/chloroform extraction as described by Hoffmann and Winston [75].

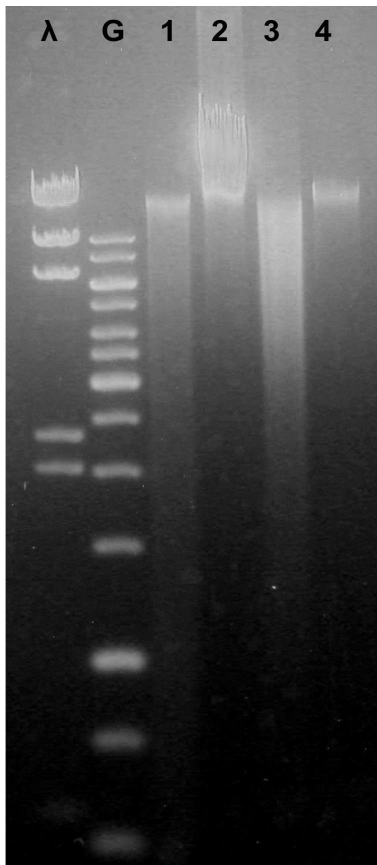

Figure 1: Comparison of *P. pastoris* gDNA quality of diverse isolation methods.
G: GeneRuler 1kb (Fermentas), λ: λ-HindIII Standard (Fermentas), 1: 5 µL of gDNA isolated using Y-PER (Novagen) and chloroform/phenol extraction followed by ethanol precipitation, 2: 5 µL of gDNA isolated according to Easy DNA™-Kit (Invitrogen), 3: 5 µL of gDNA isolated using a E.Z.N.A.® Yeast DNA Kit, 4: 5 µL of gDNA isolated using glass beads and phenol/chloroform extraction as described by Hoffmann and Winston [75].

In total, six *P. pastoris* strains carrying the GFP-Zeocin expression construct were chosen as models for the establishment of quantitative RT-PCR in *P. pastoris*. gDNA of these *P. pastoris* strains, two under the control of the wild type P_{AOX1} (named X-33 pAOX D2 and E2) and four under the control of the deletion variant P_{AOX1}d6* (named X-33 pAOXd6* A5, D10, F9 and F2) were analyzed (Table 1). Assuming that GFP expression and therefore the measured fluorescence intensity increases with the number of integrated copies we expected that *P. pastoris* X-33 pAOX D2 was a single copy strain. This strain was used for the generation of the standard curve required for the real time PCR experiments.
gDNA standard curves were generated in triplicate determinations and showed to be linear over a range from 1.2×10^4 to 3.1×10^6 copies (based on an estimated 9.7 Mbp full genome size, see material and methods) in the case of SYBR Green detection. A

tighter linear range was detected with the TaqMan method in our hands. The standard curves showed high efficiency (>0.92) and were approximately equal for the target and the reference gene, which is a prerequisite for copy number calculations using the relative $2^{-\Delta\Delta Ct}$ method [103]. As at the time when we started these studies no reports for real time PCR performed in *P. pastoris* were available, we decided to use the *ARG4* gene as an endogenous control [104] as this is a single copy gene of the haploid host. In this comparative technique, the amount of the target gene *GFP* is normalized by using the reference gene *ARG4* and set relative to the calibrator, which was in our case the single copy X-33 pAOX D2 strain (Figure 2). Hence, the *GFP/ARG4* ratio of a sample was normalized by the *GFP/ARG4* ratio of the calibrator.

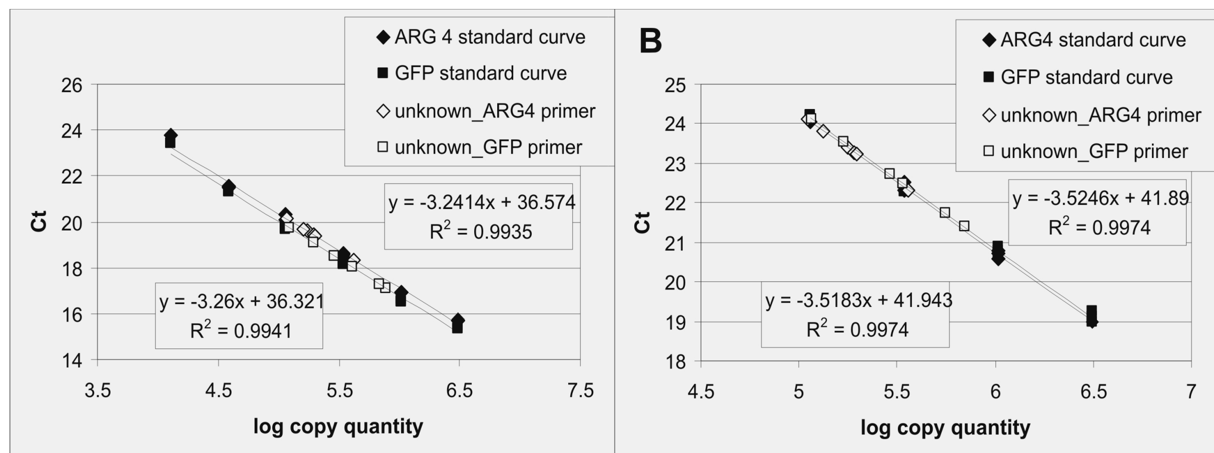

Figure 2: RT-PCR determination using SYBR Green and TaqMan methods. Target and reference gene amplification of *P. pastoris* GFP-Zeocin strains and standard curve. A: SYBR Green detection, B: TaqMan detection.

Two detection strategies were followed: 1) sequence-specific TaqMan probes and 2) SYBR Green. It was tested if the use of SYBR Green leads to similar results as the more expensive TaqMan probe based detection system, which is described to be the more sensitive detection method [105] (Table 1).

Table 1: Copy number calculation according to absolute (Abs. Q) and relative (Rel. Q) quantification.

Pichia pastoris strain	SYBR Green[a)]		TaqMan[b)]		Resulting copy number	Measured Expression GFP: [RFU], *Cal*B: [mU/L]	Expected copy number [e)]
	Abs. Q	Rel. Q	Abs Q.	Rel Q.			
X-33 pAOX D2	1.0	1.0	0.8	1.0	1 [c)]	987 ± 81	1
X-33 pAOX E2	1.7	1.8	1.6	2.0	2 [c)]	2068 ± 219	2
X-33 pAOXΔ6* A5	1.1	1.1	1.0	1.2	1 [c)]	422 ± 22	1
X-33 pAOXΔ6* D10	1.9	2.1	2.0	2.4	2 [c)]	1063 ± 189	2
X-33 pAOXΔ6* F9	2.0	2.1	1.8	2.2	2 [c)]	925 ± 58	2
X-33 pAOXΔ6* F2	4.6	4.8	4.0	5.0	4-5 [c)]	1678 ± 437	4-5
KM71H pPICK-*Cal*B H5	8.9	8.6	-	-	7-8 [d)]	0.033 ± 0.005	-
KM71H *Cal*B H5/2-L18	5.7	5.8	-	-	4-5 [d)]	0.390 ± 0.051	-
KM71H *Cal*B H5/9-H21	2.2	2.3	-	-	1 [d)]	0.853 ± 0.086	-
KM71H *Cal*B H5/10-H22	2.6	2.7	-	-	1-2 [d)]	0.876 ± 0.054	-
KM71H *Cal*B H5/10-F22	3.1	3.1	-	-	2 [d)]	0.955 ± 0.075	-

[a)] ARG(sense/antisense): 5'-TCCTCCGGTGGCAGTTCTT-3'/5'-TCCATTGACTCCCGTTTTGAG-3' and GFP(sense/antisense): 5'-AAATTTAAGGGTAAGCTTTCCGTATG-3'/5'-ATGGTGATGTTAATGGGCACAA-3' or P_{AOX1}(sense/antisense): 5´-GAAGCTGCCCTGTCTTAAACCTT-3´/5´-CAAAAGCTTGTCAATTGGAACCA-3´

[b)] GFP-probe: 5'-ATCACCTTCACCCTCTCCACTGACAGAAAA-3' , ARG-probe: 5'-TTCAACACTGGCTTCAAAGTCAAACGTTGAA-3'

[c)] based on absolute and relative quantifications results with SYBR Green and TaqMan methods.

[d)] based on results of absolute and relative quantifications with SYBR Green method. The endogenous P_{AOX1} sequence has been subtracted.

[e)] based on fluorescence measurement assuming a linear correlation between GFP intensity and the number of integrated copies.

In the range of one to five copies of the expression cassette the results were similar, no matter if SYBR Green or TaqMan probe was used. Important to mention are the higher Ct values obtained with the TaqMan detection method which is in line with the already mentioned narrow linear range compared with the SYBR Green method. In any case absolute or relative quantification delivered consistent results, although a modest deviation for the multi copy strain was detected with the TaqMan method.

The quantitation of integrated expression cassettes showed that pAOX D2 and pAOXd6* A5, although displaying different GFP intensities, are single copy transformants (or have identically integrated expression cassettes). These results confirm that the deletion d6* leads to a decrease of promoter activity to about 60% [88] under the described cultivation conditions.

Sustaining the aim of a universal tool for copy number determination, which fits for numbers of different expression constructs, good results have been obtained by using common sequences (e.g. P_{AOX1}), which are present in the integrated expression cassette and in the *P. pastoris* haploid genome. This setting resolves the identification of one single copy strain (calibrator), since the wild type strain already contains one copy of the sequence, and designed primers and probes can be used as universal tools.

To show the importance of this method in a laboratory routine, a real case scenario has been followed. The already mentioned benefits of helper protein co-expression have been focused in a so called non-rational approach. Therefore, a reporter strain was transformed with a *P. pastoris* genomic DNA library cloned under the control of the P_{GAP}. The *P. pastoris* reporter strain harbored 7-8 copies of a *Cal*B expression cassette under the control of the P_{AOX1} (named *P. pastoris* pPICK-*Cal*B H5), and displayed rather low levels of active enzyme in the culture supernatant compared to single copy strains. Previous experiments showed that certain bottlenecks during the secretion pathway led to a limited growth of the *P. pastoris Cal*B strains and low expression levels of the target gene, which could be at least in part be removed by co-expression of *Pp*PDI (S. Abad unpublished data). Following this outcome, we aimed at identifying additional putative helper proteins by a second transformation and individual co-overexpression of *P. pastoris* genomic DNA library clones. Therefore genomic *P. pastoris* DNA fragments had been ligated behind the GAP promoter of the pGAPZB expression plasmid. Putative new helper proteins from this library were

expected to lead to an increased *Ca*lB expression which can be determined by measuring the lipase activity in the culture supernatant of the reporter strain.
In total 3250 transformants were screened, which seemed to be an acceptable number for our initial purpose since four candidates (*CalB* H5/2-L18, H5/9-H21, H5/10-H22, H5/10-F22) displayed higher *Ca*lB expression levels as the reporter strain. These improved transformants were chosen for a re-screen and the expression improvement was confirmed for all candidates. Interestingly, none of the chosen co-expression strains delivered higher expression values than single copy *Ca*lB strains, and no reproducibility of the results was possible after transformation of the recovered construct into the reporter strain. The copy number determination performed with the SYBR Green method using primers designed for P_{AOX1} sequence and the *P. pastoris* KM71H strain as calibrator explained the described results. The chosen "improved" transformants showed lower numbers of integrated *Ca*lB expression cassettes. The transformants H5/2-L18 and H5/10-F22 harbored four and two copies, respectively while H5/9-H21 and H5/10-H22 displayed only one *Ca*lB expression cassette (Table1).

3.3.1.5 Concluding remarks

Quantitative PCR allows the determination of the number of integrated expression cassettes of *P. pastoris*. Strains harboring up to five copies were analyzed and delivered consistent results with two different detection methods. However, the narrow linear range of the TaqMan method should be taken into account for further experiments. This PCR based procedure provides a reliable but faster and simple alternative to standard southern blot analysis. It also allows correlations between integration events into the genome and the respective expression levels which is highly relevant for expression studies employing different promoters, redesigned genes or alternative signal sequences. Due to its simplicity it can be applied routinely.
Absolute and relative quantification showed to have no influence in the final outcome of the experiment, whereas the latter demands more precise parameters. The house-keeping gene *ARG4* showed to be a reliable endogenous control. However, caution should be taken when an Arg^- auxotroph strain is used as platform for expression experiments.

As demonstrated by application of the described method multiple transformation steps can lead to the potential loss of previously integrated expression cassettes, especially if the same locus is targeted for the genomic integration. Gene dosage control of previously integrated expression cassettes after every further transformation step is essential.

Acknowledgments: The authors acknowledge the FFG, the Province of Styria and SFG for financial support.

Conflict of Interest Statement: The authors have declared no conflict of interest.

4. Conclusions and future perspectives

For the first time an alternative expression strategy with a peroxisomal localization proved to be successful and superior to standard strategies like simply using the native sequence or pursuing secretion into the media. The systematic exchange of peroxisomal targeting sequences in combination with a codon optimized gene showed that the C-terminal tripeptide –SKL is better recognized by the *P. pastoris* peroxisomal transport machinery than the *Tv*DAO tripeptide –PNL. Additionally in this study a codon optimized *Tv*DAO gene showed higher expression rates than the native one. This highlights the positive effect of the codon used for the design of the gene, which was calculated for high expression under methanol induction conditions in *P. pastoris*. Noteworthy is the high improvement in terms of expression level, which was not expected since the yeast *T. variabilis* is closer, related to *P. pastoris* than other higher eukaryotes, and similar codon usages were assumed.

A need along the Pichia community is the efficient establishment of *P. pastoris* in the field of directed evolution. In our research group first steps like amplification of linear expression cassettes via oePCR or vector systems that force single copy integration into the genome have been successfully developed, but a strain especially engineered for this topic is still missing. Therefore during this PhD thesis a *P. pastoris* library screening platform strain has been constructed, which enables the use of shorter expression cassettes. Although the desired reduction of the expression cassette was successful, it also resulted in lower transformation efficiencies. The experimental results suggest that integration into the *AOX1* locus though homologous recombination is not the most preferred integration targeting of expression cassettes. And confirmed the presence of a well developed non homologous recombination machinery in *P. pastoris*, which is in line with results observed in other organisms (e.g. *Aspergillus niger* [106], *Sordaria macrospora* [107]) and in our research group (Näätsaari, manuscript in preparation). At this point it remains questionable if further effort should be posed on the host strain or a redirection towards the expression cassettes and other selection markers that force single copy integration is more promising.

Another drawback when working with *P. pastoris* is the evaluation of the copy number of the integrated expression cassettes into the genome for accurate interpretation of the experimental results. Also here we recognized the need of the Pichia community

for a fast and reliable method that offers an alternative to standard Southern blot analysis. Therefore a detailed protocol that describes copy number determination in *P. pastoris* via RT-PCR has been prepared, and is routinely been used in our and other research groups. Although this method offers new opportunities to analyze several strains it also needs several controls in order to give quantitative results, which reduces the number of position for the “unknown” samples in one plate. This should be the starting point for further optimizations like for example the use of several dyes in one reaction well. In this manner one sample can be used for the amplification of the control gene and at the same time the “unknown” gene. Furthermore new RT-PCR kits are now commercially available that can quantify RNA direct from cells (Invitrogen Corp.), which should be taken into account for further improvements in the field of *P. pastoris* quantitative RT-PCR.

The development of strategies and tools for protein expression in *P. pastoris* has been pursued in this thesis. The results offer new insight in this field and contribute to the already existing *P. pastoris* molecular toolbox, and demonstrate that this well established yeast system can still improve as an expression host.

5. References

1. Cereghino JL, Cregg JM: **Heterologous protein expression in the methylotrophic yeast *Pichia pastoris***. *FEMS Microbiology Reviews* 2000, **24:**45-66.

2. Cregg JM, Cereghino JL, Shi J, Higgins DR: **Recombinant protein expression in *Pichia pastoris***. *Molecular Biotechnology* 2000, **16:**23-52.

3. Sreekrishna K, Brankamp RG, Kropp KE, Blankenship DT, Tsay JT, Smith PL, Wierschke JD, Subramaniam A, Birkenberger LA: **Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris***. *Gene* 1997, **190:**55-62.

4. Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM: **Heterologous protein production using the *Pichia pastoris* expression system**. *Yeast* 2005, **22:**249-270.

5. Daly R, Hearn MT: **Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production**. *Journal of Molecular Recognition* 2005, **18:**119-138.

6. Hasslacher M, Schall M, Hayn M, Bona R, Rumbold K, Luckl J, Griengl H, Kohlwein SD, Schwab H: **High-level intracellular expression of hydroxynitrile lyase from the tropical rubber tree *Hevea brasiliensis* in microbial hosts**. *Protein Expression and Purification* 1997, **11:**61-71.

7. Gasser B, Maurer M, Gach J, Kunert R, Mattanovich D: **Engineering of *Pichia pastoris* for improved production of antibody fragments**. *Biotechnology and Bioengineering* 2006, **94:**353-361.

8. Tsai CW, Duggan PF, Shimp RL, Miller LH, Narum DL: **Overproduction of *Pichia pastoris* or *Plasmodium falciparum* protein disulfide isomerase affects expression, folding and O-linked glycosylation of a malaria vaccine candidate expressed in *P. pastoris***. *Journal of Biotechnology* 2006, **121:**458-470.

9. Inan M, Aryasomayajula D, Sinha J, Meagher MM: **Enhancement of protein secretion in *Pichia pastoris* by overexpression of protein disulfide isomerase**. *Biotechnology and Bioengineering* 2006, **93:**771-778.

10. Damasceno LM, Anderson KA, Ritter G, Cregg JM, Old LJ, Batt CA: **Cooverexpression of chaperones for enhanced secretion of a single-chain antibody fragment in *Pichia pastoris***. *Applied Microbiology and Biotechnology* 2007, **74:**381-389.

11. De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouze P, Van de Peer Y, Callewaert N: **Genome sequence of the recombinant protein production host *Pichia pastoris***. *Nature Biotechnology* 2009, **27:**561-566.

12. Gonzalez-Barroso MM, Ledesma A, Lepper S, Perez-Magan E, Zaragoza P, Rial E: **Isolation and bioenergetic characterization of mitochondria from *Pichia pastoris***. *Yeast* 2006, **23:**307-313.

13. Wriessnegger T, Leitner E, Belegratis MR, Ingolic E, Daum G: **Lipid analysis of mitochondrial membranes from the yeast *Pichia pastoris***. *Biochimica et Biophysica Acta* 2009, **1791:**166-172.

14. Jacobs PP, Geysens S, Vervecken W, Contreras R, Callewaert N: **Engineering complex-type N-glycosylation in *Pichia pastoris* using GlycoSwitch technology**. *Nature Protocols* 2009, **4:**58-70.

15. Jacobs PP, Ryckaert S, Geysens S, De Vusser K, Callewaert N, Contreras R: ***Pichia* surface display: display of proteins on the surface of glycoengineered *Pichia pastoris* strains**. *Biotechnology Letters* 2008, **30:**2173-2181.

16. Kern A, Hartner FS, Freigassner M, Spielhofer J, Rumpf C, Leitner L, Frohlich KU, Glieder A: ***Pichia pastoris* 'just in time' alternative respiration**. *Microbiology* 2007, **153:**1250-1260.

17. Hermann M, Kietzmann MU, Ivancic M, Zenzmaier C, Luiten RG, Skranc W, Wubbolts M, Winkler M, Birner-Gruenberger R, Pichler H *et al*: **Alternative pig liver esterase (APLE) - Cloning, identification and functional expression in *Pichia pastoris* of a versatile new biocatalyst**. *Journal of Biotechnology* 2008, **133:**301-310.

18. Pichler H, Gaigg B, Hrastnik C, Achleitner G, Kohlwein SD, Zellnig G, Perktold A, Daum G: **A subfraction of the yeast endoplasmic reticulum associates with the plasma membrane and has a high capacity to synthesize lipids**. *Eur J Biochem* 2001, **268:**2351-2361.

19. Schreuder MP, Mooren AT, Toschka HY, Verrips CT, Klis FM: **Immobilizing proteins on the surface of yeast cells**. *Trends in Biotechnology* 1996, **14:**115-120.

20. Marx H, Sauer M, Resina D, Vai M, Porro D, Valero F, Ferrer P, Mattanovich D: **Cloning, disruption and protein secretory phenotype of the *GAS1* homologue of *Pichia pastoris***. *FEMS Microbiology Letters* 2006, **264:**40-47.

21. Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA: **Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris***. *Molecular and Cellular Biology* 1989, **9:**1316-1323.

22. Gleeson MA, Sudbery PE: **The methylotrophic yeasts**. *Yeast* 1988, **4:**1-15.

23. van der Klei IJ, Yurimoto H, Sakai Y, Veenhuis M: **The significance of peroxisomes in methanol metabolism in methylotrophic yeast**. *Biochimica et Biophysica Acta* 2006, **1763:**1453-1462.

24. Heiland I, Erdmann R: **Biogenesis of peroxisomes - Topogenesis of the peroxisomal membrane and matrix proteins**. *FEBS Journal* 2005, **272:**2362-2372.

25. van der Klei IJ, Hilbrands RE, Kiel JA, Rasmussen SW, Cregg JM, Veenhuis M: **The ubiquitin-conjugating enzyme Pex4p of *Hansenula polymorpha* is required for efficient functioning of the PTS1 import machinery**. *EMBO Journal* 1998, **17:**3608-3618.

26. Salomons FA, Kiel JA, Faber KN, Veenhuis M, van der Klei IJ: **Overproduction of Pex5p stimulates import of alcohol oxidase and dihydroxyacetone synthase in a *Hansenula polymorpha* Pex14 null mutant**. *Journal of Biological Chemistry* 2000, **275:**12603-12611.

27. Veenhuis M, van der Klei IJ: **Peroxisomes: surprisingly versatile organelles**. *Biochim Biophys Acta* 2002, **1555:**44-47.

28. van der Klei IJ, Veenhuis M: **PTS1-independent sorting of peroxisomal matrix proteins by Pex5p**. *Biochimica et Biophysica Acta* 2006, **1763:**1794-1800.

29. Ozimek P, Kotter P, Veenhuis M, van der Klei IJ: ***Hansenula polymorpha* and *Saccharomyces cerevisiae* Pex5p's recognize different, independent peroxisomal targeting signals in alcohol oxidase**. *FEBS Letters* 2006, **580:**46-50.

30. Ozimek P, van Dijk R, Latchev K, Gancedo C, Wang DY, van der Klei IJ, Veenhuis M: **Pyruvate carboxylase is an essential protein in the assembly of yeast peroxisomal oligomeric alcohol oxidase**. *Molecular Biology of the Cell* 2003, **14:**786-797.

31. Ozimek P, Veenhuis M, van der Klei IJ: **Alcohol oxidase: a complex peroxisomal, oligomeric flavoprotein**. *FEMS Yeast Research* 2005, **5:**975-983.

32. Riethorst W, Reichert A: **An industrial view on enzymes for the cleavage of cephalosporin C**. *Chimia* 1999, **53:**600-607.

33. Pilone MS, Pollegioni L: **D-amino acid oxidase as an industrial biocatalyst**. *Biocatalysis and Biotransformation* 2002, **20:**145-159.

34. Tishkov VI, Khoronenkova SV: **D-Amino acid oxidase: Structure, catalytic mechanism, and practical application**. *Biochemistry (Moscow)* 2005, **70:**40-54.

35. Pollegioni L, Molla G, Sacchi S, Rosini E, Verga R, Pilone MS: **Properties and applications of microbial D-amino acid oxidases: current state and perspectives**. *Applied Microbiology and Biotechnology* 2008, **78:**1-16.

36. Rosini E, Molla G, Rossetti C, Pilone MS, Pollegioni L, Sacchi S: **A biosensor for all D-amino acids using evolved D-amino acid oxidase**. *Journal of Biotechnology* 2008, **135:**377-384.

37. Khoronenkova SV, Tishkov VI: **D-amino acid oxidase: physiological role and applications**. *Biochemistry (Moscow)* 2008, **73:**1511-1518.

38. Beard TM, Turner NJ: **Deracemisation and stereoinversion of α-amino acids using D-amino acid oxidase and hydride reducing agents**. *Chemical Communications* 2002, **7:**246-247.

39. Pilone MS: **D-Amino acid oxidase: new findings**. *Cellular and Molecular Life Sciences* 2000, **57:**1732-1747.

40. Betancor L, Hidalgo A, Fernandez-Lorente G, Mateo C, Rodriguez V, Fuentes M, Lopez-Gallego F, Fernandez-Lafuente R, Guisan JM: **Use of physicochemical**

tools to determine the choice of optimal enzyme: stabilization of D-amino acid oxidase. *Biotechnology Progress* 2003, **19:**784-788.

41. Kubicek-Pranz EM, Rohr M: **D-amino acid oxidase from the yeast *Trigonopsis variabilis***. *Journal of Applied Biochemistry* 1985, **7:**104-113.

42. Szwajcer E, Mosbach K: **Isolation and partial characterization of a D-amino acid oxidase active against cephalosporin C from the yeast *Trigonopsis variabilis***. *Biotechnology Letters* 1985, **7:**1-7.

43. Pollegioni L, Caldinelli L, Molla G, Sacchi S, Pilone MS: **Catalytic properties of D-amino acid oxidase in cephalosporin C bioconversion: a comparison between proteins from different sources**. *Biotechnology Progress* 2004, **20:**467-473.

44. Kubicekpranz EM, Rohr M: **Formation of D-amino acid oxidase in the yeast *Trigonopsis variabilis***. *Canadian Journal of Microbiology* 1985, **31:**625-628.

45. Neuberger G, Maurer-Stroh S, Eisenhaber B, Hartig A, Eisenhaber F: **Motif refinement of the peroxisomal targeting signal 1 and evaluation of taxon-specific differences**. *Journal of Molecular Biology* 2003, **328:**567-579.

46. Neuberger G, Maurer-Stroh S, Eisenhaber B, Hartig A, Eisenhaber F: **Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence**. *Journal of Molecular Biology* 2003, **328:**581-592.

47. Emanuelsson O, Brunak S, von Heijne G, Nielsen H: **Locating proteins in the cell using TargetP, SignalP and related tools**. *Nature Protocols* 2007, **2:**953-971.

48. Nakai K, Horton P: **PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization**. *Trends in Biochemical Sciences* 1999, **24:**34-36.

49. Hoepfner D, Schildknegt D, Braakman I, Philippsen P, Tabak HF: **Contribution of the endoplasmic reticulum to peroxisome formation**. *Cell* 2005, **122:**85-95.

50. Dib I, Stanzer D, Nidetzky B: ***Trigonopsis variabilis* D-amino acid oxidase: control of protein quality and opportunities for biocatalysis through production in *Escherichia coli***. *Applied and Environmental Microbiology* 2007, **73:**331-333.

51. Isoai A, Kimura H, Reichert A, Schorgendorfer K, Nikaido K, Tohda H, Giga-Hama Y, Mutoh N, Kumagai H: **Production of D-amino acid oxidase (DAO) of *Trigonopsis variabilis* in *Schizosaccharomyces pombe* and the characterization of biocatalysts prepared with recombinant cells**. *Biotechnology and Bioengineering* 2002, **80:**22-32.

52. Yu J, Li DY, Zhang YJ, Yang S, Li RB, Yuan ZY: **High expression of *Trigonopsis variabilis* D-amino acid oxidase in *Pichia pastoris***. *Journal of Molecular Catalysis B: Enzymatic* 2002, **18:**291-297.

53. Gonzalez FJ, Montes J, Martin F, Lopez MC, Ferminan E, Catalan J, Galan MA, Dominguez A: **Molecular cloning of *Tv*DAO1, a gene encoding a D-amino**

acid oxidase from *Trigonopsis variabilis* and its expression in *Saccharomyces cerevisiae* and *Kluyveromyces lactis*. *Yeast* 1997, **13:**1399-1408.

54. Molla G, Vegezzi C, Pilone MS, Pollegioni L: **Overexpression in *Escherichia coli* of a recombinant chimeric *Rhodotorula gracilis* D-amino acid oxidase**. *Protein Expression and Purification* 1998, **14:**289-294.

55. Horner R, Wagner F, Fischer L: **Induction of the D-amino acid oxidase from *Trigonopsis variabilis***. *Applied and Environmental Microbiology* 1996, **62:**2106-2110.

56. Dib I, Nidetzky B: **The stabilizing effects of immobilization in D-amino acid oxidase from *Trigonopsis variabilis***. *BMC Biotechnology* 2008, **8:**72.

57. Nahalka J, Dib I, Nidetzky B: **Encapsulation of *Trigonopsis variabilis* D-amino acid oxidase and fast comparison of the operational stabilities of free and immobilized preparations of the enzyme**. *Biotechnology and Bioengineering* 2008, **99:**251-260.

58. Nahalka J, Nidetzky B: **Fusion to a pull-down domain: a novel approach of producing *Trigonopsis variabilis* D-amino acid oxidase as insoluble enzyme aggregates**. *Biotechnology and Bioengineering* 2007, **97:**454-461.

59. Khang YH, Kim IW, Hah YR, Hwangbo JH, Kang KK: **Fusion protein of *Vitreoscilla hemoglobin* with D-amino acid oxidase enhances activity and stability of biocatalyst in the bioconversion process of cephalosporin C**. *Biotechnology and Bioengineering* 2003, **82:**480-488.

60. Vikartovska-Welwardova A, Michalkova E, Gemeiner P, Welward L: **Stabilization of D-amino-acid oxidase from *Trigonopsis variabilis* by manganese dioxide**. *Folia Microbiologica* 1999, **44:**380-384.

61. Zheng HB, Wang XL, Chen J, Zhu K, Zhao YH, Yang YL, Yang S, Jiang W: **Expression, purification, and immobilization of His-tagged D-amino acid oxidase of *Trigonopsis variabilis* in *Pichia pastoris***. *Applied Microbiology and Biotechnology* 2006, **70:**683-689.

62. Fernandez-Lafuente R, Rodriguez V, Guisan JM: **The coimmobilization of D-amino acid oxidase and catalase enables the quantitative transformation of D-amino acids (D-phenylalanine) into α-keto acids (phenylpyruvic acid)**. *Enzyme and Microbial Technology* 1998, **23:**28-33.

63. Buto S, Pollegioni L, D'Angiuro L, Pilone MS: **Evaluation of D-amino acid oxidase from *Rhodotorula gracilis* for the production of α-keto acids: A reactor system**. *Biotechnology and Bioengineering* 1994, **44:**1288-1294.

64. Tan Q, Song QX, Zhang YW, Wei DZ: **Characterization and application of D-amino acid oxidase and catalase within permeabilized *Pichia pastoris* cells in bioconversions**. *Applied Biochemistry and Biotechnology* 2007, **136:**279-289.

65. Monosov EZ, Wenzel TJ, Luers GH, Heyman JA, Subramani S: **Labeling of peroxisomes with green fluorescent protein in living *P. pastoris* cells**. *Journal of Histochemistry and Cytochemistry* 1996, **44:**581-589.

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

67. Tschopp JF, Brust PF, Cregg JM, Stillman CA, Gingeras TR: **Expression of the *lacZ* gene from two methanol-regulated promoters in *Pichia pastoris***. *Nucleic Acids Research* 1987, **15:**3859-3876.

68. Zhang W, Hywood Potter KJ, Plantz BA, Schlegel VL, Smith LA, Meagher MM: ***Pichia pastoris* fermentation with mixed-feeds of glycerol and methanol: growth kinetics and production improvement**. *Journal of Industrial Microbiology and Biotechnology* 2003, **30:**210-215.

69. d'Anjou MC, Daugulis AJ: **A rational approach to improving productivity in recombinant *Pichia pastoris* fermentation**. *Biotechnology and Bioengineering* 2001, **72:**1-11.

70. Lin-Cereghino GP, Lin-Cereghino J, Ilgen C, Cregg JM: **Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris***. *Current Opinion in Biotechnology* 2002, **13:**329-332.

71. Chen RR: **Permeability issues in whole-cell bioprocesses and cellular membrane engineering**. *Applied Microbiology and Biotechnology* 2007, **74:**730-738.

72. Pscheidt B, Glieder A: **Yeast cell factories for fine chemical and API production**. *Microbial Cell Factories* 2008, **7:**25.

73. Sutcliffe JG: **Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322**. *Cold Spring Harbour Symposia on Quantitative Biology* 1979, **43:**77-90.

74. Lin-Cereghino J, Wong WW, Xiong S, Giang W, Luong LT, Vu J, Johnson SD, Lin-Cereghino GP: **Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris***. *Biotechniques* 2005, **38:**44, 46, 48.

75. Hoffman CS, Winston F: **A 10 minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli***. *Gene* 1987, **57:**267-272.

76. Hellwig S, Emde F, Raven NP, Henke M, van Der Logt P, Fischer R: **Analysis of single-chain antibody production in *Pichia pastoris* using on-line methanol control in fed-batch and mixed-feed fermentations**. *Biotechnology and Bioengineering* 2001, **74:**344-352.

77. Hellwig S, Robin F, Drossard J, Raven NP, Vaquero-Martin C, Shively JE, Fischer R: **Production of carcinoembryonic antigen (CEA) N-A3 domain in *Pichia pastoris* by fermentation**. *Biotechnology and Applied Biochemistry* 1999, **30:**267-275.

78. Alexeeva M, Enright A, Dawson MJ, Mahmoudian M, Turner NJ: **Deracemization of α-methylbenzylamine using an enzyme obtained by in vitro evolution**. *Angewandte Chemie International Edition* 2002, **41:**3177-3180.

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

80. Pohn B, Gerlach J, Scheideler M, Katz H, Uray M, Bischof H, Klimant I, Schwab H: **Micro-colony array based high throughput platform for enzyme library screening**. *Journal of Biotechnology* 2007, **129:**162-170.

81. Fujio T: **Minimum genome factory: innovation in bioprocesses through genome science**. *Biotechnology and Applied Biochemistry* 2007, **46:**145-146.

82. Arnold FH, Georgiou G: **Directed Enzyme Evolution**, vol. 230. Totowa, New Jersey; 2003.

83. Goddard JP, Reymond JL: **Enzyme assays for high-throughput screening**. *Current Opinion in Biotechnology* 2004, **15:**314-322.

84. Yang G, Withers SG: **Ultrahigh-throughput FACS-based screening for directed enzyme evolution**. *ChemBioChem* 2009, **10:**2704-2715.

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

86. Wach A, Brachat A, Alberti-Segui C, Rebischung C, Philippsen P: **Heterologous HIS3 marker and GFP reporter modules for PCR-targeting in *Saccharomyces cerevisiae***. *Yeast* 1997, **13:**1065-1075.

87. Hartner FS, Glieder A, **Mutant *Pichia pastoris* AOX1 promoters**, 2006, WO2006089329

88. Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, Lin-Cereghino J, Kovar K, Cregg JM, Glieder A: **Promoter library designed for fine-tuned gene expression in *Pichia pastoris***. *Nucleic Acids Research* 2008, **36:**76.

89. Southern EM: **Detection of specific sequences among DNA fragments separated by gel electrophoresis**. *Journal of Molecular Biology* 1975, **98:**503-517.

90. Zhang N, Suen WC, Windsor W, Xiao L, Madison V, Zaks A: **Improving tolerance of *Candida antarctica* lipase B towards irreversible thermal inactivation through directed evolution**. *Protein Engineering* 2003, **16:**599-605.

91. Schroer K, Peter Luef K, Stefan Hartner F, Glieder A, Pscheidt B: **Engineering the *Pichia pastoris* methanol oxidation pathway for improved NADH regeneration during whole-cell biotransformation**. *Metab Eng* 2009.

92. Lin-Cereghino J, Cregg JM: **Heterologous protein expression in the methylotrophic yeast *Pichia pastoris***. *FEMS Microbiology Review* 2000, **24:**45-66.

93. Hohenblum H, Gasser B, Maurer M, Borth N, Mattanovich D: **Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris***. *Biotechnology and Bioengineering* 2004, **85:**367-375.

94. Vad R, Nafstad E, Dahl LA, Gabrielsen OS: **Engineering of a *Pichia pastoris* expression system for secretion of high amounts of intact human parathyroid hormone**. *Journal of Biotechnology* 2005, **116:**251-260.

95. Ballester M, Castello A, Ibanez E, Sanchez A, Folch JM: **Real-time quantitative PCR-based system for determining transgene copy number in transgenic animals**. *Biotechniques* 2004, **37:**610-613.

96. Laurendeau I, Bahuau M, Vodovar N, Larramendy C, Olivi M, Bieche I, Vidaud M, Vidaud D: **TaqMan PCR-based gene dosage assay for predictive testing in individuals from a cancer family with INK4 locus haploinsufficiency**. *Clinical Chemistry* 1999, **45:**982-986.

97. Ginzinger DG: **Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream**. *Experimental Hematology* 2002, **30:**503-512.

98. Cankar K, Stebih D, Dreo T, Zel J, Gruden K: **Critical points of DNA quantification by real-time PCR-effects of DNA extraction method and sample matrix on quantification of genetically modified organisms**. *BMC Biotechnology* 2006, **6:**37.

99. Bennett RP, Cox CA, Hoeffler JP: **Fusion of green fluorescent protein with the Zeocin™-resistance marker allows visual screening and drug selection of transfected eukaryotic cells**. *Biotechniques* 1998, **24:**478-482.

100. Crameri A, Whitehorn EA, Tate E, Stemmer WP: **Improved green fluorescent protein by molecular evolution using DNA shuffling**. *Nature Biotechnology* 1996, **14:**315-319.

101. Cregg J.M. RKA: **Pichia protocols**, vol. 103. Totowa, New Jersey: Humana Press; 1998.

102. Ohi H, Okazaki N, Uno S, Miura M, Hiramatsu R: **Chromosomal DNA patterns and gene stability of *Pichia pastoris***. *Yeast* 1998, **14:**895-903.

103. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method**. *Methods* 2001, **25:**402-408.

104. Lin Cereghino GP, Lin Cereghino J, Sunga AJ, Johnson MA, Lim M, Gleeson MA, Cregg JM: **New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of *Pichia pastoris***. *Gene* 2001, **263:**159-169.

105. Bubner B, Baldwin IT: **Use of real-time PCR for determining copy number and zygosity in transgenic plants**. *Plant Cell Reports* 2004, **23:**263-271.

106. Meyer V, Arentshorst M, El-Ghezal A, Drews AC, Kooistra R, van den Hondel CA, Ram AF: **Highly efficient gene targeting in the *Aspergillus niger kusA* mutant**. *Journal of Biotechnology* 2007, **128:**770-775.

107. Pöggeler S, Kück U: **Highly efficient generation of signal transduction knockout mutants using a fungal strain deficient in the mammalian *ku70* ortholog**. *Gene* 2006, **378:**1-10.

6. Appendix

6.1 Strain collection and VNTI files

Table 1: Strains constructed during the thesis.

BT Number	Organism	Host Strain	Name	Comments	VNTI-file
5965	*E. coli*	TOP10F´	pJET-Tv1	*Tv*DAO: codon opt., -SKL, B1-ZeoR	pJET-Tv1
5966	*E. coli*	TOP10F´	pJET-Tv3	*Tv*DAO: codon opt., -SKL, N-term citrine fusion, B1-ZeoR	pJET-Tv3
5967	*E. coli*	TOP10F´	T2-Tv1mV-M40	*Tv*DAO: codon opt., -SKL, 8 pointmutations, T2-plasmid	T2-Tv1mV-M40
5964	*E. coli*	TOP10F´	pJET-An1	*An*MAO 5M: codon opt., -SKL, 5 pointmutations, B1-ZeoR	pJET-An1
5968	*E. coli*	TOP10F´	T2-RgDAO	*Rg*DAO: codon opt., T2-plasmid	T2-RgDAO
5969	*E. coli*	TOP10F´	T2-Cr1	*Cr*LAO: codon opt., *S.cerevisiae* alpha factor fusion	T2-Cr1
3099	*E. coli*	TOP10F´	pAOXKanZeoD	platform strain construct 2, pCRII-Topo-Blunt	pAOXKanZeoD
3622	*E. coli*	TOP10F´	pAOXZeoDKan	platform strain construct 1, pCRII-Topo-Blunt	pAOXZeoDKan
3623	*E. coli*	TOP10F´	T2- Tv1	*Tv*DAO: codon opt., -SKL, T2-ZeoR	T2- Tv1
3624	*E. coli*	TOP10F´	T2-Tv1a	*Tv*DAO: native, -SKL, T2-ZeoR	T2-Tv1a
3625	*E. coli*	TOP10F´	T2-Tv5	*Tv*DAO: codon opt., -PNL, T2-ZeoR	T2-Tv5
3626	*E. coli*	TOP10F´	T2-Tv5a	*Tv*DAO: native, -PNL, T2-ZeoR	T2-Tv5a
3577	*P. pastoris*	Muts7	Tv1 A9	*Tv*DAO: codon opt., -SKL, B1-ZeoR,	pJET-Tv1
3576	*P. pastoris*	Muts7	Tv3 H2	*Tv*DAO: codon opt., -SKL, N-term citrine fusion, B1-ZeoR	pJET-Tv3
3575	*P. pastoris*	KM71H	WT1	*Tv*DAO: codon opt., 5 pointmutations, Invitrogen material!	WT1
3579	*P. pastoris*	Muts7	T2-Cr1 G2	*Cr*LAO: codon opt., *S. cerevisiae* alpha factor fusion, T2-ZeoR	T2-Cr1
3578	*P. pastoris*	Muts7	T2-RgDAO A6	*Rg*DAO: codon opt., T2-ZeoR	T2-RgDAO
3580	*P. pastoris*	Muts7	An1 B6	*An*MAO 5M: codon opt., -SKL, 5 pointmutations, B1-ZeoR	pJET-An1
3627	*P. pastoris*	X-33	platform B2	pAOXZeoDKan	pAOXKanZeoD
3628	*P. pastoris*	X-33	Platform G3	pAOXKanZeoD	pAOXZeoDKan
3629	*P. pastoris*	Muts7	T2- Tv1 B10	*Tv*DAO: codon opt., -SKL, T2-ZeoR	T2- Tv1
3630	*P. pastoris*	Muts7	T2-TV1a G11	*Tv*DAO: native, -SKL, T2-ZeoR	T2-Tv1a
3631	*P. pastoris*	Muts7	T2-Tv5 H3	*Tv*DAO: codon opt., -PNL, T2-ZeoR	T2-Tv5
3632	*P. pastoris*	Muts7	T2-Tv5a G4	*Tv*DAO: native, -PNL, T2-ZeoR	T2-Tv5a
3197	*P. pastoris*	KM71H	pHRP-PDI B11	pHRP-AOX1 pPICKd6*-PDI B11	pPICKd6*-PDI
3198	*P. pastoris*	KM71H	pHRP-PDI D7	pHRP-AOX1 pPICK-PDI D7	pPICK-PDI