Mudassar Ahmad, M. Phil. Biotechnology

Development of tools and methods for genetic engineering of the recombinant protein production host *Pichia pastoris*

DOCTORAL THESIS

to achieve the university degree of

Doktor der technischen Wissenschaften

submitted to

Graz University of Technology

Supervisor

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

Institute of Molecular Biotechnology

Graz, October, 2015

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present doctoral thesis.

Date

Signature

Acknowledgement

First of all, I would like to thank my supervisor, Prof. Helmut Schwab, for giving me the opportunity to work in his lab, for support, advice and freedom to work on different projects. Secondly, I would like to express my appreciation to Dr. Harald Pichler for his valuable and timely suggestions in planning and conducting this research work, for his critical reading and correction of manuscripts.

I would always be grateful to Dr. Ingund Rosales Rodriguez, Tamara Nudl and Patrizia Inschlag for helping me, during the start of my PhD studies, with different laboratory methods. I would like to express my great appreciation to Dr. Christine Winkler, Dr. Thomas Vogl, Markus Kolmbauer, Manuela Killinger and Melanie Hirz for their help and support in carrying out the research work and writing of this thesis.

I would also like to thank Dr. Andrea Mellitzer, Sarah Trunk, Simone Scharl, Steffen Gruber, for their continuous encouragement during the course of my studies.

At the end, I would like to thank my family for their continuous support throughout my life.

TABLE OF CONTENTS

Abstract

During the past two decades, *Pichia pastoris* has emerged as a robust heterologous protein production host for industrial and pharmaceutical bioprocesses and for research applications. In the present study, we have developed several tools and methods for optimized recombinant protein production in this yeast. In first part of this thesis, a set of novel *Pichia* pool expression vectors was constructed and used to compare the secretory potential of *P. pastoris* alpha mating factor secretion signal (*Pp*_αMF) and *S. cerevisiae* alpha mating factor secretion signal (*Sc*_αMF). Strikingly, the heterologous *Sc*_αMF turned out to be superior to endogenous *Pp*_αMF in directing secretion of *B. subtilis* levanase and horseradish peroxidase. It also appears that cleavage of propeptide of *Pp*_αMF from fusion protein is more dependent on *KEX2* protease activity than for *Sc*_αMF.

In the second part of this thesis, a novel set of knockout vectors based on the *FLP/FRT* recombinase system to selectively target and delete genes was developed. These knockout vectors can be adapted to any gene or strain background with a single cloning step. This knockout system was used to create clean and marker free knockout strains for a number of biosynthetic and protease genes.

In the third part of this thesis, , based on type IIS restriction enzymes, a set of 40 expression plasmids for intracellular and secretory expression of recombinant proteins was constructed. A single PCR product of the gene of interest can be cloned in frame in all the constructed vectors irrespective of upstream or downstream DNA sequences, allowing the efficient study of effects of promoters, secretion signals, N and C-terminal tags on recombinant protein production in *P.pastoris*. In the fourth and last part of this thesis, a novel insertion mutagenesis method to randomly target and disrupt genes in *Pichia pastoris* is described. This method was used to identify

genes that affect secretion of horseradish peroxidase in this yeast. Briefly, linear DNA marker cassettes were used to randomly disrupt genes and mutant strains showing altered secretion levels of HRP were selected by medium throughput screening. Subsequently, integration loci were identified by template blocking PCR method for genome walking and further characterized by creating clean deletions of identified genes in CBS 7435 Δ*his4* strain background. A number of highly interesting genes affecting the secretion of a number of unrelated recombinant proteins were identified.

Zusammenfassung

In den letzten zwei Jahrzehnten hat sich *Pichia pastoris* als robuster Wirt für heterologe Proteinproduktion für den industriellen und pharmazeutischen Bereich sowie für Forschungszwecke entwickelt. In der vorliegenden Studie wurden verschiedene Werkzeuge und Methoden zur optimierten Produktion rekombinanter Proteine in dieser Hefe entwickelt. Im ersten Teil der Arbeit wurde eine Reihe von neuartigen Pichia Pool Expressionsvektoren konstruiert und verwendet, um das sekretorische Potential der Sekretionssignale von *P. pastoris* alpha-Paarungsfaktor (*Pp*_αMF) und *S. cerevisiae* alpha-Paarungsfaktor (*Sc*_αMF) zu vergleichen. Auffallend ist, dass *S. c* αMF im Vergleich zu *Pp*_αMF überlegen hinsichtlich der Sekretion von *B. subtilis* Levanase und Meerrettich-Peroxidase ist. Es scheint auch, dass die Abspaltung des Propeptids von *Pp_αMF* stärker abhängig von *KEX2*-Protease Aktivität ist als bei *Sc*_αMF.

Im zweiten Teil dieser Arbeit haben wir, beruhend auf dem *FLP/FRT* Rekombinase-System, eine Reihe von neuen knockout-Vektoren entwickelt um selektiv Gene zu adressieren oder zu deletieren. Die knockout-Vektoren können für jedes Gen oder jeden Stammhintergrund mit einem einzigen Klonierungsschritt adaptiert werden. Dieses knockout-System wurde genutzt, um saubere und Marker-freie knockout-Stämme für eine Reihe von Biosynthese- und Protease-Genen zu erstellen.

Im dritten Teil dieser Arbeit wurde, basierend auf Type IIS Restriktionsenzymen, eine Reihe von 40-Expressionsplasmiden für die intrazelluläre und sekretorische Expression von rekombinanten Proteinen konstruiert. Ein einzelnes PCR-Produkt des Gens von Interesse kann „in-frame“ in allen konstruierten Vektoren unabhängig von vor- oder nachgeschalteten DNA Sequenzen kloniert werden. Dies ermöglicht die effiziente Untersuchung der Wirkung von Promotoren,

Sekretionssignalen, N und C-terminalen Tags auf die Produktion rekombinanter Proteine in *P. pastoris*.

Im vierten und letzten Teil dieser Arbeit wird ein neuartiges Insertions-Mutagenese Verfahren beschrieben, mit dem in zufälliger Weise Gene in *Pichia pastoris* adressiert werden können. Diese Methode wurde benutzt, um Gene zu identifizieren, die die Sekretion von Meerrettich-Peroxidase (HRP) in dieser Hefe beeinflussen. Kurz gesagt, wurden lineare DNA-Marker-Kassetten verwendet, um nach dem Zufallsprinzip Gene zu stören. Mutantenstämme, die veränderte Sekretion von HRP zeigen, werden durch Medium-Throughput-Screening ausgewählt. Anschließend wurden Integrationsorte mittels „Template blocking PCR method for genome walking“ identifiziert und näher durch die Schaffung von sauberen Deletionen der identifizierten Gene im *P. pastoris* CBS 7435 *Δhis4* Stammhintergrund charakterisiert. Auf diese Weise konnte eine Anzahl von höchst interessanten Genen identifiziert werden, die Einfluß auf die Sekretion einer Reihe von nicht verwandten rekombinanten Proteinen zeigen.

Introduction and Outline

Pichia pastoris, a methylotrophic yeast, has been genetically engineered to express thousands of heterologous proteins both intracellularly and extracellularly for a variety of reasons. Special features of this yeast such as the ability to grow to high cell densities on cheap media, the availability of tightly regulated and constitutive promoters, the ability to carry out post-translational protein modifications, the ease of handling and genetic manipulations as well as commercially available expression vectors and strains make it an ideal host for recombinant protein production ***(1–3)***.

In 1970s, Philips Petroleum Company developed high density fermentation protocols for *P. pastoris* using methanol as a sole carbon and energy source. The company wanted to develop this yeast as single cell protein additive for animal feed. However, increase in the production costs of methanol, due to the oil crisis of 1973, made this process uneconomical. In 1980s, with the advent of molecular techniques to engineer yeast, *P.pastoris* was developed as a host for recombinant protein production using the strong and tightly regulated alcohol oxidase 1 promoter (P_{AOX1}) ***(4, 5)***. This promoter is repressed when cells are grown on glucose, glycerol or ethanol and becomes de-repressed on depletions of these carbon sources. Alcohol oxidase is the first enzyme in the methanol assimilation pathway, which is encoded by two genes i.e., *AOX1* and *AOX2*. Due to the difference in the regulatory sequences of both genes, the majority of alcohol oxidase enzyme is expressed from *AOX1*. After induction with methanol ~5% of poly-adenylated RNA and 35% of total cellular protein are derived from *AOX1* promoter. Therefore, deletion of the *AOX1* gene greatly reduces the ability of *P. pastoris* to metabolize methanol, as cells have to rely on weaker expression levels from the *AOX2* gene, resulting in methanol utilization slow (Mut^S) phenotype ***(6–8)***. Better expression levels of heterologous proteins have been reported from strains showing

this phenotype, presumably due to the reduced growth rate ***(9)***. The oxidation of methanol into formaldehyde and hydrogen peroxide takes place inside specialized organelles called peroxisomes. When cells are shifted from other carbon sources to methanol, as a sole carbon and energy source, these organelles proliferate and vice versa. Therefore, *P.pastoris* has been used extensively to study peroxisome biogenesis and function ***(10, 11)***.

During the past decade, a number of milestones have been achieved with the *P.pastoris* expression system i.e., humanization of the glycosylation pathway ***(12–17)***, sequencing of the genome ***(18–20)***, availability of engineered promoters for fine-tuned gene expression ***(7, 21, 22)***, GRAS (generally regarded as safe) status from the Food and Drug Administration (FDA), and FDA approval of recombinant biopharmaceuticals (Kalbitor®, a kallikrein inhibitor and Jetrea, a treatment for vitreomacular traction).

Chapter 1, a review paper, summarizes the classical and novel developments of expression vectors with special emphasis on the newly developed constitutive and inducible promoter systems. Additionally, few aspects of secretion as well as different strategies applied for the improvement of recombinant protein secretion from this yeast are discussed. The second part of this review describes developments of engineered host strains i.e., auxotrophic strains, protease deficient strains and glycol-engineered strains for recombinant protein production. The third part of this review, discusses different strategies available for screening of transformants for high level expression and describes the application of the *P. pastoris* expression system for production of industrial, biopharmaceutical and membrane proteins. Lastly are provided future perspectives for this expression system keeping in mind the new developments in cell and molecular biology ***(2)***.

Selection of expression vectors and host strains play a critical role in protein expression experiments. Several choices have to be made in order to avoid pitfalls and delays later on

regarding recombinant protein production in *P. pastoris*. Which expression vector or promoter would be suitable for a particular protein? Which selection marker should be used? Should the protein be expressed intracellularly or should it be secreted? What kind of post translation modification does a particular protein require? Which marker is suitable for multicopy selection? These are some of the important questions that should be answered before embarking on any protein expression experiment. C**hapter 2** addresses these questions and provides a comprehensive list of classical and novel expression vectors and host strains for heterologous protein expression. Additionally, it also describes methods and strategies for generating Mut^{+}, Mut^{S} and multicopy strains.

Chapter 3, describes the construction of novel '*Pichia* pool' expression vectors –for intracellular and secretory expression of recombinant proteins – based on the CBS 7435 strain background. For this strain, patent protection has expired and there are no material rights pending, therefore, expression strains and vectors based on it can be used for commercial applications without paying licensing fees. We have used these newly constructed expression vectors to compare the potential of the α-mating factor secretory leader sequences of *P. pastoris* (*Pp*_αMF) and of S. cerevisiae (*Sc*_αMF) using *B. subtilis* levanase and horseradish peroxidase as reporter proteins. These studies show that *Pp*_αMF is inferior compared to *Sc*_αMF in directing protein secretion from this yeast, at least for the used reporter proteins. Furthermore, it appears that the Glu-Ala spacer sequence plays a more important role for processing of the *Pp*_αMF.

Targeted gene knockouts plays a critical role in assigning specific functions to genes. In **chapter 4**, a simple yet potent system for creating knockout cassettes for gene deletion in *P. pastoris* is described. This system can be adapted to any target gene and strain background with a single cloning step. The cloning of homologous sequences for gene targeting has been made efficient by

recombinase like properties of the *SfiI* restriction enzyme ***(23)***. The selection marker for cassette integration can be eliminated by induced expression of Flippase recombination enzyme, resulting in marker free clean knockout strains. We applied our knockout system to delete genes involved in biosynthetic pathway i.e., *LYS2* [29], *MET2* ***(25)***, *TYR1* ***(26)***, *SUB2* ***(27)***, *PEP4, PRB1* ***(28)*** *PRC1* ***(29)*** *YPS1, YPS2, YPS7* ***(30)***, *KEX1* ***(31)*** and *KEX2* ***(32)***. We describe for the first time the targeting of putative proteases *PrtP, CTSE, KPX1-KPX9 (Knockout Protease X)* in addition to biosynthetic genes *PHA2* and *PRO3*, in the latter case creating proline auxotrophy in *P. pastoris.* Deletion of *PHA2*, the gene encoding the key enzyme for phenylalanine biosynthesis in *S. cerevisiae* ***(33)***, resulted in a bradytroph phenotype. Therefore, it seems that an alternative but less efficient biosynthetic pathway for phenylalanine exits in *P. pastoris*. Contrary to previous reports, we were not able to achieve *PEP4* and *KEX2* deletion using Zeocin™ as a selection marker ***(32, 34)***. We reasoned that deletion of these protease genes reduces the viability of knockout strains, which results in increased sensitivity to Zoecin™ antibiotic. We were able to delete these genes by substituting the Zeocin™ with *HIS4* selection marker. Furthermore, a fast pooling method to identify the rare recombination events for multiple gene deletions in parallel was developed. By adding innovative details to previously defined approaches, gene targeting in *P. pastoris* has become easier and less laborious for researchers.

Over the years, several different promoters, secretion signals, N and C-terminal tags have been used to facilitate recombinant protein expression, secretion, detection and purification in *P. pastoris.* It is almost impossible to predict beforehand, which promoter, secretion signal or tag would work best for a particular protein. **Chapter 5** describes a novel restriction site free cloning (RSFC) strategy based on type II S restriction enzymes. This strategy allows a simple, seamless cloning of a single PCR product irrespective of upstream or downstream DNA sequences. Based

on this strategy, a set of 40 expression vectors was constructed for testing the effects of promoters, secretion signals, N and C-terminal tags on recombinant protein production in *P. pastoris*. This strategy could also be easily applied to other prokaryotic and eukaryotic expression systems. The expression vectors constructed during this study feature different promoters (P_{AOX1}, P_{GAP}), *Sc*_αMF variants (with and without Glu-Ala repeats), fusion partners (eGFP, MBP), N and C-terminal tags (Myc, FLAG, His, Strep) provided for intracellular and secretory expression. Additionally the effects of these features on secretion of horseradish peroxidase from *P. pastoris* are demonstrated ***(35)***.

C**hapter 6** reports on a novel insertion mutagenesis method to randomly target and disrupt genes in *Pichia pastoris*. The developed method was applied to identify genomic loci involved in the secretion of proteins in this yeast. Briefly, a Zeocin™ resistance cassette, with no apparent homology to the *P. pastoris* genome, was used to disrupt genomic loci in a *P. pastoris* strain expressing horseradish peroxidase as a reporter gene. Medium throughput screening of Zeocin™ positive transformants identified a number of genomic loci, which effect recombinant protein secretion. It is shown for the first time that deletion of Rim pathway genes (*Rim101*, *Rim20*, *Rim13*), which are responsible for response to alkaline pH in *S. cerevisiae*, and *Sgt2*, a member of the GET-complex, increases secretion of HRP. We also identified an uncharacterized protein, *Kep1* (Knockout enhances protein secretion 1), which upon deletion increases the secretion of multiple unrelated recombinant proteins i.e., HRP, alternative pig liver esterase (APLE), and human growth hormone (hGH).

References:

1. J.L. Cereghino and J.M. Cregg (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*., FEMS microbiology reviews. 24, 45–66.

2. M. Ahmad, M. Hirz, H. Pichler, et al. (2014), Protein expression in *Pichia pastoris*: Recent achievements and perspectives for heterologous protein production, http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4047484&tool=pmcentrez&rendertype=abstract.

3. S. Macauley-Patrick, M.L. Fazenda, B. McNeil, et al. (2005) Heterologous protein production using the *Pichia pastoris* expression system., Yeast (Chichester, England). 22, 249–270.

4. J.M.M. Cregg, K.J.J. Barringer, A.Y. Hessler, et al. (1985) *Pichia pastoris* as a host system for transformations., Molecular and Cellular …. 5, 3376–3385.

5. M. Hasslacher, M. Schall, M. Hayn, et al. (1997) High-level intracellular expression of hydroxynitrile lyase from the tropical rubber tree *Hevea brasiliensis* in microbial hosts., Protein expression and purification. 11, 61–71.

6. J.M. Cregg, K.R. Madden, K.J. Barringer, et al. (1989) Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*., Molecular and cellular biology. 9, 1316–1323.

7. F.S. HartnerF.S., C. Ruth, D. Langenegger, et al. (2008) Promoter library designed for fine-tuned gene expression in *Pichia pastoris*., Nucleic acids research. 36, e76.

8. J.F. Tschopp, P.F. Brust, J.M. Cregg, et al. (1987) Expression of the *lacZ* gene from two methanol-regulated promoters in *Pichia pastoris*, Nucleic Acids Research. 15, 3859–3876.

9. F.W. Krainer, C. Dietzsch, T. Hajek, et al. (2012) Recombinant protein expression in *Pichia pastoris* strains with an engineered methanol utilization pathway., Microbial cell factories. 11, 22.

10. I.J. van der Klei, H. Yurimoto, Y. Sakai, et al. (2006) The significance of peroxisomes in methanol metabolism in methylotrophic yeast, Biochimica et Biophysica Acta - Molecular Cell Research. 1763, 1453–1462.

11. M. a Johnson, H.R. Waterham, G.P. Ksheminska, et al. (1999) Positive selection of novel peroxisome biogenesis-defective mutants of the yeast *Pichia pastoris*., Genetics. 151, 1379–1391.

12. S. Wildt and T.U. Gerngross (2005) The humanization of N-glycosylation pathways in yeast., Nature reviews. Microbiology. 3, 119–128.

13. H. Li, N. Sethuraman, T. a Stadheim, et al. (2006) Optimization of humanized IgGs in glycoengineered *Pichia pastoris*., Nature biotechnology. 24, 210–215.

14. S.R. Hamilton, R.C. Davidson, N. Sethuraman, et al. (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins., Science (New York, N.Y.). 313, 1441–1443.

15. B.-K. Choi, P. Bobrowicz, R.C. Davidson, et al. (2003) Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast *Pichia pastoris*., Proceedings of the National Academy of Sciences of the United States of America. 100, 5022–5027.

16. R.K. Bretthauer (2003) Genetic engineering of Pichia pastoris to humanize N-glycosylation of proteins., Trends in biotechnology. 21, 459–462.

17. S.R. Hamilton and T.U. Gerngross (2007) Glycosylation engineering in yeast: the advent of fully humanized yeast., Current opinion in biotechnology. 18, 387–392.

18. A. Küberl, J. Schneider, G.G. Thallinger, et al. (2011) High-quality genome sequence of *Pichia pastoris* CBS7435., Journal of biotechnology. 154, 312–320.

19. K. De Schutter, Y.-C. Lin, P. Tiels, et al. (2009) Genome sequence of the recombinant protein production host *Pichia pastoris*., Nature biotechnology. 27, 561–566.

20. D. Mattanovich, A. Graf, J. Stadlmann, et al. (2009) Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*., Microbial cell factories. 8, 29.

21. X. Qin, J. Qian, G. Yao, et al. (2011) GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*., Applied and environmental microbiology. 77, 3600–3608.

22. E. Nevoigt, J. Kohnke, C.R. Fischer, et al. (2006) Engineering of promoter replacement cassettes for fine-tuning of gene expression in *Saccharomyces cerevisiae*., Applied and environmental microbiology. 72, 5266–73.

23. L.M. Wentzell, T.J. Nobbs, and S.E. Halford (1995) The *SfiI* restriction endonuclease makes a four-strand DNA break at two copies of its recognition sequence., Journal of molecular biology. 248, 581–595.

24. R.J. Austin, R.E. Kuestner, D.K. Chang, et al. (2011) SILAC compatible strain of *Pichia pastoris* for expression of isotopically labeled protein standards and quantitative proteomics, Journal of Proteome Research. 10, 5251–5259.

25. D. Thor, S. Xiong, C.C. Orazem, et al. (2005) Cloning and characterization of the *Pichia pastoris* MET2 gene as a selectable marker., FEMS yeast research. 5, 935–942.

26. M.M. Whittaker and J.W. Whittaker (2005) Construction and characterization of *Pichia pastoris* strains for labeling aromatic amino acids in recombinant proteins., Protein expression and purification. 41, 266–74.

27. K. Salamin, D. Sriranganadane, B. Léchenne, et al. (2010) Secretion of an endogenous subtilisin by *Pichia pastoris* strains GS115 and KM71, Applied and Environmental Microbiology. 76, 4269–4276.

28. M. a Gleeson, C.E. White, D.P. Meininger, et al. (1998) Generation of protease-deficient strains and their use in heterologous protein expression., Methods in molecular biology (Clifton, N.J.). 103, 81–94.

29. H. Ohi, W. Ohtani, N. Okazaki, et al. (1996) Cloning and characterization of the *Pichia pastoris PRC1* gene encoding carboxypeptidase Y, Yeast. 12, 31–40.

30. B. Guan, J. Lei, S. Su, et al. (2012) Absence of *Yps7p*, a putative glycosylphosphatidylinositol-linked aspartyl protease in *Pichia pastoris*, results in aberrant cell wall composition and increased osmotic stress resistance., FEMS yeast research. 12, 969–979.

31. T. Boehm, S. Pirie-Shepherd, L.B. Trinh, et al. (1999) Disruption of the *KEX1* gene in *Pichia pastoris* allows expression of full-length murine and human endostatin., Yeast (Chichester, England). 15, 563–572.

32. M.W.T. Werten and F. a. De Wolf (2005) Reduced proteolysis of secreted gelatin and *Yps1*-mediated alpha-factor leader processing in a *Pichia pastoris kex2* disruptant., Applied and environmental microbiology. 71, 2310–2317.

33. G.H. Braus (1991) Aromatic amino acid biosynthesis in the yeast *Saccharomyces cerevisiae*: a model system for the regulation of a eukaryotic biosynthetic pathway., Microbiological reviews. 55, 349–70.

34. R. Pan, J. Zhang, W.-L.L. Shen, et al. (2011) Sequential deletion of *Pichia pastoris* genes by a self-excisable cassette, FEMS Yeast Research. 11, 292–298.

35. T. Vogl, M. Ahmad, F.W. Krainer, et al. (2015) Restriction site free cloning (RSFC) plasmid family for seamless, sequence independent cloning in *Pichia pastoris*., Microbial cell factories. 14, 103.

Chapter 1

Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production

Mudassar Ahmad[1*], Melanie Hirz[1*], Harald Pichler[1,2] and Helmut Schwab[1,2]

[1]*Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, A-8010 Graz, Austria,*

[2]*Austrian Centre of Industrial Biotechnology (ACIB), Petersgasse 14, A-8010 Graz, Austria*

[*]***equal contributions***

Contribution:

MA*, MH*, HP and HS wrote the manuscript (*equal contribution). All Authors read and approved the final manuscript.

Appl Microbiol Biotechnol (2014) 98:5301–5317
DOI 10.1007/s00253-014-5732-5

MINI-REVIEW

Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production

Mudassar Ahmad · Melanie Hirz · Harald Pichler · Helmut Schwab

Received: 27 January 2014 / Revised: 25 March 2014 / Accepted: 26 March 2014 / Published online: 18 April 2014

Abstract *Pichia pastoris* is an established protein expression host mainly applied for the production of biopharmaceuticals and industrial enzymes. This methylotrophic yeast is a distinguished production system for its growth to very high cell densities, for the available strong and tightly regulated promoters, and for the options to produce gram amounts of recombinant protein per litre of culture both intracellularly and in secretory fashion. However, not every protein of interest is produced in or secreted by *P. pastoris* to such high titres. Frequently, protein yields are clearly lower, particularly if complex proteins are expressed that are hetero-oligomers, membrane-attached or prone to proteolytic degradation. The last few years have been particularly fruitful because of numerous activities in improving the expression of such complex proteins with a focus on either protein engineering or on engineering the protein expression host *P. pastoris*. This review refers to established tools in protein expression in *P. pastoris* and highlights novel developments in the areas of expression vector design, host strain engineering and screening for high-level expression strains. Breakthroughs in membrane protein expression are discussed alongside numerous commercial applications of *P. pastoris* derived proteins.

Keywords Yeast · *Pichia pastoris* · Protein expression · Protein secretion · Protease-deficient strains · Chaperone

Mudassar Ahmad and Melanie Hirz contributed equally to this work.

M. Ahmad · M. Hirz · H. Pichler · H. Schwab (✉)
Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14/5, 8010 Graz, Austria
e-mail: helmut.schwab@tugraz.at

H. Pichler · H. Schwab
Austrian Centre of Industrial Biotechnology (ACIB), Petersgasse 14, 8010 Graz, Austria

Introduction

The methylotrophic yeast *Pichia pastoris*, currently reclassified as *Komagataella pastoris*, has become a substantial workhorse for biotechnology, especially for heterologous protein production (Kurtzman 2009). It was introduced more than 40 years ago by Phillips Petroleum for commercial production of single cell protein (SCP) as animal feed additive based on a high cell density fermentation process utilizing methanol as carbon source. However, the oil crisis in 1973 increased the price for methanol drastically and made SCP production uneconomical. In the 1980s, *P. pastoris* was developed as a heterologous protein expression system using the strong and tightly regulated *AOX1* promoter (Cregg et al. 1985). In combination with the already developed fermentation process for SCP production, the *AOX1* promoter provided exceptionally high levels of heterologous proteins. One of the first large-scale industrial production processes established in the 1990s was the production of the plant-derived enzyme hydroxynitrile lyase at >20 g of recombinant protein per litre of culture volume (Hasslacher et al. 1997). This enzyme is used as biocatalyst for the production of enantiopure *m*-phenoxybenzaldehyde cyanohydrin — a building block of synthetic pyrethroids — on the multi-ton scale.

Through a far-sighted decision this expression system, initially patented by Phillips Petroleum, was made available to the scientific community for research purposes. A major breakthrough was the publication of detailed genome sequences of the original SCP production strain CBS7435 (Küberl et al. 2011), the first host strain developed for heterologous protein expression GS115 (De Schutter et al. 2009), as well as of the related *P. pastoris* DSMZ 70382 strain (Mattanovich et al. 2009b). Equally important breakthroughs for the commercial application of the *P. pastoris* cell factory were the Food and Drug Administration (FDA) GRAS (generally recognized as safe) status for a protein used in animal

feed, phospholipase C (Ciofalo et al. 2006), and the FDA approval of a recombinant biopharmaceutical product, Kalbitor®, a kallikrein inhibitor (Thompson 2010).

The classical *P. pastoris* expression system has been extensively reviewed over the years (Cereghino and Cregg 2000; Daly and Hearn 2005; Gasser et al. 2013; Jin et al. 2006; Macauley-Patrick et al. 2005). In this review, we focus on recent developments for heterologous protein production and describe examples for the commercial use of this expression system. In the first chapter, we refer to the established basic vector systems and elaborate on developments thereof with an emphasis on newly developed promoter systems. Herein, also some aspects of secretion will be summarized. The second part is devoted to the most recent developments regarding host strain development. As a specific novelty, a new platform based on the CBS7435 strain is described, for which patent protection has ceased and no specific material rights are pending. In the third chapter, we describe specific strategies for obtaining high-level expression strains and summarize important applications of *P. pastoris* for production of biopharmaceuticals, membrane proteins and industrial proteins. The last section provides an outlook on future perspectives covering recent progress in molecular and cell biology of *P. pastoris* and possibilities for implementing new strategies in expression strain development.

Basic systems for cloning and expression in *P. pastoris*

When devising strategies for cloning and expression of heterologous proteins in *P. pastoris* some points need to be considered from the start, that is, the choice of promoter–terminator combinations, suitable selection markers and application of vector systems for either intracellular or secreted expression including selection of proper secretion signals (Fig. 1). The choice of the proper expression vector and complementary host strain are a most important prerequisite for successful recombinant protein expression.

Promoters

The use of tightly regulated promoters such as the alcohol oxidase (*AOX1*) promoter holds advantages for overexpression of proteins. By uncoupling the growth from the production phase, biomass is accumulated prior to protein expression. Therefore, cells are not stressed by the accumulation of recombinant protein during growth phase, and even the production of proteins that are toxic to *P. pastoris* is possible. Furthermore, it may be desirable to co-express helper proteins like chaperones at defined time points, for example, before the actual target protein is formed. On the other hand, use of constitutive promoters may ease process handling. Constitutive promoters are usually also applied to express selection markers. Metabolic pathway engineering strategies might further take advantage of fine-tuned constitutive promoters to ensure a controlled flux of metabolites. An extensive summary of promoters used for heterologous expression in *P. pastoris* has recently been published by Vogl and Glieder (2013). An overview of broadly used and extensively studied as well as recently examined promoters is given in Table 1.

Inducible promoters

The tightly regulated *AOX1* promoter (P_{AOX1}), which was first employed for heterologous gene expression by Tschopp et al. (1987a), is still the most commonly used promoter (Lünsdorf et al. 2011; Sigoillot et al. 2012; Yu et al. 2013). P_{AOX1} is strongly repressed when *P. pastoris* is grown on glucose, glycerol or ethanol (Inan and Meagher 2001). Upon depletion of these carbon sources, the promoter is de-repressed, but is fully induced only upon addition of methanol. Several studies have identified multiple regulatory elements in the P_{AOX1} sequence (Hartner et al. 2008; Kranthi et al. 2006, 2009; Ohi et al. 1994; Parua et al. 2012; Staley et al. 2012; Xuan et al. 2009). Positively and negatively acting elements have been described (Kumar and Rangarajan 2012; Lin-Cereghino et al. 2006; Polupanov et al. 2012), but the molecular details of P_{AOX1} regulation are still not completely elucidated.

Methanol is a highly flammable and hazardous substance and, therefore, undesirable for large-scale fermentations. Alternative inducible promoters or P_{AOX1} variants, which can be induced without methanol but still reach high expression levels, are desired. A recently published patent application describes such a method, wherein expression is controlled by methanol-inducible promoters, such as *AOX1*, methanol oxidase (*MOX*) or formate dehydrogenase (*FMDH*), without the addition of methanol (Takagi et al. 2008). This was achieved by constitutively co-expressing the positively acting transcription factor Prm1p from either of the *GAP*, *TEF* or *PGK* promoters. The relative activity of a phytase reporter protein was 3-fold increased without addition of methanol as compared to a control strain with *PRM1* under its native promoter. However, phytase expression levels were not compared for standard methanol induction and constitutive Prm1p expression conditions. Hartner et al. have constructed a synthetic *AOX1* promoter library by deleting or duplicating transcription factor binding sites for fine-tuned expression in *P. pastoris* (Hartner et al. 2008). Using EGFP as reporter, some promoter variants were found to confer even higher expression levels than the native P_{AOX1} spanning a range between 6 % and 160 % of the native promoter activity. These P_{AOX1} variants have also proven to behave similarly when industrially relevant enzymes such as horseradish peroxidase and hydroxynitrile lyases were expressed.

Numerous further controllable promoters are currently being investigated for their ability to promote high-level

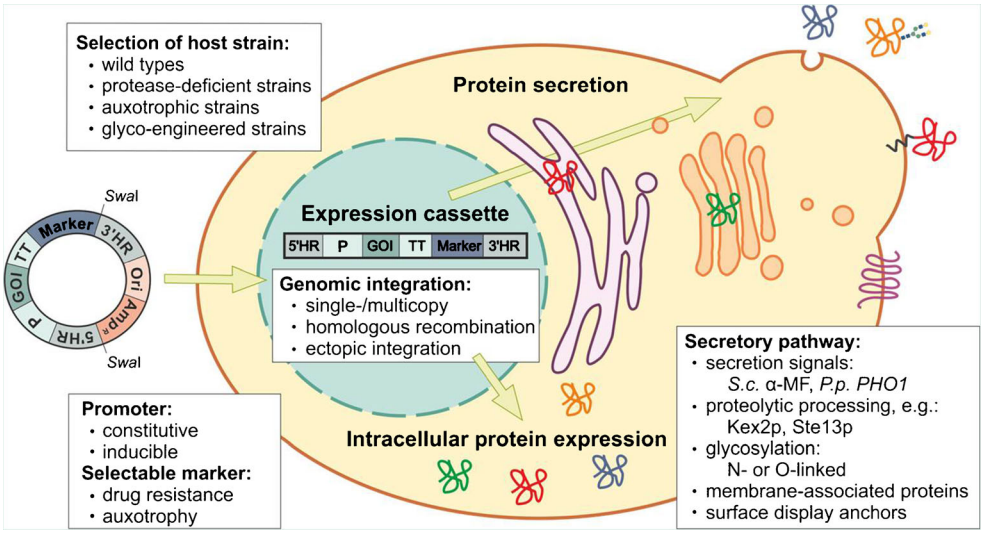


Fig. 1 General considerations for heterologous gene expression in *P. pastoris*. Expression plasmids harbouring the gene(s) of interest (*GOI*) are linearized prior to transformation. Selectable markers (e.g., Amp^R) and origin of replication (*Ori*) are required for plasmid propagation in *E. coli*. The expression level of the protein of interest may depend on (i) the chromosomal integration locus, which is targeted by the 5′ and 3′ homologous regions (5′HR and 3′HR), and (ii) on the gene copy number. A representative promoter (*P*) and transcription terminator (*TT*) pair are shown. Proper signal sequences will guide recombinant protein for intracellular or secretory expression, and will govern membrane integration or membrane anchoring

expression (Table 1). For example, a recently published patent application describes the use of three novel inducible promoters from *P. pastoris*, *ADH1* (alcohol dehydrogenase), *GUT1* (glycerol kinase) and *ENO1* (enolase), showing interesting regulatory features (Cregg and Tolstorukov 2012). However, due to a lack of absolute expression values the performance of these novel promoters cannot be compared to the widely used *AOX1* and *GAP* promoters.

Table 1 The most prominently used and very recently established promoters for heterologous expression in *P. pastoris*

Inducible	Corresponding gene	Regulation	Reference
AOX1	Alcohol oxidase 1	Inducible with MeOH	(Tschopp et al. 1987a)
DAS	Dihydroxyacetone synthase	Inducible with MeOH	(Ellis et al. 1985; Tschopp et al. 1987a)
FLD1	Formaldehyde dehydrogenase 1	Inducible with MeOH or methylamine	(Shen et al. 1998)
ICL1	Isocitrate lyase	Repressed by glucose, induction in absence of glucose/by addition of ethanol	(Menendez et al. 2003)
PHO89	Putative Na^+/phosphate symporter	Induction upon phosphate starvation	(Ahn et al. 2009)
THI11	Thiamine biosynthesis gene	Repressed by thiamin	(Stadlmayr et al. 2010)
ADH1	Alcohol dehydrogenase	Repressed on glucose and methanol, induced on glycerol and ethanol	(Cregg and Tolstorukov 2012)
ENO1	Enolase	Repressed on glucose, methanol and ethanol, induced on glycerol	(Cregg and Tolstorukov 2012)
GUT1	Glycerol kinase	Repressed on methanol, induced on glucose, glycerol and ethanol	(Cregg and Tolstorukov 2012)
Constitutive	**Corresponding gene**	**Regulation**	**Reference**
GAP	Glyceraldehyde-3-P dehydrogenase	Constitutive expression on glucose, to a lesser extent on glycerol and methanol	(Waterham et al. 1997)
TEF1	Translation elongation factor 1	Constitutive expression on glycerol and glucose	(Ahn et al. 2007)
PGK1	3-Phosphoglycerate kinase	Constitutive expression on glucose, to a lesser extent on glycerol and methanol	(de Almeida et al. 2005)
GCW14	Potential glycosyl phosphatidyl inositol (GPI)-anchored protein	Constitutive expression on glycerol, glucose and methanol	(Liang et al. 2013b)
G1	High affinity glucose transporter	Repressed on glycerol, induced upon glucose limitation	(Prielhofer et al. 2013)
G6	Putative aldehyde dehydrogenase	Repressed on glycerol, induced upon glucose limitation	(Prielhofer et al. 2013)

Constitutive promoters

Constitutive expression eases process handling, omits the use of potentially hazardous inducers and provides continuous transcription of the gene of interest. For this purpose, the glyceraldehyde-3-phosphate promoter (P_{GAP}) is commonly used, which — on glucose — reaches almost the same expression levels as methanol-induced P_{AOX1} (Waterham et al. 1997). Expression levels from P_{GAP} drop to about one half on glycerol and to one third when cells are grown on methanol (Cereghino and Cregg 2000). Alternative constitutive promoters and promoter variants have been described recently (Table 1). The constitutive P_{GCW14} promoter, for example, was described to be a stronger promoter than the *GAP* and *TEF1* promoters, which was assessed by secretory expression of EGFP (Liang et al. 2013b). It was found that EGFP expression from P_{GCW14} yielded in a 10-fold increase compared to P_{GAP} driven expression when cells were cultivated on glycerol or methanol, and a 5-fold increase on glucose.

A recent DNA microarray study identified novel promoters that are repressed on glycerol, but are being induced upon shift to glucose-limited media (Prielhofer et al. 2013). Supposedly, the most interesting promoters discovered by this approach control expression of a high-affinity glucose transporter, *HGT1*, and of a putative aldehyde dehydrogenase. The former promoter was reported to drive EGFP expression to even higher levels than could be reached with P_{GAP}. In glycerol fed-batch fermenter cultures, human serum album was expressed from the novel promoter to a 230 % increase in specific product yield as compared to P_{GAP} driven expression.

In some cases, it is desired that expression levels can be fine-tuned in order to (1) co-express accessory proteins facilitating recombinant protein expression and secretion or (2) provide protein post-translational modifications as well as to (3) engineer whole metabolic pathways consisting of a cascade of different enzymatic steps. For such applications, a library of *GAP* promoter variants with relative strengths ranging from 0.6 % to 16.9-fold of the wild type promoter activity was developed and tested using three different reporter proteins, yEGFP, β-galactosidase and methionine acetyltransferase (Qin et al. 2011).

Vectors

The standard setup of vectors is a bi-functional system enabling replication in *E. coli* and maintenance in *P. pastoris* using as selection markers either auxotrophy markers (e.g., *HIS4*, *MET2*, *ADE1*, *ARG4*, *URA3*, *URA5*, *GUT1*) or genes conferring resistance to drugs such as Zeocin™, geneticin (G418) and blasticidin S. Although there are some reports of using episomal plasmids for heterologous protein expression or for the screening of mutant libraries in *P. pastoris* (Lee et al. 2005; Uchima and Arioka 2012), stable integration into the host genome is the most preferred method. Unlike in *Saccharomyces cerevisiae*, where homologous recombination (HR) predominates, non-homologous end-joining (NHEJ) is a frequent process in *P. pastoris.* The ratio of NHEJ and HR can be shifted towards HR by elongating the length of the homologous regions flanking the actual expression cassettes and by suppressing NHEJ efficiency (Näätsaari et al. 2012).

The standard vector systems for intracellular and secretory expression provided by Life Technologies (Carlsbad, CA, USA) include constitutive (P_{GAP}) and inducible promoters triggered by methanol or methylamine (P_{AOX1}, P_{FLD}). The recently introduced PichiaPink™ expression kit for intracellular or secreted expression enables easy selection of multicopy integration clones by differences in colour formation based on *ade2* knockout strains and truncated *ADE2* promoters of varying strengths in front of the *ADE2* marker gene (Du et al. 2012; Nett 2010).

Additionally, BioGrammatics (Carlsbad, CA, USA) holds licences for selling standard *P. pastoris* expression vectors and strains and also provides GlycoSwitch® vectors for humanized glycosylation of target proteins (Table 2). Several vectors for disruption of *OCH1* and expression of different glycosidases or glycosyltransferases are available to achieve mammalian-type N-glycan structures in *P. pastoris*. These vectors harbour, for example, the human GlcNAc transferase I, the mannosidase II from rat, or the human galactosyl transferase I. A detailed protocol for humanizing the glycosylation pattern using the GlycoSwitch® vectors is provided (Jacobs et al. 2009).

James Cregg's laboratory at the Keck Graduate Institute, Claremont, CA, USA, has developed a set of plasmids for protein secretion and intracellular expression in *P. pastoris* containing the strong *AOX1* promoter. These vectors are based on different auxotrophy markers, such as *ARG4*, *ADE1*, *URA3* and *HIS4*, for selection necessitating the use of the appropriate host strains (see section "Host strain development"). The vectors contain restriction sites for linearization within the marker genes to target the expression cassettes to the desired locus as well as for multicopy integration (Lin-Cereghino et al. 2001). Moreover, a set of integration vectors for sequential disruption of *ARG1*, *ARG2*, *ARG3*, *HIS1*, *HIS2*, *HIS5* and *HIS6* in *P. pastoris* was applied to provide the host strains for engineering the protein glycosylation pathway (Nett et al. 2005).

The Institute of Molecular Biotechnology, Graz University of Technology, Austria, provides vectors and strains to the *P. pastoris* community through the so-called '*Pichia* Pool'. The p*Pp* plasmids described by Näätsaari et al. (2012) comprise vectors containing the *GAP* or *AOX1* promoters and, for secretory expression, the *S. cerevisiae* α-mating factor (α-MF) secretion signal. The antibiotic selection marker cassettes were placed under the control of *ADH1* or *ILV5* promoters in the p*Pp*B1 and p*Pp*T4 vectors, respectively. It is

Table 2 Commercial vector systems

Supplier	Promoter	Signal sequences	Selection in yeast	Selection in bacteria	Comments
Life Technologies™	*AOX1, FLD1, GAP*	*S. cerevisiae* α-MF; *P. pastoris PHO1*	Blasticidin, G418, Zeocin™, *HIS4*	Zeocin™, Ampicillin, Blasticidin	c-myc epitope, V5 epitope, C-terminal 6× His-tag available for detection/purification
Life Technologies –PichiaPink™	*AOX1*	α-MF; set of eight different signal sequences – not ready to use[a]	*ADE2*	Ampicillin	Low- and high-copy vectors available, *TRP2* sequence for targeting
BioGrammatics	*AOX1*	α-MF	Zeocin™, G418, Nourseothricin	Ampicillin	Intracellular or secreted expression
BioGrammatics – GlycoSwitch®	*GAP*	–	Zeocin™, G418, Hygromycin, *HIS4*, Nourseothricin	Zeocin™, Ampicillin, Kanamycin, Nurseothricin	Human GlcNAc transferase I, rat Mannosidase II, human Gal transferase I
DNA2.0	*AOX1*	Ten different signal sequences – ready to use[b]	Zeocin™, G418	Zeocin™, Ampicillin	Intracellular or secreted

[a] The different secretion signals have to be cloned into the vector by a three-way ligation step

[b] The α-MF secretion signal is provided once with Kex2p (KR) and Ste13p cleavage sites (EAEA), once lacking EA repeats, and once as truncated version (pre-region only)

described that the p*Pp*T4-based vectors usually lead to lower gene copies in the cell as compared to the p*Pp*B1-based vectors.

Further vectors based on either the *GAP* or the *AOX1* promoter and a series of strains have recently been added to this pool, both for intracellular and secretory protein expression (M. Ahmad, unpublished results). For intracellular expression, cloning of the target genes is accomplished by using *Eco*RI and *Not*I, whereby the Kozak consensus sequence has to be restored for efficient translation initiation (Fig. 2a). A special characteristic of these vectors is that the *Eco*RI site has been introduced by a single point mutation directly into the *AOX1* promoter sequence without changing the promoter activity. Thereby, the gene of interest may be fused to the promoter without having additional nucleotides between the promoter and the start codon. Another advantage is the use of the short *ARG4* promoter for the expression of the selection markers. The weaker *ARG4* promoter used for selection marker cassettes enables selection at lower concentrations of Zeocin™ (i.e., 25 instead of 100 μg/ml) without obtaining false-positive clones. For secretory expression governed by the *S. cerevisiae* α-MF signal sequence, *Xho*I and/or *Not*I sites are used for cloning the genes of interest (Fig. 2b).

Aspects of secretory expression

One of the main advantages of using *P. pastoris* as a protein production host is its ability to secrete high titres of properly folded, post-translationally processed and active recombinant proteins into the culture media. As a rule of thumb, proteins secreted in their native hosts will also be secreted in *P. pastoris*. However, there are also some reports of successful secretion of typically intracellular proteins such as GFP or human catalase (Eiden-Plach et al. 2004; Shi et al. 2007). The most commonly employed secretion signals in *P. pastoris* are derived from *S. cerevisiae* α-MF, *S. cerevisiae* invertase (*SUC2*) and the *P. pastoris* endogenous acid phosphatase (*PHO1*) (Daly and Hearn 2005). As listed in Table 2, commercial kits also provide vectors with different secretion signals, which allows for screening of the best-suited signal sequence.

The α-MF signal sequence is composed of a pre- and pro-region and has proven to be most effective in directing protein through the secretory pathway in *P. pastoris*. The pre-region is responsible for directing the nascent protein post-translationally into the endoplasmic reticulum (ER) and is cleaved off subsequently by signal peptidase (Waters et al. 1988). The pro-region is thought to play a role in transferring the protein from ER to Golgi compartment and is finally cleaved at the dibasic KR site by the endo-protease Kex2p (Julius et al. 1984). The two EA repeats are subsequently trimmed by the *STE13* gene product (Brake et al. 1984). One of the common problems encountered while using the α-MF secretion signal is non-homogeneity of the N-termini of the recombinant proteins due to incomplete *STE13* processing. Constructs without the EA repeats may enhance homogeneity at the N termini of recombinant proteins. However, the removal of these sequences may affect protein yield. While no reports on enhanced co-expression of *STE13* are available, co-overexpression of *HAC1*, a transcription factor in the unfolded protein response (UPR) pathway, with the membrane protein adenosine A2 receptor had a positive effect on proper processing of the α-MF signal sequence (Guerfal

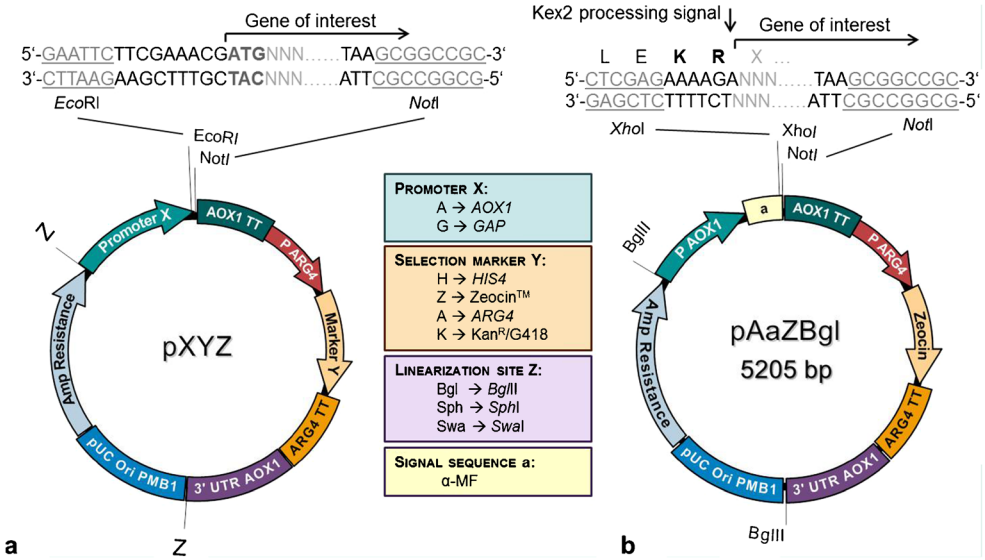


Fig. 2 Novel '*Pichia* Pool' plasmid sets for intracellular and secretory expression. **a** General features of pXYZ vector for intracellular expression. Letters refer to the choice of promoters (*X*), selection markers (*Y*), and restriction enzymes (*Z*) for linearization. Available elements are shown in boxes. The vector backbone harbours an ampicillin resistance marker and origin of replication for maintenance of the plasmid in *E. coli*. The GOI is *Eco*RI–*Not*I cloned directly after the promoter of choice. The Kozak consensus sequence for yeast (i.e., CGAAACG), should be restored between the *Eco*RI cloning site and the start codon of the GOI in order to achieve optimal translation. In addition, sequence variation within this region will allow fine-tuning translation initiation efficiency. Expression in *P. pastoris* is driven either by the methanol inducible *AOX1* or the constitutive *GAP* promoter. Positive clones can be selected for by antibiotic resistance (i.e., to Zeocin™ or geneticin sulphate) or by selection for His or Arg prototrophy. Selection marker expression is uniformly driven by the *ARG4* promoter–terminator pair. **b** Plasmid pAaZBgl from '*Pichia* Pool' is shown as an example of a vector made for secretory expression encoding *S. cerevisiae* α-MF signal sequence in front of the GOI cloning site. The Kex2 processing site AAAAGA should be restored between the *Xho*I cloning site and the fusion point of the GOI

et al. 2010). Recently, Yang et al. (2013) reported enhanced secretory protein production by optimizing the amino acid residues at the Kex2 P1′ site.

Multiple strategies have been followed to enhance the secretory potential of the α-MF signal sequence including codon optimization (Kjeldsen et al. 1998), directed evolution (Rakestraw et al. 2009), insertion of spacers and deletion mutagenesis (Lin-Cereghino et al. 2013). Directed evolution of the α-MF signal sequence in *S. cerevisiae* resulted in up to 16-fold enhanced full-length IgG_1 secretion as compared to the wild type. Furthermore, when this improved leader sequence was combined with strain engineering strategies comprising *PDI* overexpression and elimination of proteins involved in vacuolar targeting, up to 180-fold enhanced secretion of the reporter protein was observed (Rakestraw et al. 2009). Deletion mutagenesis based on a predicted structure model of α-MF signal peptide resulted in 50 % increased secretion of horseradish peroxidase and *C. antarctica* lipase B (CALB) in *P. pastoris* (Lin-Cereghino et al. 2013). It appears that decreasing the hydrophobicity of the leader sequence by deleting hydrophobic residues or substituting them with more polar or charged residues increased the flexibility of the α-MF signal sequence structure, which enhanced the overall secretory capacity of the pro-region. Alternative signal sequences used to direct protein secretion and their features and applications are summarized in Table 3.

Beyond the choice of the secretion signals there are several other factors that govern efficient protein secretion. The newly synthesized proteins are translocated co- or post-translationally into the ER lumen through the Sec61p translocon. Then, proteins may undergo one or several post-translational modifications, folding into the native state, disulphide-bond formation, glycosylation and membrane-anchoring. When the recombinant protein fails to fold into its native state or protein expression exceeds the folding capacity of the ER (Sha et al. 2013), unfolded proteins may start to aggregate, triggering the UPR pathway. UPR is responsible for induction of genes that are involved in protein folding. In parallel to UPR pathway, ER-associated degradation (ERAD) by the proteasome may relieve blocks in protein secretion (recently reviewed by Idiris et al. 2010 and Damasceno et al. 2012). Inappropriate mRNA structure and gene copy numbers, limits in transcription, translation and protein translocation into the ER, incomplete protein folding and inefficient protein targeting to the exterior of the cell are major bottlenecks encountered in secretory expression of heterologous proteins. Commonly used strategies to overcome such secretory bottlenecks comprise the overexpression of folding helper proteins like BiP/Kar2p, DnaJ, PDI, PPIs and Ero1p or, alternatively, overexpression of *HAC1*, a transcriptional regulator of the UPR pathway genes. Unlike in *S. cerevisiae*, Guerfal et al. (2010) reported that *HAC1* is

Table 3 Signal sequences used to secrete the protein into the extracellular space

Secretion signal	Source	Target protein(s)	Length	Reference
α-MF	*S.c.* α-mating factor	Most commonly used secretion signal in *P. pastoris*	85 aa, with or without EA repeats	(Brake et al. 1984)
PHO1	*P.p.* acid phosphatase	Mouse 5-HT5A, porcine pepsinogen,	15 aa	(Payne et al. 1995; Weiss et al. 1995; Yoshimasu et al. 2002)
SUC2	*S.c.* Invertase	Human interferon, α-amylase, α-1-antitrypsin	19 aa	(Moir and Dumais 1987; Paifer et al. 1994; Tschopp et al. 1987b)
PHA-E	Phytohemagglutinin	GNA, GFP and native protein	21 aa	(Raemaekers et al. 1999)
KILM1	Kl toxin	CM cellulase	44 aa	(Skipper et al. 1985)
pGKL	pGKL killer protein	Mouse α-amylase	20 aa	(Kato et al. 2001)
CLY and CLY-L8	C-lysozyme and syn. leucin-rich peptide	Human lysozyme	18 and 16 aa	(Oka et al. 1999)
K28 pre-pro-toxin	K28 virus toxin	Green fluorescent protein	36 aa	(Eiden-Plach et al. 2004)
Scw, Dse and Exg	*P.p.* Endogenous signal peptides	CALB and EGFP	19, 20 and 23 aa	(Liang et al. 2013a)
Pp Pir1	*P.p.* Pir1p	EGFP and Human α1-antitrypsin	61 aa	(Khasa et al. 2011)
HBFI and HBFII	Hydrophobins of *Trichoderma reesei*	EGFP	16 and 15 aa	(Kottmeier et al. 2011)

constitutively expressed and spliced in *P. pastoris* under normal growth conditions, which may explain the higher titers of secreted proteins obtainable with this organism. A contradictory observation was reported by Whyteside et al. (2011). Unspliced *HAC1* mRNA was detected under normal growth conditions and splicing of *HAC1* mRNA was only detected when cells were grown in presence of dithiothreitol (DTT) to activate the UPR. It should be mentioned, though, that sometimes overexpression of folding helpers actually reduced protein secretion or did not have any effect (van der Heide et al. 2002).

Host strain development

Elucidation of full genome sequences and gene annotation were great steps toward rational strain engineering, identifying new promoters and progressing in the (systems) biology of *P. pastoris* (Küberl et al. 2011; Mattanovich et al. 2009a; De Schutter et al. 2009). Two online databases (http://bioinformatics.psb.ugent.be/orcae/overview/Picpa and http://www.pichiagenome.org) provide convenient access to genome sequences and annotations. Frequently used commercially available strains are the *his4* strain GS115, the reconstituted prototrophic strain X-33, the *aox1* knockout strains KM71 and KM71H as well as protease-deficient strains SMD1168 and SMD1168H and the *ade2* auxotrophic PichiaPink™ strain. Use of these strains for commercial applications, however, is restricted by patent protection and/or materials ownership policy. Strains derived from *P. pastoris* CBS7435, in contrast, are not covered by patent protection and, therefore represent an alternative for production purposes. Furthermore, the CBS7435 Mut^S strain provided by the Graz *Pichia* Pool has the advantage of being marker-free as it was constructed using the Flp/FRT recombinase system for marker removal (Näätsaari et al. 2012). Using the same strategy, *ade1* and *his4* knockout strains were created along with the CBS7435 *ku70* strain (CBS 12694), which is impaired in the NHEJ mechanism, thereby enhancing the efficiency of HR. A selection of most relevant strains is compiled in Table 4.

Auxotrophic strains

Several auxotrophic strains (e.g., *ade1*, *arg4*, *his4*, *ura3*, *met2*), and combinations thereof are available together with vectors harbouring the respective genes as selectable markers (Lin-Cereghino et al. 2001; Thor et al. 2005, Graz *Pichia* Pool). Auxotrophic strains have been useful for in vivo labelling of proteins, for example in the global fluorination of *Candida antarctica* lipase B (CALB) in a *P. pastoris* X-33 *aro1* strain deficient in tryptophan, tyrosine, and phenylalanine biosynthesis (Budisa et al. 2010). Fluorinated analogues of these amino acids were supplemented and incorporated into the heterologous protein, thereby, for example, prolonging CALB shelf-life but lowering its lipase activity. The proteolytic pattern of CALB was retained, though. Another example is the use of a *lys2 arg4* double knockout strain for stable isotope labelling by amino acids in cell culture (SILAC) (Austin et al. 2011).

Protease-deficient strains

Undesired proteolysis of heterologous proteins expressed in *P. pastoris* does not only lower the product yield or biological

Table 4 *P. pastoris* host strains

Strain	Genotype	Phenotype	Source
Wild-type strains			
CBS7435 (NRRL Y-11430)	WT	WT	Centraalbureau voor Schimmelcultures, the Netherlands
CBS704 (DSMZ 70382)	WT	WT	Centraalbureau voor Schimmelcultures, the Netherlands
X-33	WT	WT	Life Technologies™
Auxotrophic strains			
GS115	*his4*	His^-	Life Technologies™
PichiaPink™ 1	*ade2*	Ade^-	Life Technologies™
KM71	*his4, aox1::ARG4, arg4*	His^-, Mut^S	Life Technologies™
KM71H	*aox1::ARG4, arg4*	Mut^S	Life Technologies™
BG09	*arg4::nourseo*^{R} *Δlys2::hyg*^{R}	Lys^-, Arg^-, $Nourseothricin^R$, $Hygromycin^R$	BioGrammatics
GS190	*arg4*	Arg^-	(Cregg et al. 1998)
GS200	*arg4 his4*	His^-, Arg^-	(Waterham et al. 1996)
JC220	*ade1*	Ade^-	(Cregg et al. 1998)
JC254	*ura3*	Ura^-	(Cregg et al. 1998)
JC227	*ade1 arg4*	Ade^- Arg^-	(Lin-Cereghino et al. 2001)
JC300-JC308	Combinations of *ade1 arg4 his4 ura3*	Combinations of Ade^-, Arg^-, His^-, Ura^-	(Lin-Cereghino et al. 2001)
YJN165	*ura5*	Ura^-	(Nett and Gerngross 2003)
CBS7435 *his4*[a]	*his4*	His^-	(Näätsaari et al. 2012)
CBS7435 Mut^S *his4*[a]	*aox1, his4*	Mut^S, His^-	(Näätsaari et al. 2012)
CBS7435 Mut^S *arg4*[a]	*aox1, arg4*	Mut^S, Arg^-	(Näätsaari et al. 2012)
CBS7435 *met2*[a]	*met2*	Met^-	(*Pp*7030)[b]
CBS7435 *met2 arg4*[a]	*met2 arg4*	Met^- Arg^-	(*Pp*7031)[b]
CBS7435 *met2 his4*[a]	*met2 his4*	Met^- His^-	(*Pp*7032)[b]
CBS7435 *lys2*[a]	*lys2*	Lys^-	(*Pp*7033)[b]
CBS7435 *lys2 arg4*[a]	*lys2 arg4*	Lys^- Arg^-	(*Pp*7034)[b]
CBS7435 *lys2 his4*[a]	*lys2 his4*	Lys^- His^-	(*Pp*7035)[b]
CBS7435 *pro3*[a]	*pro3*	Pro^-	(*Pp*7036)[b]
CBS7435 *tyr1*[a]	*tyr1*	Tyr^-	(*Pp*7037)[b]
Protease-deficient strains			
SMD1163	*his4 pep4 prb1*	His^-	(Gleeson et al. 1998)
SMD1165	*his4 prb1*	His^-	(Gleeson et al. 1998)
SMD1168	*his4 pep4::URA3 ura3*	His^-	Life Technologies™
SMD1168H	*pep4*		Life Technologies™
SMD1168 kex1::SUC2	*pep4::URA3 kex1::SUC2 his4 ura3*	His^-	(Boehm et al. 1999)
PichiaPink 2-4	Combinations of *prb1/pep4*	Ade^-	Life Technologies™
BG21	*sub2*		BioGrammatics
CBS7435 *prc1*[a]	*prc1*		(*Pp*6676)[b]
CBS7435 *sub2*[a]	*sub2*		(*Pp*6668)[b]
CBS7435 *sub2*[a]	*his4 pep4*	His^-	(*Pp*6911)[b]
CBS7435 *prb1*[a]	*prb1*		(*Pp*6912)[b]
CBS7435 *his4 pep4 prb1*	*his4 pep4 prb1*	His^-	(*Pp*7013)[b]
Glyco-engineered strains			
$SuperMan_5$	*his4 och1::pGAPTrα1,2-mannosidase*	His^-, $Blasticidin^R$	BioGrammatics
	och1::pGAPTrα1,2-mannosidase	$Blasticidin^R$	BioGrammatics
	pep4 och1::pGAPTrα1,2-mannosidase	$Blasticidin^R$	BioGrammatics

Table 4 (continued)

Strain	Genotype	Phenotype	Source
Other strains			
GS241	*fld1*	Growth defect on methanol as sole C-source or methylamine as sole N-source	(Shen et al. 1998)
MS105	*his4 fld1*	See GS241; His^-	(Shen et al. 1998)
MC100-3	*his4 arg4 aox1::ScARG4 aox2::PpHIS4*	Mut^-	(Cregg et al. 1989)
CBS7435 *ku70* [a]	*ku70*	WT	(Näätsaari et al. 2012)
CBS7435 *ku70 his4* [a]	*ku70, his4*	His^-	(Näätsaari et al. 2012)
CBS7435 *ku70 gut1*	*ku70, gut1*	Growth defect on glycerol; $Zeocin^R$	(Näätsaari et al. 2012)
CBS7435 *ku70 ade1*	*ku70, ade1*	Ade^-, $Zeocin^R$	(Näätsaari et al. 2012)

[a] These *P. pastoris* CBS7435 derived strains are marker-free knockouts

[b] Strains from '*Pichia* Pool' of TU Graz (M. Ahmad, unpublished results)

activity, but also complicates downstream processing of the intact product as the degradation products will have similar physicochemical and affinity properties. Proteolysis may occur either during vesicular transport of recombinant protein by secretory pathway-resident proteases (Werten and de Wolf 2005; Ni et al. 2008) or in the extracellular space by proteases being secreted, cell wall-associated (Kang et al. 2000) or released into the culture medium as a result of cell disruption during high cell density cultivation (Sinha et al. 2005). Different strategies have been employed to address the proteolysis problem, namely, modifying fermentation parameters (pH, temperature and specific growth rate), changing the media composition (rich medium, addition of casamino acids or peptone as competing substrates), lowering the salt concentration and addition of soytone (Zhao et al. 2008), applying protein engineering strategies (Gustavsson et al. 2001) and engineering of the expression host to obtain protease-deficient strains (reviewed by Idiris et al. 2010 and Macauley-Patrick et al. 2005). However, in some cases, optimization of the fermentation media and protein engineering strategies failed to alleviate the proteolysis problem and tuning the expression host itself was the only viable option (Li et al. 2010). The use of protease-deficient strains such as SMD1163 (*Δhis4 Δpep4 Δprb1*), SMD1165 (*Δhis4 Δprb1*) and SMD1168 (*Δhis4 Δpep4*) has been well documented for the expression of protease-sensitive proteins (Gleeson et al. 1998). *PEP4* encodes a major vacuolar aspartyl protease which is able to activate itself as well as further proteases such as carboxypeptidase Y (*PRC1*) and proteinase B (*PRB1*). The use of protease-deficient strains other than the above mentioned (e.g., *yps1*, *kex1*, *kex2*) was reported with variable success (Ni et al. 2008; Werten and de Wolf 2005; Wu et al. 2013; Yao et al. 2009). A general conclusion from these studies is that in many cases several proteases are involved in degradation events and, therefore, it is not an easy task to optimize protein expression by knocking out just a single one. However, the *pep4* and *prb1* knockout strains are still the most effective ones in preventing recombinant protein degradation, and, hence, also the most widely applied. Although it has been reported that protease-deficient strains show typically slower growth rates, lower transformation efficiencies and reduced viability (Lin-Cereghino and Lin-Cereghino 2007), experiments in our laboratory showed robust growth behaviour of 28 protease-deficient strains that were recently created (M. Ahmad, unpublished results).

Glyco-engineered strains

When yeasts such as *P. pastoris* are chosen for production of therapeutic proteins, N- and O-linked glycosylation are of tremendous relevance. Although the assembly of the core glycans, that is, $(Man)_8$-$(GlcNAc)_2$, in the ER is highly conserved in mammals and yeasts, mammals provide a much higher diversity in the ultimate glycan structure assembled in the Golgi cisternae. Yeasts, in contrast, produce high mannose glycan structures, which may lead to decreased serum half-life and may trigger allergic reactions in the human body (Ballou 1990). While in *P. pastoris* the hyper-mannosylation is not as prominent as in *S. cerevisiae*, it is still a problem that needs to be tackled, and is therefore a target for intensive strain engineering. A very detailed summary of the glycosylation machinery and the targets for glyco-engineering in different yeast species, including *P. pastoris*, has been given recently (De Pourcq et al. 2010). To sum up briefly, engineering strategies included the introduction of a *Trichoderma reesei* α-1,2-mannosidase (Callewaert et al. 2001), the knockout of the highly conserved yeast Golgi protein α-1,6-mannosyltransferase encoded by *OCH1*, which is responsible for hyperglycosylation (Choi et al. 2003; Vervecken et al. 2004), as well as co-overexpression of several glycosyltransferases and glycosidases carrying proper targeting signals (Hamilton et al. 2003). Terminally sialylated glycoproteins

produced for the first in *P. pastoris* were obtained by introducing a complex sialic acid pathway (Hamilton et al. 2006). Key to success was the correct localization of the heterologous glycosyltransferases and glycosidases in the ER and Golgi networks. Combinatorial genetic libraries and high throughput screening methods were successfully applied to find the best targeting signal/enzyme combinations for N-linked glycoengineering (Nett et al. 2011). Furthermore, a useful guide to glyco-engineering in *P. pastoris* by using the GlycoSwitch® technology was described by Jacobs et al. (2009). These strategies, altogether, enable the production of valuable biopharmaceuticals with a more homogeneous, 'humanized' N-glycosylation pattern.

However, as yeasts also carry out O-glycosylation that differs structurally from the mammalian type (Strahl-Bolsinger et al. 1999), O-glycosylation has also been an interesting target for engineering. In *P. pastoris*, O-linked glycosylation is initiated with a mannose monosaccharide, which is further elongated by α-1,2-mannose residues and finally capped with β- or phospho-mannose residues. Until lately, the engineering strategies were limited to the use of an inhibitor of the major ER located protein-O-mannosyltransferases (PMTs) as the deletion of these genes did not yield robust and viable strains. The characterization of the *P. pastoris* PMT gene family was an important step forward in O-glycosylation engineering (Nett et al. 2013). In this study, the knockout of PMTs as well as the use of PMT inhibitors led to a reduced number of O-mannosylation events and, furthermore, to reduced chain lengths of the O-glycans. A follow-up study described the production of a TNFR2:Fc[1] fusion protein carrying sialylated O-linked glycans in *P. pastoris* (Hamilton et al. 2013). Therein, an α-1,2-mannosidase as well as a protein-O-linked-mannose β-1,2-N-acetylglucosaminyl-transferase 1 (PomGnT1) were co-expressed in a *P. pastoris* strain, that was already engineered in its N-glycosylation pathway. Hence, the mannose residues were first trimmed to single O-linked mannose residues, which were then capped with N-acetylglucosamine. This structure was extended with sialic acid residues to achieve human-like O-glycan residues similar to the α-dystroglycan-type. However, there is still room for improvement, for example by engineering *P. pastoris* towards human mucin-type O-glycosylation.

Expression strategies and industrial applications

Screening for high level expression

Subsequent to the choice of suitable expression vectors and proper host strains, and transformation of the expression cassettes, it is important to select for transformants which show high expression levels of the desired protein. Single copy transformants can be easily generated by targeting the linear expression cassettes to the *AOX1* locus resulting in gene replacement events. Ectopic integrations may simultaneously occur, however. Transformants resulting from gene replacement at the *AOX1* locus have methanol utilization slow phenotype (Mut^S) and can be easily identified by replica-plating on minimal methanol plates. The most commonly applied strategy to screen for high-yielding *P. pastoris* transformants focusses on screening for clones having multicopy integrations of the expression cassette. A recent detailed review describes the methods applied to obtain strains containing multiple expression cassettes and provides a summary of published data showing correlations between copy number and expression levels of intracellular as well as secreted proteins. It also highlights the problem of genetic instability of the integration cassettes that might be encountered when cultivating multicopy strains. Due to the highly recombinogenic nature of *P. pastoris*, expression cassettes might be excised through loop-out recombination. This effect seems to be more pronounced the more copies are integrated (Aw and Polizzi 2013).

Regarding the correlation between copy number and expression level, a number of recent studies have shown a direct correlation especially for intracellular expression (Marx et al. 2009; Vassileva et al. 2001). The direct correlation of expression level and gene copy number is, however, not necessarily valid when the protein is directed to the secretory pathway. The most commonly employed method of generating multicopy expression strains in *P. pastoris* is based on plating the transformation mixture directly on selection plates containing increasing concentrations of antibiotics (e.g., 100 to 2,000 μg/ml of Zeocin™). The majority of transformants will have a single copy of the expression vector integrated into the genome, and numerous clones will have to be screened to find high-copy transformants (Lin-Cereghino and Lin-Cereghino 2007). Therefore, several high-throughput methods have been established to screen a large number of clones based on small-scale cultivation in deep well plates (Mellitzer et al. 2012; Weinhandl et al. 2012; Weis et al. 2004). The selected clones, however, might not perform as well in fermenter cultivations due to different cultivation conditions. A further pronounced problem of resistance marker based screening is a high prevalence of false-positive colonies. This so-called high transformation background is supposedly caused by cell stress and cell rupture. Depending on the mechanism of antibiotic resistance conferred by the resistance marker, un-transformed cells may survive in the vicinity of ruptured transformants. This problem was addressed by constructing expression vectors based on marker gene expression driven by the weak *ARG4* promoter (*Pichia* Pool, Fig. 2). This ensures basal levels of expression, thereby allowing handlers to select single copy to

[1] Ectodomain of tumor necrosis factor 2 with crystallizable fragment of IgG1 (Fc)

multicopy strains by plating the transformants directly on low concentrations of Zeocin™ (i.e., 25 μg/ml for single copy and up to 400 μg/ml for multi-copy transformants). Thus, transformants having 1 to 20 (±5) copies can be selected. To reduce the chances of having single copy transformants, regeneration time should be kept short and transformants should be plated directly on increased concentrations of antibiotic. By employing this method, only few transformants survive on high concentrations of antibiotic, but will most likely contain multiple copies, which can be determined by quantitative (qPCR) or Southern blot analysis (M. Ahmad, unpublished results). Performance can then be tested directly under production conditions in bioreactor cultivations instead of small-scale cultivations in deep well plates or shake flasks.

Membrane protein expression

P. pastoris has been shown to produce 15+ g of soluble recombinant protein per litre of culture intracellularly (Hasslacher et al. 1997) or in secretory mode (Werten et al. 1999). Key to such high titres is the ability of *P. pastoris* to grow to very high cell densities reaching up to 150 g cell dry weight per litre of fermentation broth in fed-batch bioreactor cultivations (Jahic et al. 2006). At very high cell densities, even proteins that are present in limited entities per single cell can be produced with reasonable volumetric yields in *P. pastoris*. Typical examples of non-abundant proteins with high scientific and commercial relevance are integral membrane proteins. Being the targets of >50 % of drugs applied on humans (Arinaminpathy et al. 2009), only very few membrane proteins have been characterized on the molecular level regarding structure–function relationships. The simple reason is that it is difficult to obtain sufficient purified membrane protein for structural and biochemical studies, unless affinity-tagged membrane proteins are obtained at reasonable yield. Actually, *P. pastoris* has been applied routinely to produce affinity-tagged membrane proteins for protein purification and subsequent biochemical studies (Cohen et al. 2005; Haviv et al. 2007; Lifshitz et al. 2007). Furthermore, *P. pastoris* has been the expression host of choice for elucidating the crystal structures of membrane proteins from diverse origins, even from higher eukaryotes (Brohawn et al. 2012; Hino et al. 2012; Ho et al. 2009).

Evolutionary proximity of a heterologous expression host and the origin of an expressed membrane protein are beneficial for successful recombinant expression (Grisshammer and Tateu 2009). In addition to the intramolecular forces and bonds, ions, cofactors and interacting proteins that stabilize soluble proteins, membrane proteins are usually interacting with and are partially also stabilized by the lipids of the surrounding bilayers (Adamian et al. 2011). As *P. pastoris* and other yeast expression hosts do significantly differ in their membrane compositions from bacterial, plant or animal cells (Wriessnegger et al. 2007, 2009; Zinser and Daum 1995), heterologous membrane proteins may face stability issues upon expression in distantly related hosts. Thus, multiple approaches have been undertaken to improve *P. pastoris* host strains and expression conditions for membrane protein production. Applying similar tools as for the optimisation of soluble protein expression — that is, manipulation of expression conditions, addition of chemical chaperones, co-expression of chaperones or of proteins activating UPR, use of protease deficient strains, etc. — has been showing some, however often target-specific success in membrane protein expression. A novel approach is the engineering of *P. pastoris* cellular membranes for improved accommodation of heterologous membrane proteins. In the first reported example, a cholesterol-producing *P. pastoris* strain was shown to stably express an enhanced level of ligand-binding human Na,K-ATPase moieties on the cell surface (Hirz et al. 2013).

Products on — or on the way to — the market

The *P.* pastoris expression system has gained importance for industrial application as highlighted by the number of patents published on heterologous expression in and cell engineering of *P. pastoris* (Bollok et al. 2009). Products obtained by heterologous expression in *P. pastoris* have already found their way to the market, as FDA approved biopharmaceuticals or industrial enzymes have shown. The www.pichia.com web page provides a list of proteins produced in *P. pastoris* with the commercial expression system licensed by Research Corporation Technologies (RCT) and their applications: Phytase (Phytex, Sheridan, IN, USA) is applied as animal feed additive to cleave plant derived phytate, thereby providing a source of phosphate. Trypsin (Roche Applied Science, Germany) is used, for example, as protease in proteomics research to obtain peptide patterns for MS analysis. Further examples listed are nitrate reductase (The Nitrate Elimination Co., Lake Linden, MI, USA), used for water testing and treatment, phospholipase C (Verenium, San Diego, CA, USA/DSM, The Netherlands), used for degumming of vegetable oils, and Collagen (Fibrogen, San Francisco, CA, USA), used in medical research and as dermal filler. Thermo Scientific (Waltham, MA, USA) sells recombinant *Tritirachium album* Proteinase K produced in *P. pastoris*. Concerning biopharmaceuticals, a famous example is Kalbitor® (ecallantide), produced in *P. pastoris* by Dyax (Cambridge, MA, USA). Kalbitor® is a plasma kallikrein inhibitor indicated against hereditary angioedema. This product was the first biopharmaceutical to be approved by the FDA for market release in 2009 (Walsh 2010). As can be found on the web page of RCT (www.rctech.com), *Pichia*-manufactured Jetrea®, a drug used for treatment of symptomatic vitreomacular adhesion, was recently approved by the FDA and the European Commission. Other *Pichia*-

derived products provided by the Indian company Biocon are recombinant human insulin and analogues thereof (Insulin, Glargine). Products under development, such as Elastase inhibitor against Cystic fibrosis or Nanobody® ALX antibody fragments developed by Ablynx (Belgium), are also listed by Gerngross (2004) and on www.pichia.com. In 2008, Novozymes (Denmark), which found a highly active antimicrobial agent, the plectasin peptide derivative NZ2114 (Andes et al. 2009; Mygind et al. 2005), granted Sanofi-Aventis (France) an exclusive licence for the production and commercialisation of this compound in *P. pastoris.* This might be the first antimicrobial peptide approved for the market in the future.

Although not yet approved for medical use, many products can be found on the market for research purposes. GenScript (Piscataway, NJ, USA) provides recombinant cytokines and growth factors, such as human HSA-IFN-Alpha 2b, human Stem Cell Factor SCF, murine TNF-α and ovine IFN-τ, to name just a few examples. Recombinant human angiostatin can be found for instance in the reagents offered by Sigma-Aldrich (St. Louis, MO, USA).

Future perspectives — outlook

Successful expression of many industrial enzymes as well as pharmaceutically relevant proteins has rendered the methylotrophic yeast *P. pastoris* one of the most suitable and powerful protein production host systems. It is also an emerging host for the expression of membrane proteins (Hirz et al. 2013) and of small bioactive and antimicrobial peptides, which could be a forthcoming alternative to chemical synthesis (Zhang et al. 2014). Although many basic elements of this expression system are now well developed and one can make use of a broad variety of vectors and host strains, there is still space for further optimization of protein expression and secretion, which, in many cases, will be highly dependent on the desired product. One general interest is to find effective alternatives for induction to replace methanol for industrial scale fermentations (Delic et al. 2013; Prielhofer et al. 2013; Stadlmayr et al. 2010).

Improving protein secretion performance is one of the first and foremost goals for engineering *P. pastoris*. There is still potential to increase yields, for example, by employing different secretion signals (Vadhana et al. 2013) or mutating *S. cerevisiae* α-MF (Lin-Cereghino et al. 2013). In contrast to the well-studied secretory pathway of *S. cerevisiae*, *P. pastoris* still is a black box regarding factors influencing secretion efficiency. Current studies try to identify these factors by mutagenesis approaches and screening for enhanced secretion of reporter proteins (Larsen et al. 2013; C. Winkler and H. Pichler, unpublished results). The well-developed tools for strain engineering, including marker-free integration and deletion of desired genes, will provide a powerful set of engineered designer host strains in the near future. These will provide optimized cell factories by fine-tuned co-expression of important homologous or heterologous protein functions needed for efficient and accurate functional expression, secretion and post-translational modification of proteins. Moreover, knockout or knockdown of undesired functions such as proteolytic decay will increase product quality and process performance. Considering the scope of this review on heterologous protein expression, it was not feasible to address all possible applications for *P. pastoris* as production organism, such as metabolic engineering for production of small molecules and metabolites, or for whole-cell biocatalysis. However, developments in these fields may also be relevant for constructing improved host strains dedicated for protein production. There are several recent reviews and research articles describing advances in these fields in detail (Abad et al. 2010; Araya-Garay et al. 2012; Wriessnegger and Pichler 2013).

References

Abad S, Nahalka J, Bergler G, Arnold SA, Speight R, Fotheringham I, Nidetzky B, Glieder A (2010) Stepwise engineering of a *Pichia pastoris* D-amino acid oxidase whole cell catalyst. Microb Cell Factories 9:24. doi:10.1186/1475-2859-9-24

Adamian L, Naveed H, Liang J (2011) Lipid-binding surfaces of membrane proteins: evidence from evolutionary and structural analysis. Biochim Biophys Acta 1808:1092–1102. doi:10.1016/j.bbamem.2010.12.008

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

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

Andes D, Craig W, Nielsen LA, Kristensen HH, (2009) *In vivo* pharmacodynamic characterization of a novel plectasin antibiotic, NZ2114, in a murine infection model. Antimicrob Agents Chemother 53:3003–3009. doi:10.1128/AAC.01584-08

Araya-Garay JM, Feijoo-Siota L, Rosa-dos-Santos F, Veiga-Crespo P, Villa TG (2012) Construction of new *Pichia pastoris* X-33 strains for production of lycopene and β-carotene. Appl Microbiol Biotechnol 93:2483–2492. doi:10.1007/s00253-011-3764-7

Arinaminpathy Y, Khurana E, Engelman DM, Gerstein MB (2009) Computational analysis of membrane proteins: the largest class of drug targets. Drug Discov Today 14:1130–1135. doi:10.1016/j.drudis.2009.08.006

Austin RJ, Kuestner RE, Chang DK, Madden KR, Martin DB (2011) SILAC compatible strain of *Pichia pastoris* for expression of

isotopically labeled protein standards and quantitative proteomics. J Proteome Res 10:5251–5259. doi:10.1021/pr200551e
Aw R, Polizzi KM (2013) Can too many copies spoil the broth? Microb Cell Factories 12:128. doi:10.1186/1475-2859-12-128
Ballou CE (1990) Isolation, characterization, and properties of *Saccharomyces cerevisiae mnn* mutants with nonconditional protein glycosylation defects. Methods Enzymol 185:440–470
Boehm T, Pirie-Shepherd S, Trinh LB, Shiloach J, Folkman J (1999) Disruption of the *KEX1* gene in *Pichia pastoris* allows expression of full-length murine and human endostatin. Yeast 15:563–572. doi:10.1002/(SICI)1097-0061(199905)15:7<563::AID-YEA398>3.0.CO;2-R
Bollok M, Resina D, Valero F, Ferrer P (2009) Recent patents on the *Pichia pastoris* expression system: expanding the toolbox for recombinant protein production. Recent Pat Biotechnol 3:192–201. doi:10.2174/187220809789389126
Brake AJ, Merryweather JP, Coit DG, Heberlein UA, Masiarz FR, Mullenbach GT, Urdea MS, Valenzuela P, Barr PJ (1984) Alpha-factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 81:4642–4646
Brohawn SG, del Mármol J, MacKinnon R (2012) Crystal structure of the human K2P TRAAK, a lipid- and mechano-sensitive K^+ ion channel. Science 335:436–441. doi:10.1126/science.1213808
Budisa N, Wenger W, Wiltschi B (2010) Residue-specific global fluorination of *Candida antarctica* lipase B in *Pichia pastoris*. Mol BioSyst 6:1630–1639. doi:10.1039/c002256j
Callewaert N, Laroy W, Cadirgi H, Geysens S, Saelens X, Min Jou W, Contreras R (2001) Use of HDEL-tagged *Trichoderma reesei* mannosyl oligosaccharide 1,2-alpha-D-mannosidase for N-glycan engineering in *Pichia pastoris*. FEBS Lett 503:173–178
Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol Rev 24:45–66. doi:10.1111/j.1574-6976.2000.tb00532.x
Choi B-K, Bobrowicz P, Davidson RC, Hamilton SR, Kung DH, Li H, Miele RG, Nett JH, Wildt S, Gerngross TU (2003) Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast *Pichia pastoris*. Proc Natl Acad Sci U S A 100:5022–5027. doi:10.1073/pnas.0931263100
Ciofalo V, Barton N, Kreps J, Coats I, Shanahan D (2006) Safety evaluation of a lipase enzyme preparation, expressed in *Pichia pastoris*, intended for use in the degumming of edible vegetable oil. Regul Toxicol Pharmacol 45:1–8. doi:10.1016/j.yrtph.2006.02.001
Cohen E, Goldshleger R, Shainskaya A, Tal DM, Ebel C, le Maire M, Karlish SJD (2005) Purification of Na+, K+-ATPase expressed in *Pichia pastoris* reveals an essential role of phospholipid-protein interactions. J Biol Chem 280:16610–16618. doi:10.1074/jbc.M414290200
Cregg J, Tolstorukov I (2012) *P. pastoris* ADH promoter and use thereof to direct expression of proteins. United States patent US 8,222,386
Cregg JM, Barringer KJ, Hessler AY, Madden KR (1985) *Pichia pastoris* as a host system for transformations. Mol Cell Biol 5:3376–3385. doi:10.1128/MCB.5.12.3376
Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA (1989) Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. Mol Cell Biol 9:1316–1323
Cregg JM, Shen S, Johnson M, Waterham HR (1998) Classical genetic manipulation. Methods Mol Biol 103:17–26. doi:10.1385/0-89603-421-6:17
Daly R, Hearn MTW (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. J Mol Recognit 18:119–138. doi:10.1002/jmr.687
Damasceno LM, Huang C-J, Batt CA (2012) Protein secretion in *Pichia pastoris* and advances in protein production. Appl Microbiol Biotechnol 93:31–39. doi:10.1007/s00253-011-3654-z
de Almeida JRM, de Moraes LMP, Torres FAG (2005) Molecular characterization of the 3-phosphoglycerate kinase gene (*PGK1*) from the methylotrophic yeast *Pichia pastoris*. Yeast 22:725–737. doi:10.1002/yea.1243
De Pourcq K, De Schutter K, Callewaert N (2010) Engineering of glycosylation in yeast and other fungi: current state and perspectives. Appl Microbiol Biotechnol 87:1617–1631. doi:10.1007/s00253-010-2721-1
De Schutter K, Lin Y-C, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouzé P, Van de Peer Y, Callewaert N (2009) Genome sequence of the recombinant protein production host *Pichia pastoris*. Nat Biotechnol 27:561–566. doi:10.1038/nbt.1544
Delic M, Mattanovich D, Gasser B (2013) Repressible promoters — a novel tool to generate conditional mutants in *Pichia pastoris*. Microb Cell Factories 12:6. doi:10.1186/1475-2859-12-6
Du M, Battles MB, Nett JH (2012) A color-based stable multi-copy integrant selection system for *Pichia pastoris* using the attenuated *ADE1* and *ADE2* genes as auxotrophic markers. Bioeng Bugs 3:32–37. doi:10.4161/bbug.3.1.17936
Eiden-Plach A, Zagorc T, Heintel T, Carius Y, Breinig F, Schmitt MJ (2004) Viral preprotoxin signal sequence allows efficient secretion of green fluorescent protein by *Candida glabrata*, *Pichia pastoris*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*. Appl Environ Microbiol 70:961–966. doi:10.1128/AEM.70.2.961-966.2004
Ellis SB, Brust PF, Koutz PJ, Waters AF, Harpold MM, Gingeras TR (1985) Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*. Mol Cell Biol 5:1111–1121
Gasser B, Prielhofer R, Marx H, Maurer M, Nocon J, Steiger M, Puxbaum V, Sauer M, Mattanovich D (2013) *Pichia pastoris*: protein production host and model organism for biomedical research. Future Microbiol 8:191–208. doi:10.2217/fmb.12.133
Gerngross TU (2004) Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. Nat Biotechnol 22:1409–1414. doi:10.1038/nbt1028
Gleeson MA, White CE, Meininger DP, Komives EA (1998) Generation of protease-deficient strains and their use in heterologous protein expression. Methods Mol Biol 103:81–94. doi:10.1385/0-89603-421-6:81
Grisshammer R, Tateu CG (2009) Overexpression of integral membrane proteins for structural studies. Q Rev Biophys 28:315. doi:10.1017/S0033583500003504
Guerfal M, Ryckaert S, Jacobs PP, Ameloot P, Van Craenenbroeck K, Derycke R, Callewaert N (2010) The *HAC1* gene from *Pichia pastoris*: characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins. Microb Cell Factories 9:49. doi:10.1186/1475-2859-9-49
Gustavsson M, Lehtio J, Denman S, Teeri TT, Hult K, Martinelle M (2001) Stable linker peptides for a cellulose-binding domain–lipase fusion protein expressed in *Pichia pastoris*. Protein Eng Des Sel 14:711–715. doi:10.1093/protein/14.9.711
Hamilton SR, Bobrowicz P, Bobrowicz B, Davidson RC, Li H, Mitchell T, Nett JH, Rausch S, Stadheim TA, Wischnewski H, Wildt S, Gerngross TU (2003) Production of complex human glycoproteins in yeast. Science 301:1244–1246. doi:10.1126/science.1088166
Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi B, Hopkins D, Wischnewski H, Roser J, Mitchell T, Strawbridge RR, Hoopes J, Wildt S, Gerngross TU (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins. Science 313:1441–1443. doi:10.1126/science.1130256
Hamilton SR, Cook WJ, Gomathinayagam S, Burnina I, Bukowski J, Hopkins D, Schwartz S, Du M, Sharkey NJ, Bobrowicz P, Wildt S, Li H, Stadheim TA, Nett JH (2013) Production of sialylated O-linked glycans in *Pichia pastoris*. Glycobiology 23:1192–1203. doi:10.1093/glycob/cwt056

Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, Lin-Cereghino J, Kovar K, Cregg JM, Glieder A (2008) Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. Nucleic Acids Res 36:e76. doi:10.1093/nar/gkn369

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

Haviv H, Cohen E, Lifshitz Y, Tal DM, Goldshleger R, Karlish SJD (2007) Stabilization of Na(+), K(+)-ATPase purified from Pichia pastoris membranes by specific interactions with lipids. Biochemistry 46:12855–12867. doi:10.1021/bi701248y

Hino T, Arakawa T, Iwanari H, Yurugi-Kobayashi T, Ikeda-Suno C, Nakada-Nakura Y, Kusano-Arai O, Weyand S, Shimamura T, Nomura N, Cameron AD, Kobayashi T, Hamakubo T, Iwata S, Murata T (2012) G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. Nature 482:237–240. doi:10.1038/nature10750

Hirz M, Richter G, Leitner E, Wriessnegger T, Pichler H (2013) A novel cholesterol-producing *Pichia pastoris* strain is an ideal host for functional expression of human Na, K-ATPase α3β1 isoform. Appl Microbiol Biotechnol 97:9465–9478. doi:10.1007/s00253-013-5156-7

Ho JD, Yeh R, Sandstrom A, Chorny I, Harries WEC, Robbins RA, Miercke LJW, Stroud RM (2009) Crystal structure of human aquaporin 4 at 1.8 A and its mechanism of conductance. Proc Natl Acad Sci U S A 106:7437–7442. doi:10.1073/pnas.0902725106

Idiris A, Tohda H, Kumagai H, Takegawa K (2010) Engineering of protein secretion in yeast: strategies and impact on protein production. Appl Microbiol Biotechnol 86:403–417. doi:10.1007/s00253-010-2447-0

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

Jacobs PP, Geysens S, Vervecken W, Contreras R, Callewaert N (2009) Engineering complex-type N-glycosylation in *Pichia pastoris* using GlycoSwitch technology. Nat Protoc 4:58–70. doi:10.1038/nprot.2008.213

Jahic M, Veide A, Charoenrat T, Teeri T, Enfors S-O (2006) Process technology for production and recovery of heterologous proteins with *Pichia pastoris*. Biotechnol Prog 22:1465–1473. doi:10.1021/bp060171t

Jin F, Xu X, Zhang W, Gu D (2006) Expression and characterization of a housefly cecropin gene in the methylotrophic yeast, *Pichia pastoris*. Protein Expr Purif 49:39–46. doi:10.1016/j.pep.2006.03.008

Julius D, Brake A, Blair L, Kunisawa R, Thorner J (1984) Isolation of the putative structural gene for the lysine–arginine-cleaving endopeptidase required for processing of yeast prepro-alpha-factor. Cell 37:1075–1089

Kang HA, Choi ES, Hong WK, Kim JY, Ko SM, Sohn JH, Rhee SK (2000) Proteolytic stability of recombinant human serum albumin secreted in the yeast *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 53:575–582

Kato S, Ishibashi M, Tatsuda D, Tokunaga H, Tokunaga M (2001) Efficient expression, purification and characterization of mouse salivary alpha-amylase secreted from methylotrophic yeast, *Pichia pastoris*. Yeast 18:643–655. doi:10.1002/yea.714

Khasa YP, Conrad S, Sengul M, Plautz S, Meagher MM, Inan M (2011) Isolation of *Pichia pastoris PIR* genes and their utilization for cell surface display and recombinant protein secretion. Yeast 28:213–226. doi:10.1002/yea.1832

Kjeldsen T, Hach M, Balschmidt P, Havelund S, Pettersson AF, Markussen J (1998) Prepro-leaders lacking N-linked glycosylation for secretory expression in the yeast *Saccharomyces cerevisiae*. Protein Expr Purif 14:309–316. doi:10.1006/prep.1998.0977

Kottmeier K, Ostermann K, Bley T, Rödel G (2011) Hydrophobin signal sequence mediates efficient secretion of recombinant proteins in *Pichia pastoris*. Appl Microbiol Biotechnol 91:133–141. doi:10.1007/s00253-011-3246-y

Kranthi BV, Balasubramanian N, Rangarajan PN (2006) Isolation of a single-stranded DNA-binding protein from the methylotrophic yeast, *Pichia pastoris* and its identification as zeta crystallin. Nucleic Acids Res 34:4060–4068. doi:10.1093/nar/gkl577

Kranthi BV, Kumar R, Kumar NV, Rao DN, Rangarajan PN (2009) Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in *Pichia pastoris*. Biochim Biophys Acta 1789:460–468. doi:10.1016/j.bbagrm.2009.05.004

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

Kumar NV, Rangarajan PN (2012) The zinc finger proteins Mxr1p and repressor of phosphoenolpyruvate carboxykinase (ROP) have the same DNA binding specificity but regulate methanol metabolism antagonistically in *Pichia pastoris*. J Biol Chem 287:34465–34473. doi:10.1074/jbc.M112.365304

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

Larsen S, Weaver J, de Sa CK, Bulahan R, Nguyen J, Grove H, Huang A, Low L, Tran N, Gomez S, Yau J, Ilustrisimo T, Kawilarang J, Lau J, Tranphung M, Chen I, Tran C, Fox M, Lin-Cereghino J, Lin-Cereghino GP (2013) Mutant strains of *Pichia pastoris* with enhanced secretion of recombinant proteins. Biotechnol Lett. doi:10.1007/s10529-013-1290-7

Lee CC, Williams TG, Wong DWS, Robertson GH (2005) An episomal expression vector for screening mutant gene libraries in *Pichia pastoris*. Plasmid 54:80–85. doi:10.1016/j.plasmid.2004.12.001

Li Z, Leung W, Yon A, Nguyen J, Perez VC, Vu J, Giang W, Luong LT, Phan T, Salazar KA, Gomez SR, Au C, Xiang F, Thomas DW, Franz AH, Lin-Cereghino J, Lin-Cereghino GP (2010) Secretion and proteolysis of heterologous proteins fused to the *Escherichia coli* maltose binding protein in *Pichia pastoris*. Protein Expr Purif 72:113–124. doi:10.1016/j.pep.2010.03.004

Liang S, Li C, Ye Y, Lin Y (2013a) Endogenous signal peptides efficiently mediate the secretion of recombinant proteins in *Pichia pastoris*. Biotechnol Lett 35:97–105. doi:10.1007/s10529-012-1055-8

Liang S, Zou C, Lin Y, Zhang X, Ye Y (2013b) Identification and characterization of P GCW14 : a novel, strong constitutive promoter of *Pichia pastoris*. Biotechnol Lett 35:1865–1871. doi:10.1007/s10529-013-1265-8

Lifshitz Y, Petrovich E, Haviv H, Goldshleger R, Tal DM, Garty H, Karlish SJD (2007) Purification of the human alpha2 Isoform of Na, K-ATPase expressed in *Pichia pastoris*. Stabilization by lipids and FXYD1. Biochemistry 46:14937–14950. doi:10.1021/bi701812c

Lin-Cereghino J, Lin-Cereghino GP (2007) Vectors and strains for expression. Methods Mol Biol 389:11–26. doi:10.1007/978-1-59745-456-8_2

Lin-Cereghino GP, Lin-Cereghino J, Jay Sunga A, Johnson MA, Lim M, Gleeson MAG, Cregg JM (2001) New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of *Pichia pastoris*. Gene 263:159–169. doi:10.1016/S0378-1119(00)00576-X

Lin-Cereghino GP, Godfrey L, de la Cruz BJ, De JS, Khuongsathiene S, Tolstorukov I, Yan M, Lin-Cereghino J, Veenhuis M, Subramani S, Cregg JM (2006) Mxr1p, a key regulator of the methanol utilization

pathway and peroxisomal genes in *Pichia pastoris*. Mol Cell Biol 26:883–897. doi:10.1128/MCB.26.3.883-897.2006

Lin-Cereghino GP, Stark CM, Kim D, Chang JWJ, Shaheen N, Poerwanto H, Agari K, Moua P, Low LK, Tran N, Huang AD, Nattestad M, Oshiro KT, Chavan A, Tsai JW, Lin-Cereghino J (2013) The effect of α-mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*. Gene 519:311–317. doi:10.1016/j.gene.2013.01.062

Lünsdorf H, Gurramkonda C, Adnan A, Khanna N, Rinas U (2011) Virus-like particle production with yeast: ultrastructural and immunocytochemical insights into *Pichia pastoris* producing high levels of the hepatitis B surface antigen. Microb Cell Factories 10:48. doi:10.1186/1475-2859-10-48

Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the *Pichia pastoris* expression system. Yeast 22:249–270. doi:10.1002/yea.1208

Marx H, Mecklenbräuker A, Gasser B, Sauer M, Mattanovich D (2009) Directed gene copy number amplification in *Pichia pastoris* by vector integration into the ribosomal DNA locus. FEMS Yeast Res 9:1260–1270. doi:10.1111/j.1567-1364.2009.00561.x

Mattanovich D, Callewaert N, Rouzé P, Lin Y-C, Graf A, Redl A, Tiels P, Gasser B, De Schutter K (2009a) Open access to sequence: browsing the *Pichia pastoris* genome. Microb Cell Factories 8:53. doi:10.1186/1475-2859-8-53

Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, Kleinheinz M, Sauer M, Altmann F, Gasser B (2009b) Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*. Microb Cell Factories 8:29. doi:10.1186/1475-2859-8-29

Mellitzer A, Glieder A, Weis R, Reisinger C, Flicker K (2012) Sensitive high-throughput screening for the detection of reducing sugars. Biotechnol J 7:155–162. doi:10.1002/biot.201100001

Menendez J, Valdes I, Cabrera N (2003) The *ICL1* gene of *Pichia pastoris*, transcriptional regulation and use of its promoter. Yeast 20:1097–1108. doi:10.1002/yea.1028

Moir DT, Dumais DR (1987) Glycosylation and secretion of human alpha-1-antitrypsin by yeast. Gene 56:209–217

Mygind PH, Fischer RL, Schnorr KM, Hansen MT, Sönksen CP, Ludvigsen S, Raventós D, Buskov S, Christensen B, De Maria L, Taboureau O, Yaver D, Elvig-Jørgensen SG, Sørensen MV, Christensen BE, Kjaerulff S, Frimodt-Moller N, Lehrer RI, Zasloff M, Kristensen H-H (2005) Plectasin is a peptide antibiotic with therapeutic potential from a saprophytic fungus. Nature 437:975–980. doi:10.1038/nature04051

Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A (2012) Deletion of the *Pichia pastoris KU70* homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS ONE 7:e39720. doi:10.1371/journal.pone.0039720

Nett J (2010) Yeast strains for protein production. United States patent US 2010/0279348

Nett JH, Gerngross TU (2003) Cloning and disruption of the *PpURA5* gene and construction of a set of integration vectors for the stable genetic modification of *Pichia pastoris*. Yeast 20:1279–1290. doi:10.1002/yea.1049

Nett JH, Hodel N, Rausch S, Wildt S (2005) Cloning and disruption of the *Pichia pastoris ARG1*, *ARG2*, *ARG3*, *HIS1*, *HIS2*, *HIS5*, *HIS6* genes and their use as auxotrophic markers. Yeast 22:295–304. doi:10.1002/yea.1202

Nett JH, Stadheim TA, Li H, Bobrowicz P, Hamilton SR, Davidson RC, Choi B-K, Mitchell T, Bobrowicz B, Rittenhour A, Wildt S, Gerngross TU (2011) A combinatorial genetic library approach to target heterologous glycosylation enzymes to the endoplasmic reticulum or the Golgi apparatus of *Pichia pastoris*. Yeast 28:237–252. doi:10.1002/yea.1835

Nett JH, Cook WJ, Chen M-T, Davidson RC, Bobrowicz P, Kett W, Brevnova E, Potgieter TI, Mellon MT, Prinz B, Choi B-K, Zha D, Burnina I, Bukowski JT, Du M, Wildt S, Hamilton SR (2013) Characterization of the *Pichia pastoris* protein-O-mannosyltransferase gene family. PLoS ONE 8:e68325. doi:10.1371/journal.pone.0068325

Ni Z, Zhou X, Sun X, Wang Y, Zhang Y (2008) Decrease of hirudin degradation by deleting the *KEX1* gene in recombinant *Pichia pastoris*. Yeast 25:1–8. doi:10.1002/yea.1542

Ohi H, Miura M, Hiramatsu R, Ohmura T (1994) The positive and negative *cis*-acting elements for methanol regulation in the *Pichia pastoris AOX2* gene. Mol Gen Genet 243:489–499

Oka C, Tanaka M, Muraki M, Harata K, Suzuki K, Jigami Y (1999) Human lysozyme secretion increased by alpha-factor pro-sequence in *Pichia pastoris*. Biosci Biotechnol Biochem 63:1977–1983. doi:10.1271/bbb.63.1977

Paifer E, Margolles E, Cremata J, Montesino R, Herrera L, Delgado JM (1994) Efficient expression and secretion of recombinant alpha amylase in *Pichia pastoris* using two different signal sequences. Yeast 10:1415–1419. doi:10.1002/yea.320101104

Parua PK, Ryan PM, Trang K, Young ET (2012) *Pichia pastoris* 14-3-3 regulates transcriptional activity of the methanol inducible transcription factor Mxr1 by direct interaction. Mol Microbiol 85:282–298. doi:10.1111/j.1365-2958.2012.08112.x

Payne WE, Gannon PM, Kaiser CA (1995) An inducible acid phosphatase from the yeast *Pichia pastoris*: characterization of the gene and its product. Gene 163:19–26. doi:10.1016/0378-1119(95)00379-K

Polupanov AS, Nazarko VY, Sibirny AA (2012) *Gss1* protein of the methylotrophic yeast *Pichia pastoris* is involved in glucose sensing, pexophagy and catabolite repression. Int J Biochem Cell Biol 44:1906–1918. doi:10.1016/j.biocel.2012.07.017

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

Qin X, Qian J, Yao G, Zhuang Y, Zhang S, Chu J (2011) GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*. Appl Environ Microbiol 77:3600–3608. doi:10.1128/AEM.02843-10

Raemaekers RJM, de Muro L, Gatehouse JA, Fordham-Skelton AP (1999) Functional phytohemagglutinin (PHA) and *Galanthus nivalis* agglutinin (GNA) expressed in *Pichia pastoris*. Correct N-terminal processing and secretion of heterologous proteins expressed using the PHA-E signal peptide. Eur J Biochem 265:394–403. doi:10.1046/j.1432-1327.1999.00749.x

Rakestraw JA, Sazinsky SL, Piatesi A, Antipov E, Wittrup KD (2009) Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in *Saccharomyces cerevisiae*. Biotechnol Bioeng 103:1192–1201. doi:10.1002/bit.22338

Sha C, Yu X-W, Li F, Xu Y (2013) Impact of gene dosage on the production of lipase from *Rhizopus chinensis* CCTCC M201021 in *Pichia pastoris*. Appl Biochem Biotechnol 169:1160–1172. doi:10.1007/s12010-012-0050-9

Shen S, Sulter G, Jeffries TW, Cregg JM (1998) A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast *Pichia pastoris*. Gene 216:93–102

Shi X-L, Feng M-Q, Shi J, Shi Z-H, Zhong J, Zhou P (2007) High-level expression and purification of recombinant human catalase in *Pichia pastoris*. Protein Expr Purif 54:24–29. doi:10.1016/j.pep.2007.02.008

Sigoillot M, Brockhoff A, Lescop E, Poirier N, Meyerhof W, Briand L (2012) Optimization of the production of gurmarin, a sweet-taste-suppressing protein, secreted by the methylotrophic yeast *Pichia pastoris*. Appl Microbiol Biotechnol 96:1253–1263. doi:10.1007/s00253-012-3897-3

Sinha J, Plantz BA, Inan M, Meagher MM (2005) Causes of proteolytic degradation of secreted recombinant proteins produced in methylotrophic yeast *Pichia pastoris*: case study with recombinant

ovine interferon-tau. Biotechnol Bioeng 89:102–112. doi:10.1002/bit.20318
Skipper N, Sutherland M, Davies RW, Kilburn D, Miller RC, Warren A, Wong R (1985) Secretion of a bacterial cellulase by yeast. Science 230:958–960. doi:10.1126/science.230.4728.958
Stadlmayr G, Mecklenbräuker A, Rothmüller M, Maurer M, Sauer M, Mattanovich D, Gasser B (2010) Identification and characterisation of novel *Pichia pastoris* promoters for heterologous protein production. J Biotechnol 150:519–529. doi:10.1016/j.jbiotec.2010.09.957
Staley CA, Huang A, Nattestad M, Oshiro KT, Ray LE, Mulye T, Li ZH, Le T, Stephens JJ, Gomez SR, Moy AD, Nguyen JC, Franz AH, Lin-Cereghino J, Lin-Cereghino GP (2012) Analysis of the 5′ untranslated region (5′UTR) of the alcohol oxidase 1 (*AOX1*) gene in recombinant protein expression in *Pichia pastoris*. Gene 496:118–127. doi:10.1016/j.gene.2012.01.006
Strahl-Bolsinger S, Gentzsch M, Tanner W (1999) Protein O-mannosylation. Biochim Biophys Acta Gen Subj 1426:297–307. doi:10.1016/S0304-4165(98)00131-7
Takagi S, Tsutsumi N, Terui Y, Kong XY (2008) Method for methanol independent induction from methanol inducible promoters in *Pichia*. United States patent US 8,143,023
Thompson CA (2010) FDA approves kallikrein inhibitor to treat hereditary angioedema. Am J Health Syst Pharm 67:93. doi:10.2146/news100005
Thor D, Xiong S, Orazem CC, Kwan A-C, Cregg JM, Lin-Cereghino J, Lin-Cereghino GP (2005) Cloning and characterization of the *Pichia pastoris MET2* gene as a selectable marker. FEMS Yeast Res 5:935–942. doi:10.1016/j.femsyr.2005.03.009
Tschopp JF, Brust PF, Cregg JM, Stillman CA, Gingeras TR (1987a) Expression of the *lacZ* gene from two methanol-regulated promoters in *Pichia pastoris*. Nucleic Acids Res 15:3859–3876
Tschopp JF, Sverlow G, Kosson R, Craig W, Grinna L (1987b) High-level secretion of glycosylated invertase in the methylotrophic yeast, *Pichia pastoris*. Nat Biotechnol 5:1305–1308. doi:10.1038/nbt1287-1305
Uchima CA, Arioka M (2012) Expression and one-step purification of recombinant proteins using an alternative episomal vector for the expression of N-tagged heterologous proteins in *Pichia pastoris*. Biosci Biotechnol Biochem 76:368–371. doi:10.1271/bbb.110628
Vadhana AKP, Samuel P, Berin RM, Krishna J, Kamatchi K, Meenakshisundaram S (2013) Improved secretion of *Candida antarctica* lipase B with its native signal peptide in *Pichia pastoris*. Enzym Microb Technol 52:177–183. doi:10.1016/j.enzmictec.2013.01.001
van der Heide M, Hollenberg CP, van der Klei IJ, Veenhuis M (2002) Overproduction of BiP negatively affects the secretion of *Aspergillus niger* glucose oxidase by the yeast *Hansenula polymorpha*. Appl Microbiol Biotechnol 58:487–494. doi:10.1007/s00253-001-0907-2
Vassileva A, Chugh DA, Swaminathan S, Khanna N (2001) Effect of copy number on the expression levels of hepatitis B surface antigen in the methylotrophic yeast *Pichia pastoris*. Protein Expr Purif 21:71–80. doi:10.1006/prep.2000.1335
Vervecken W, Kaigorodov V, Callewaert N, Geysens S, De Vusser K, Contreras R (2004) In vivo synthesis of mammalian-like, hybrid-type N-glycans in *Pichia pastoris*. 70:2639–46. doi:10.1128/AEM.70.5.2639
Vogl T, Glieder A (2013) Regulation of *Pichia pastoris* promoters and its consequences for protein production. Nat Biotechnol 30:385–404. doi:10.1016/j.nbt.2012.11.010
Walsh G (2010) Biopharmaceutical benchmarks 2010. Nat Biotechnol 28:917–924. doi:10.1038/nbt0910-917
Waterham HR, de Vries Y, Russel KA, Xie W, Veenhuis M, Cregg JM (1996) The *Pichia pastoris PER6* gene product is a peroxisomal integral membrane protein essential for peroxisome biogenesis and has sequence similarity to the Zellweger syndrome protein PAF-1. Mol Cell Biol 16:2527–2536
Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM (1997) Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene 186:37–44. doi:10.1016/S0378-1119(96)00675-0
Waters MG, Evans EA, Blobel G (1988) Prepro-alpha-factor has a cleavable signal sequence. J Biol Chem 263:6209–6214
Weinhandl K, Winkler M, Glieder A, Camattari A (2012) A novel multi-enzymatic high throughput assay for transaminase activity. Tetrahedron 68:7586–7590. doi:10.1016/j.tet.2012.06.058
Weis R, Luiten R, Skranc W, Schwab H, Wubbolts M, Glieder A (2004) Reliable high-throughput screening with *Pichia pastoris* by limiting yeast cell death phenomena. FEMS Yeast Res 5:179–189. doi:10.1016/j.femsyr.2004.06.016
Weiss HM, Haase W, Michel H, Reiländer H (1995) Expression of functional mouse 5-HT_{5A} serotonin receptor in the methylotrophic yeast *Pichia pastoris*: pharmacological characterization and localization. FEBS Lett 377:451–456
Werten MWT, de Wolf FA (2005) Reduced proteolysis of secreted gelatin and Yps1-mediated alpha-factor leader processing in a *Pichia pastoris kex2* disruptant. Appl Environ Microbiol 71:2310–2317. doi:10.1128/AEM.71.5.2310-2317.2005
Werten MWT, van den Bosch TJ, Wind RD, Mooibroek H, de Wolf FA (1999) High-yield secretion of recombinant gelatins by *Pichia pastoris*. Yeast 15:1087–1096. doi:10.1002/(SICI)1097-0061(199908)15:11<1087::AID-YEA436>3.0.CO;2-F
Whyteside G, Nor RM, Alcocer MJC, Archer DB (2011) Activation of the unfolded protein response in *Pichia pastoris* requires splicing of a *HAC1* mRNA intron and retention of the C-terminal tail of Hac1p. FEBS Lett 585:1037–1041. doi:10.1016/j.febslet.2011.02.036
Wriessnegger T, Pichler H (2013) Yeast metabolic engineering—targeting sterol metabolism and terpenoid formation. Prog Lipid Res 52:277–293. doi:10.1016/j.plipres.2013.03.001
Wriessnegger T, Gübitz G, Leitner E, Ingolic E, Cregg J, de la Cruz BJ, Daum G (2007) Lipid composition of peroxisomes from the yeast *Pichia pastoris* grown on different carbon sources. Biochim Biophys Acta 1771:455–461. doi:10.1016/j.bbalip.2007.01.004
Wriessnegger T, Leitner E, Belegratis MR, Ingolic E, Daum G (2009) Lipid analysis of mitochondrial membranes from the yeast *Pichia pastoris*. Biochim Biophys Acta 1791:166–172. doi:10.1016/j.bbalip.2008.12.017
Wu M, Shen Q, Yang Y, Zhang S, Qu W, Chen J, Sun H, Chen S (2013) Disruption of *YPS1* and *PEP4* genes reduces proteolytic degradation of secreted HSA/PTH in *Pichia pastoris* GS115. J Ind Microbiol Biotechnol 40:589–599. doi:10.1007/s10295-013-1264-8
Xuan Y, Zhou X, Zhang W, Zhang X, Song Z, Zhang Y (2009) An upstream activation sequence controls the expression of *AOX1* gene in *Pichia pastoris*. FEMS Yeast Res 9:1271–1282. doi:10.1111/j.1567-1364.2009.00571.x
Yang S, Kuang Y, Li H, Liu Y, Hui X, Li P, Jiang Z, Zhou Y, Wang Y, Xu A, Li S, Liu P, Wu D (2013) Enhanced production of recombinant secretory proteins in *Pichia pastoris* by optimizing Kex2 P1′ site. PLoS ONE 8:e75347. doi:10.1371/journal.pone.0075347
Yao XQ, Zhao HL, Xue C, Zhang W, Xiong XH, Wang ZW, Li XY, Liu ZM (2009) Degradation of HSA-AX15(R13K) when expressed in *Pichia pastoris* can be reduced via the disruption of *YPS1* gene in this yeast. J Biotechnol 139:131–136. doi:10.1016/j.jbiotec.2008.09.006
Yoshimasu MA, Ahn J-K, Tanaka T, Yada RY (2002) Soluble expression and purification of porcine pepsinogen from *Pichia pastoris*. Protein Expr Purif 25:229–236
Yu XW, Sha C, Guo YL, Xiao R, Xu Y (2013) High-level expression and characterization of a chimeric lipase from *Rhizopus oryzae* for

biodiesel production. Biotechnol Biofuels 6:29. doi:10.1186/1754-6834-6-29

Zhang Y, Teng D, Mao R, Wang X, Xi D, Hu X, Wang J (2014) High expression of a plectasin-derived peptide NZ2114 in *Pichia pastoris* and its pharmacodynamics, postantibiotic and synergy against *Staphylococcus aureus*. Appl Microbiol Biotechnol 98:681–694. doi:10.1007/s00253-013-4881-2

Zhao HL, Xue C, Wang Y, Yao XQ, Liu ZM (2008) Increasing the cell viability and heterologous protein expression of *Pichia pastoris* mutant deficient in *PMR1* gene by culture condition optimization. Appl Microbiol Biotechnol 81:235–241. doi:10.1007/s00253-008-1666-0

Zinser E, Daum G (1995) Isolation and biochemical characterization of organelles from the yeast, *Saccharomyces cerevisiae*. Yeast 11:493–536. doi:10.1002/yea.320110602

Chapter 2

Strains and vectors for protein expression in *P. pastoris*

Melanie Hirz[1*], **Mudassar Ahmad[1*]** and Helmut Schwab[1, 2]

[1]Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, A-8010 Graz, Austria,

[2]Austrian Centre of Industrial Biotechnology (ACIB), Petersgasse 14, A-8010 Graz, Austria

equal contributions

Contribution:

MH*, MA* and HS wrote the manuscript (*equal contribution). All Authors read and approved the final manuscript.

Strains and vectors for protein expression in *P. pastoris*

Melanie Hirz[1*], Mudassar Ahmad[1*] and Helmut Schwab[1, 2]

[1]*Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, A-8010 Graz, Austria,*

[2]*Austrian Centre of Industrial Biotechnology (ACIB), Petersgasse 14, A-8010 Graz, Austria*

[*]*equal contributions*

Corresponding author:

Prof. Helmut Schwab

Institute of Molecular Biotechnology

Graz University of Technology

Petersgasse 14/V

A-8010 Graz

Austria

Tel.: +43 (0)316 873 – 4070

Fax: +43 (0)316 873 – 4071

E-mail: helmut.schwab@tugraz.at

Abstract

Successful expression of heterologous proteins in *P. pastoris* is dependent on several diverse factors. One major choice to be made in the beginning is the selection of a proper host strain/vector-system that is desirable for expression. In this chapter, we will provide a summary of basic and novel vector systems available for intracellular and secreted expression. Furthermore, strains will be described, which are optimized for heterologous protein expression by providing different auxotrophies for selection or protease-deficient strains for reduced proteolysis of the desired product. Additionally, selection of Mut^+ and Mut^S strains after transformation and methods for multicopy integration will be discussed.

Key words

P. pastoris host strains, expression vectors, multicopy integration

1. Introduction

In the beginning of each expression experiment with *P. pastoris*, several questions arise to the investigator. In fact, it can have a large impact on the success of heterologous protein expression in *P. pastoris*, if the right choices concerning host strains and vectors are made from the start. Which promoter should be chosen? Which selection marker is suitable for the experiment, and does it allow for screening of multicopy integration events? Which host strain is desirable for the successful outcome of the project? In this book chapter, these questions are addressed by discussing major advantages, and – if there are any – disadvantages of using specific host strains and vectors. We provide tables that list standard strains and vectors together with recently developed alternatives. These new developments include i) clean, marker-free auxotrophic and protease-deficient strains, ii) a novel selection strategy based on *ade2* mutants, simultaneously enabling screening for multicopy strains, iii) glyco-engineered platform strains and vectors for a human-like N-glycan structure, iv) generally applied strategies for strain engineering for improved protein production and v) novel vector systems using different secretion signals. Furthermore, the influence of methanol utilization (Mut) phenotypes on protein expression will be discussed, as well as the screening for strains having multiple copies of the desired gene integrated.

2. Materials

2.1 Strains

Since *P. pastoris* has become a work horse in biotechnology, a diverse set of different strains was developed for different purposes, e.g. to reduce proteolytic activity or to alter the glycosylation pattern towards human-like N-glycan structures. These strains were all derived from the wild type strain *P. pastoris* NRRL Y-11430 (Northern Regional Research Laboratories, IL, USA), also known as CBS7435 (Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands) which was recently reclassified as *Komagataella phaffii* ***(1)***.

There are sets of strains available from companies such as Life Technologies™ and BioGrammatics, however, with the limitation that license fees have to be paid if the strains are used for industrial purposes. The wild-type *P. pastoris* CBS7435 strain, which was initially used by Philips Petroleum Company for single cell protein production, is now free to use for the scientific community as well as in industry. Therefore, this strain is of particular interest for companies for heterologous protein production. An overview of engineered strains of interest can be found in Table 1. Conveniently, there are already full genome sequences and annotations available for *P. pastoris* CBS7435 ***(2)***, *P. pastoris* GS115 ***(3)***, and *P. pastoris* CBS 704 ***(4)***. which can be accessed through *Pichia* genome browsers (http://www.pichiagenome.org and http://bioinformatics.psb.ugent.be/orcae/overview/Picpa) ***(5)***. This knowledge greatly facilitated host strain development.

2.1.1 Auxotrophic strains

During the last years, several auxotrophic strains have become available, which can be conveniently and cost-effectively used for DNA transformation and selection ***(6, 7)***. Strains with a deleted or mutated histidinol dehydrogenase gene *HIS4* are still most widely used for selection due to the broad variety of vectors available harbouring the intact *HIS4* gene for complementation. It has to be mentioned, that the *his4* auxotrophic *P. pastoris* GS115 strain was created by nitrosoguanidine mutagenesis ***(8)*** and it is therefore possible that the strain spontaneously reverts the mutation in the *HIS4* gene without having the expression cassette integrated, resulting in false-positive clones. So-called “clean” knockouts, generated by completely disrupting the coding sequence with a knockout-cassette via homologous recombination and subsequent marker recycling using the Flp-FRT recombinase system are more stable and therefore favourable ***(9)***. Clean knock-out strains auxotrophic for *his4, arg4*, *met2, lys2, pro3* and *tyr1* are available from the TU Graz *Pichia* Pool (see Table 1). Additional auxotrophic strains which were constructed by mutagenesis are available at the Keck Graduate Institute together with vectors containing the complementing biosynthetic gene ***(10)***.

Transformation experiments using auxotrophic markers are usually highly efficient with almost no background colonies appearing, as compared to antibiotic selection markers. It is, however, not directly possible to screen for multicopy strains by using standard auxotrophic markers, and vectors need to have an additional antibiotic selection marker such as Geneticin^{R} disulphate (G418). The PichiaPink^{TM} system has overcome this limitation by exploiting the feature that *ade2* defect strains appear pink due to accumulation of products of the adenine biosynthetic pathway. By using high copy plasmids, which have a truncated and thus weaker *ADE2* promoter, colonies can be directly screened on plate for multicopy integration events (see section 3.4.2) ***(11)***.

The section on vectors describes in more detail, which vectors are available harbouring auxotrophic selection markers.

2.1.2 Protease-deficient strains

In some cases, researchers might run into the problem that the desired protein is not stable under secretory expression conditions in *P. pastoris*. Proteolytic activity, arising from vacuolar, secreted or intracellular proteases after cell lysis, can be problematic and strongly exacerbate downstream processing from culture supernatants, leading to a loss of final product yield. Especially during high cell density fermentation with *P. pastoris*, it might come to cell lysis, liberating proteases into the culture medium. Over the years, several strategies have been developed to combat the proteolytic degradation of heterologous proteins, such as modification of fermentation conditions, media optimization by addition of casamino acids or soytone, and protein engineering ***(12, 13)***. Additionally, strains can be used for expression, which are deficient for the major vacuolar proteases Pep4p and Prb1p. These strains called SMD1168 (Δ*his4* Δ*pep4*), SMD1165 (Δ*his4* Δ*prb1*) and SMD1163 (Δ*his4* Δ*pep4 Δprb1*) are available from Life Technologies^{TM}. The *PEP4* gene product, an aspartyl protease, is responsible for activating itself and other proteases such as proteinase B (*PRB1*) and carboxypeptidase Y (*PRC1*). Strains deficient in *pep4* and *prb1* have therefore a strongly reduced proteolytic activity. There are several studies reporting that the use of these strains lead to increased expression of intact

protein, e.g. for expression of Human Insulin-Like Growth Factor I (IGF-I) ***(14)***, mouse 5-HT5A serotonin receptor ***(15)*** or mouse epidermal growth factor ***(12)***. However, there are also studies reporting no additional beneficial effect of using a protease-deficient strain as compared to wild type strains ***(16–18)***.

The construction of additional strains deficient for certain proteases, such as Yps1p ***(19–21)***, Kex1p ***(22, 23)*** and Kex2p ***(21)***, for improved protein expression are also reported with variable success. Knockout of the serine carboxypeptidase Kex1p, which is specific for basic amino acid residues, can be beneficial for proteins that are prone to C-terminal degradation. Typically, the beneficial effect of using protease-deficient strains is strongly dependent on the protein of interest. One cannot generally advise the use of a certain protease-deficient strain from the start, because it is reported that these strains are not as robust as WT strains and there are often several different proteases involved in degradation of the end product. Therefore, the knockout of just a single one might often not be sufficient. However, if there are problems with proteolytic degradation, it is definitely a viable option to try using protease-deficient strains for expression. Table 1 gives an overview of the protease-deficient strains available.

2.1.3 Glycoengineered strains

The early steps of N-glycosylation of proteins in the ER leading to the core glycan structure $(Man)_8(GlcNac)_2$ are highly conserved in higher eukaryotic species and yeasts. In mammals, however, final N-glycan structures are more diverse and complex as they contain besides N-acetyl glucosamine and mannose also galactose, fucose and terminal sialic acid. Yeasts such as *P. pastoris* tend to attach high-mannose glycan structures to proteins that enter the secretory pathway, although hyper-mannosylation is not as pronounced as in *S. cerevisiae*. Still, this can have severe impacts on the properties of the proteins, especially in the case of therapeutic proteins due to possible immunogenic reactions and decreased serum half-life ***(24)***. Several successful attempts have been made recently to change the glycosylation towards a more human-like N-glycan structure ***(25–27)***. BioGrammatics now offers commercially available GlycoSwitch strains, which are derived from GS115 and are also available

as *HIS4* prototrophic, protease-deficient or Mut^S variants. The major engineering steps in these strains were the knockout of Golgi-resident Och1p, an α-1,6-mannosyl-transferase located in the Golgi apparatus, which hinders the extensive addition of mannose residues, and the introduction of α-1,2-mannosidase from *Trichoderma reesei*. These modifications trim the typical yeast high-mannose structure back to a more homogenous glycan structure. Additional plasmids are available for overexpression of different glycosyltransferases to further modify glycan structures. A comprehensive guide to produce complex human-like N-glycan structures in *P. pastoris* strains using the GlycoSwitch technology is provided by Jacobs et al. ***(28)***.

Recently, Krainer et al. described the construction of a clean Och1p knockout strain in the *P. pastoris* CBS7435 Mut^S background, which has been proven to be a versatile host for secretory expression of a more uniformly mannosylated horseradish peroxidase, despite the observed growth defects ***(29)***.

Chapters 14 and 15 will give a deeper insight into post-translational modifications including N-glycosylation in *P. pastoris*.

2.1.4 General strain engineering strategies

Besides using protease-deficient strains or glyco-egnineered strains, there are several other strategies to generate efficient, high-yielding *P. pastoris* production strains. To improve folding capacity during protein secretion, co-overexpression of folding helpers such as the ER resident chaperone protein-disulfide isomerase (PDI) from either *S. cerevisiae* or *P. pastoris* turned out to be of advantage in several reported cases e. g. for secretion of human parathyroid hormone ***(18)***, *Necator americanus* secretory protein (Na-ASP1) ***(30)***, *Rhizopus chinensis* lipase ***(31)***, or *P. falciparum* transmission-blocking vaccine candidate Pfs25 ***(32)***. Although it was favourable in these cases, it seems not to be a general applicable strategy, as there are also studies reporting no or adverse effects on protein production, e.g. for A33 single chain antibody fragment secretion ***(33)***. The same study describes, however, the positive effect of overexpressing immunoglobulin binding protein (BiP), an Hsp70 class heat shock protein, on protein secretion. There are two *P. pastoris* CBS7435 Mut^S platform strains available from

VTU and TU Graz, having 1 or more copies of PDI genomically integrated. As it was already shown in several cases to be of advantage, overexpression of folding helpers is an engineering strategy worth trying.

Another strain engineering strategy was applied for the expression of mammalian membrane proteins. For expression of such complex proteins it can be favourable to offer the proper membrane surrounding directly inside the yeast cell. As yeast contains mainly ergosterol, whereas in mammalian cells the major sterol is cholesterol, an engineering approach was described where the sterol pathway was redirected towards production of cholesterol ***(34)***. Hence, expression of the mammalian Na,K-ATPase αβ1 was improved in terms of protein stability and activity. This approach might be applicable for other membrane proteins such as GPCRs as well, which need a special lipid environment to be fully functional and stable.

Table 1. *P. pastoris* host strains

Strain	Genotype	Phenotype	Source
Wildtype strains			
CBS7435 (NRRL Y-11430)	WT	WT	Centraalbureau voor Schimmelcultures, the Netherlands
CBS704 (DSMZ 70382)	WT	WT	Centraalbureau voor Schimmelcultures, the Netherlands
X-33	WT	WT	Life Technologies™
Auxotrophic strains			
GS115	*his4*	His^-	Life Technologies™
PichiaPink™ 1	*ade2*	Ade^-	Life Technologies™
KM71	*his4, aox1::ARG4, arg4*	His^-, Mut^S	Life Technologies™
KM71H	*aox1::ARG4, arg4*	Mut^S	Life Technologies™
BG09	*arg4::nourseo^R Δlys2::hyg^R*	Lys^-, Arg^-, Nourseothricin^R, Hygromycin^R	BioGrammatics
GS190	*arg4*	Arg^-	***(35)***
GS200	*arg4 his4*	His^-, Arg^-	(Waterham et al. 1996)
JC220	*ade1*	Ade^-	***(35)***
JC254	*ura3*	Ura^-	***(35)***
JC227	*ade1 arg4*	Ade^- Arg^-	***(6)***
JC300-JC308	Combinations of *ade1 arg4 his4 ura3*	Combinations of Ade^- , Arg^-, His^- , Ura^-	***(6)***
YJN165	*ura5*	Ura^-	***(36)***
[a]CBS7435 *his4*	*his4*	His^-	***(9)***
[a]CBS7435 Mut^S *his4*	*aox1, his4*	Mut^S, His^-	***(9)***
[a]CBS7435 Mut^S *arg4*	*aox1, arg4*	Mut^S, Arg^-	***(9)***
[a]CBS7435 *met2*	*met2*	Met^-	([b]*Pp*7030)
[a]CBS7435 *met2 arg4*	*met2 arg4*	Met^- Arg^-	([b]*Pp*7031)
[a]CBS7435 *met2 his4*	*met2 his4*	Met^- His^-	([b]*Pp*7032)
[a]CBS7435 *lys2*	*lys2*	Lys^-	([b]*Pp*7033)
[a]CBS7435 *lys2 arg4*	*lys2 arg4*	Lys^- Arg^-	([b]*Pp*7034)

[a]CBS7435 *lys2 his4*	*lys2 his4*	Lys$^-$ His$^-$	([b]*Pp*7035)
[a]CBS7435 *pro3*	*pro3*	Pro$^-$	([b]*Pp*7036)
[a]CBS7435 *tyr1*	*tyr1*	Tyr$^-$	([b]*Pp*7037)
Protease-deficient strains			
SMD1163	*his4 pep4 prb1*	His$^-$	***(37)***
SMD1165	*his4 prb1*	His$^-$	***(37)***
SMD1168	*his4 pep4::URA3 ura3*	His$^-$	Life Technologies™
SMD1168H	*pep4*		Life Technologies™
SMD1168 kex1::SUC2	*pep4::URA3 kex1::SUC2 his4 ura3*	His$^-$	(Boehm et al. 1999)
PichiaPink™ 2-4	Combinations of *prb1/pep4*	Ade$^-$	Life Technologies™
BG21	*sub2*		BioGrammatics
[a]CBS7435 *prc1*	*prc1*		([b]*Pp*6676)
[a]CBS7435 *sub2*	*sub2*		([b]*Pp*6668)
[a]CBS7435 *sub2*	*his4 pep4*	His$^-$	([b]*Pp*6911)
[a]CBS7435 *prb1*	*prb1*		([b]*Pp*6912)
[a]CBS7435 *his4 pep4 prb1*	*his4 pep4 prb1*	His$^-$	([b]*Pp*7013)
Glyco-engineered strains			
SuperMan$_5$	*his4 och1::pGAPTrα1,2-mannosidase*	His$^-$, Blasticidin	BioGrammatics
	och1::pGAPTrα1,2-mannosidase	Blasticidin	BioGrammatics
	pep4 och1::pGAPTrα1,2-mannosidase	Blasticidin	BioGrammatics
***Pp*FWK3**	*aox1 och1*	Mut^S, defective for mannosyltransferase	***(29)***
Other Strains			
GS241	*fld1*	Growth defect on methanol as sole C-source or methylamine as sole N-source	***(38)***
MS105	*his4 fld1*	See GS241; His$^-$	***(38)***
MC100-3	*his4 arg4 aox1::ScARG4 aox2::PpHIS4*	Mut$^-$	***(39)***
[a]CBS7435 *ku70*	*ku70*	WT	***(9)***
[a]CBS7435 *ku70 his4*	*ku70 his4*	His$^-$	***(9)***
CBS7435 *ku70 gut1*	*ku70 gut1*	Growth defect on glycerol; Zeocin™	***(9)***

CBS7435 *ku70 ade1*	*ku70 ade1*	Ade^{-}, Zeocin^{TM}	***(9)***
CBS7435 cholesterol strain	*Ku70 his4 erg5::DHCR7^{Zeo} erg6::DHCR24^{G418}*	His^{-}, Zeocin^{TM}, Geneticin^{R}	***(34)***

2.2 Vectors

Selection of an expression vector and a corresponding host strain is one crucial factor that could influence the final outcome of heterologous protein expression in *P. pastoris*. Genomic integration of expression constructs is the preferred method in *P. pastoris* due to instability issues of autonomously replicating plasmids. Therefore, all expression vectors are based on a bi-functional setup, i.e. the cloning and amplification of an expression vector is carried out in *E.coli*, followed by linearization and transformation of the expression cassette to generate expression strains. For this purpose, vectors are equipped with an origin of replication and a marker cassette for plasmid maintenance and selection in *E. coli*. In addition, each vector contains a marker cassette for selection in *P. pastoris* and an expression cassette, which is composed of a promoter (in most cases P_{Aox1} or P_{GAP}), a multiple cloning site (MCS) and a terminator sequence. The gene of interest can be cloned by using any of the restriction sites. However, it is recommended to use restriction sites that introduce least interfering nucleotides between the promoter sequence and the start codon of the heterologous gene. Some vectors also contain an additional 3' untranslated region of the *AOX1* gene (3'UTR) for targeting the expression cassette via homologous recombination into the *AOX1* locus to generate strains with a slow methanol utilization phenotype (Mut^S). To allow secretion of heterologous proteins, different secretion signals, e.g. from *S. cerevisiae* alpha mating factor or *PHO1* are fused N-terminally with the protein of interest.

Recently, a new breed of vectors has emerged, where cloning is performed based on type IIS restriction enzymes. This cloning strategy enables fusion of the gene of interest seamlessly with upstream or downstream sequences to circumvent any potential problems arising from intervening nucleotides (Vogl et al. 2014, manuscript under revision). This section will give an overview of standard and novel promoters, biosynthetic or antibiotic markers as well as secretion signals commonly used in *P. pastoris* expression vectors. A non-exhaustive list of commonly used vectors is provided in the table 2.

2.2.1 Promoters

Alcohol oxidase 1 promoter (P_{AOX1}) is by far the most studied and commonly used promoter to drive protein expression in *P. pastoris*. P_{AOX1} is a tightly regulated promoter which is repressed in the presence of glucose and can be induced up to 1000-fold by growing cells on methanol as a sole carbon source ***(40)***. The higher degree of process control makes this promoter ideal for heterologous protein expression by uncoupling the growth phase from production phase, particularly in case of toxic proteins. Alternative strong, methanol inducible promoters such as the formaldehyde dehydrogenase (*FLD1*) promoter and the dihydroxy acetone synthase (*DAS*) promoter have also been used to express proteins in *P. pastoris* ***(38, 41)***. In addition to methanol, the P_{FLD1} promoter can be induced by methylamine as a single nitrogen source in presence of glucose as a carbon source ***(38)***.

In some cases, the use of a strong promoter may lead to undesirable results, especially for secretory expression, as more time is required for the proper protein folding and processing of recombinant proteins through the secretory pathway ***(42, 43)***. Promoters having a similar regulatory profile as of P_{AOX1} but moderate expression levels such as the alcohol oxidase 2 promoter (P_{AOX2}) can be used for these particular applications. Another constitutively expressed promoter derived from peroxisomal matrix protein (Pex8) has also been rarely used to this end. P_{PEX8} can transcribe proteins on glucose at low but considerable levels and can be induced up to 3-5 times with methanol or oleate ***(43)***. In addition, several variants of P_{AOX1} are available for “fine-tuned” expression of heterologous genes. These promoter variants have been shown to possess a range of 6% to 160% of the wild type promoter activity ***(44)***. Some variants also show de-repression under glycerol feeding and were employed recently to secrete 18 g/L of TcHB1 in small scale fermenter cultivations ***(44–46)***. These promoter variants and respective expression vectors are available from VTU technology (www.vtu-technology.com).

However, methanol is an extremely toxic and combustible substance and its use in large scale fermentations requires specialized and costly handling procedures. In addition, being a derivative of

petroleum, its use is not appropriate in food industry. Furthermore, during high density fermentations, excess of methanol can promote cell death phenomena, releasing intracellular proteins into the culture broth, which not only complicates the downstream processing and purification but may also result in undesirable proteolysis of secreted recombinant proteins ***(47)***. Therefore, several constitutive promoters such as P_{GAP}, P_{TEF1}, and P_{YPT1} have been used to express proteins ***(48–50)***. P_{GAP} is the most commonly used alternative promoter in place of P_{AOX1} and has been shown to express heterologous proteins to similar levels like P_{AOX1} ***(48)***. A major advantage of using constitutive promoters is that it eliminates the need for switching carbon sources, thereby reducing production time, effort and increasing overall productivity. However, constitutive promoters can only be used to express proteins that are not toxic to the host cell. An extensive review of available *P. pastoris* promoters and there regulatory properties has been recently published by Vogl et al. ***(51)***.

2.2.2 Selectable markers

The genetic modifications and amplification of all *P. pastoris* vectors are carried out in *E. coli*. The majority of vectors contain the *bla* gene encoding ampicillin resistance for selection in *E. coli*. However, several new vectors are available either from Life Technologies™ or TU Graz that contain a single resistance marker gene, *Sh ble* from *Streptoalloteichus hindustanus*, which confers resistance to Zeocin™ in both organisms. The two most frequently used selection markers are *HIS4* and Zeocin™. In addition, several other auxotrophic and dominant selection markers are available, which are discussed below.

2.2.3 Auxotrophic selection markers

Auxotrophic selection markers are preferred over dominant selection markers due to their ease of handling, cost effectiveness and superior genetic stability of generated expression clones. However, they can only be used with corresponding auxotrophic strains. Secondly, due to multiple markers for selection in bacteria and yeast, auxotrophic marker based expression vectors are usually larger in size than Zeocin™/Blasticidin based vectors, which might complicate the cloning and transformation. A

number of auxotrophic selection markers, readily cloned in expression vectors, are available for genetic manipulation of *P. pastoris* e.g., *HIS4* (histidinol dehydrogenase) ***(8)***, *ARG4* (argininosuccinate lyase), *ADE1* (PR-amidoimidazolesuccinocarboxamide synthase), *URA3* (orotidine 5'-phosphate decarboxylase) ***(6)***, *ADE2* (phosphoribosylaminoimidazole carboxylase) ***(11)***, *URA5* (orotate-phosphoribosyl transferase) ***(36)***, *MET2* (homoserine-O-transacetylase) ***(7)***, and *GUT1* (glycerol kinase 1) ***(9)***. In addition, a set of expression vectors containing *ARG1*, *ARG2*, *ARG3* and *HIS1*, *HIS2* and *HIS5* as auxotrophic markers were constructed by Nett et al ***(52)***. These vectors can be used to disrupt Arginine pathway genes with Histidine pathway genes and vice versa in a sequential manner. During the process, multiple heterologous genes can be integrated at defined loci into the genome of *P. pastoris*. The main disadvantage of this strategy is the need of time-consuming screening of transformants to find the auxotrophic strains with targeted integration of the expression cassette, which can be then used for further transformations.

2.2.3.1 Dominant selection markers

One of the major advantages of dominant selection markers is that they are not limited to a complementary genetic host and can therefore be used for genetic modifications of wild type or industrial production strains. Another advantage of many of these dominant markers is the significantly reduced size of the expression vectors due the possibility to use one single marker for selection in both, *E. coli* and *P. pastoris*. Furthermore, certain markers can be conveniently used for screening of multicopy clones (see section 3.4). There are several dominant selection markers available conferring resistance to Zeocin^{TM} (*She ble*) ***(53)***, Geneticin^{R}/G418 (*Tn903kan*^{r}) ***(54, 55)***, Blasticidin S (*BSD*) ***(56)***, Formaledyde (*FLD1)* ***(57)***, hygromycin (*HPH*) ***(58)*** and nourseothricin (*NAT1*) ***(59)***.

2.2.4 Secretion signals

One of the main attractive features of *P. pastoris* is its ability to secrete properly processed and active recombinant proteins into the culture media. *P. pastoris* secretes only low levels of endogenous proteins and as a result the secreted protein is often a major protein in the supernatant, greatly

reducing the downstream processing otherwise needed for the purification of intracellular proteins ***(60)***. The decision to target a recombinant protein to the secretory pathway depends upon the native situation of the protein in its natural host. Thus, protein secretion in *P. pastoris* is worth trying if the protein is naturally secreted. Intracellular proteins are usually more problematic and therefore intracellular expression is most likely more promising. However, there are some rare examples reported in literature where researchers succeeded in secreting an intracellular protein ***(61, 62)***. After the decision has been made to secrete a protein, the second point to consider is the choice of the secretion signal. The most commonly used strategy is to design at least two expression constructs, one with the native secretion signal and one with the *S. cerevisiae* α-mating factor secretion signal (α-MF) ***(63–65)***.

Several other foreign secretion signals have also been used to direct proteins to the secretory pathway in *P. pastoris* with varying success such as PHO1 (acid phosphatase secretion signal) ***(66)***, SUC2 (Invertase signal sequence) ***(65, 67)***, PHA-E (Phaseolus vulgaris agglutinin- E form) ***(68)*** and Killer Toxin Prepro signal sequence ***(61, 69)***. However, these secretion signals have not been used extensively as there is either limited data available or results have been variable for a broad range of proteins. In addition, there is a set of eight secretion signals available from Life Technologies™ to trouble shoot for the best working signal sequence (www.lifetechnologies.com). The most commonly, and by far the most successfully used secretion signal is the α-mating factor (α-MF) prepro signal peptide, which is readily available in most of the expression vectors available from Life Technologies™, TU Graz, Biogrammatics and Keck Graduate Institute. In order to have authentic N-terminal amino acids of the secreted recombinant protein, *Xho*I or compatible *Sal*I sites can be used to clone the heterologous gene in frame with α-MF, but the *KEX2* cleavage site (Lys-Arg) needs to be restored.

There are two common problems reported for the α-MF secretion signal. The first one is the occurrence of partly incomplete signal peptide processing at the *KEX2* cleavage site (Lys-Arg) due to inefficient *KEX2* protease activity. This might be overcome by including Glu-Ala repeats between the *KEX2* cleavage site and the coding sequence of the gene of interest or by optimizing the *KEX2*

recognition site ***(70, 71)***. The second problem can be incomplete processing of Glu-Ala residues by *STE13* protease in the Golgi complex, resulting in heterogeneous N-termini of the recombinant protein. Four different variants of the α-MF signal sequence along with seven other alternative secretion signals are readily available from DNA 2.0 (www.dna20.com).

3. Methods

3.1 Cloning strategies

Most of the conventional vector systems are equipped with a multiple cloning site (MCS) for cloning of the gene of interest based on type II restriction enzymes, which often leaves a cloning scar resulting in a suboptimal 5' untranslated region (5' UTR). Mutations in 5' UTR have been shown to negatively affect the translation efficiency of heterologous gene expression ***(72, 73)***. Therefore, several new vector systems have been developed to clone the gene of interest seamlessly with flanking regulatory sequences e.g. promoter, terminator, secretion signals and fusion tags. A set of expression vectors termed "*Pichia* pool 2 plasmid family" is available from TU Graz, where cloning is performed via a single *Eco*RI site, which has been introduced into the promoter region. Thereby, the immediate 5' region upstream of ATG is identical to the natural *AOX1* gene. The Kozak Consensus sequence (TTCGAAACG) between *Eco*RI and the start ATG has to be added to the coding region of the gene when vectors for intracellular expression are used. ***(63)***. There are also some vector systems available from TU Graz, Biogrammatics (www.biogrammatics.com) and DNA 2.0 (www.dna20.com) based on type II S restriction enzymes, which cleave outside of their recognition sequence. Recently, a set of 40 expression vectors based on a novel cloning strategy termed restriction site free cloning (RSFC) was developed employing type II S endonuclease, Mly II, in our lab. A single PCR product can be cloned in frame with multiple promoters, secretion signals and N and C- terminal tags to screen for optimal protein expression and purification (Vogl et al. 2014, manuscript under revision). However, the proper orientation of the cloned gene has to be confirmed prior to transformation into the expression host.

When vectors for secretory expression are used, it is of great importance to maintain the *KEX2* cleavage site. If restriction enzymes are chosen, which cut out the Lys-Arg residues, they need to be added again in frame, to ensure the proper processing of the secreted protein.

3.2 Mut$^+$/Mut^S screening

P. pastoris contains two alcohol oxidase genes, *AOX1* and *AOX2*, which are necessary for the cells to grow on methanol as sole carbon source ***(39, 74)***. The two peroxisomal enzymes catalyse the first step in the methanol assimilation pathway by oxidizing methanol to hydrogen peroxide and formaldehyde. Despite the high similarity of Aox proteins, the *AOX1* gene product is synthesized to a much higher extend due to the very strong *AOX1* promoter, resulting in approximately 30% of total cell protein upon growth on methanol ***(75)***. After a heterologous gene has been integrated into the *AOX1* locus via double-cross over, the cells become defective for *AOX1* and have to rely solely on *AOX2* for methanol utilization, which results in a slow growth phenotype on methanol containing media (Mut^S, methanol utilization slow). If the *AOX1* gene remains intact, cells grow very well on methanol (Mut$^+$, methanol utilization plus).

There are several reports that the use of a Mut^S over a Mut$^+$ strain is of advantage, for example in the case of horseradish peroxidase ***(76)*** or the antibody single chain variable fragment scFvA33 ***(77)***. On the contrary, there are also studies showing high expression levels using Mut$^+$ strains, e.g. for *Coprinus cinereus* peroxidase ***(78)***, or studies where the methanol utilization phenotype seemed not to be of particular importance for product yields, like in the case of tetanus toxin fragment C ***(79)***. In the case of *Rhizopus oryzae* lipase, the maximum lipase activity and the specific activity were higher using a Mut^S strain, but the productivity was higher for the Mut$^+$ strain, meaning that more enzyme was produced in shorter fermentation periods with the same amount of biomass. The same study describes furthermore a different behaviour for multicopy strains. The Mut$^+$ strains seemed to be more robust when expressing lipase from multiple gene copies ***(80)***. A clear advantage of using a Mut^S strain is, however, that less amounts of methanol are necessary and the fermentation processes become easier to control.

It has to be evaluated for each protein of interest, which phenotype performs best during fermentation. After transformation of the vector, clones should be investigated for growth behaviour

on minimal media containing either gluose or methanol to determine their methanol utilization phenotype. This can be easily done by streaking clones on minimal methanol and minimal dextrose plates. If it is already known for a protein, that Mut^S is the preferred phenotype, it is possible to use *P. pastoris* Mut^S strains from the start, which are available from Life Technologies^{TM} (KM71) or from the TU Graz *Pichia* Pool.

3.3 Single copy integration – targeting a defined locus

Sometimes it is desired to integrate one single copy of a heterologous gene into a specific locus. This can be the case if Mut^S strains are desired for heterologous expression, if overexpression of other helper-proteins such as PDI has positive effects on expression, or if it is already known that multiple copies do not enhance protein expression efficiency.

The problems with generating multicopy strains might be genetic instability and, as the integration locus is often random and therefore unknown, pleiotropic effects might arise. It is often difficult to compare engineered *P. pastoris* strains to detect positive effects of gene overexpression, if the exact genotypes of the strains are not known. For that reason, it is necessary to target a desired locus to be able to obtain comparable results. This can be achieved by integration into *AOX1* locus followed by screening for Mut^S mutants (see chapter 3.3) or by Integration into *HIS4* locus followed by screening for histidine auxotrophic mutants. Vectors for integration into the *AOX1* or the *HIS4* locus are available from the TU Graz *Pichia* Pool. The advantage of integration into the *HIS4* locus is furthermore that by generation of a histidine auxotrophic strain an additional marker becomes available for further transformations. Correct integration should be also confirmed by PCR. To ensure, that there are no further copies integrated randomly into the genome, copy numbers can also be determined.

3.4 Multicopy integration

Integration of linear expression cassettes into the genome of *P. pastoris* is mostly preferred over autonomously replicating plasmids due to superior genetic stability of final expression strains. One of

the key strategies to achieve maximal protein expression in *P. pastoris*, beside minimising the negative effects of suboptimal 5' untranslated region, mRNA secondary structure, protein stability and locus of integration, is by increasing the copy number of heterologous genes. In case of intracellular expression, there seems to be a correlation of increased copy numbers with increased expression levels. However, this may not hold true for secretory expression due to a possible overload of the secretory pathway. Several well established protocols are available to generate multicopy expression strains and have been reviewed in detail elsewhere ***(81, 82)***. Briefly, one of these methods includes direct screening for spontaneously occuring multicopy strains based on protein expression levels using sodium dodecyl sulphate-polyacylamide gel electrophoresis (SDS-PAGE), immunoblotting, direct enzymatic assays, or by DNA hybridization techniques at genomic level.

Another strategy is based on *in vitro* multimerziation of expression cassettes using different expression plasmids provided with specific restriction sites. Plasmids for this purpose are available from either Life Technologies™ (i.e. pAO815) or from Keck Graduate Institute ***(6, 7)*** (see Tables 2 and 3). The main disadvantage of this strategy is the time and effort required to perform increasingly difficult cloning steps and problems associated with transformation to generate expression strains due to the extremely large size of the resulting plasmid. However, this strategy is ideal for situations where absolutely defined conditions are required. This could also be combined with *in vivo* multimerziation using post-translational vector amplification (PTVA) ***(83)*** or integration into the rDNA locus to generate extremely high copy number clones ***(84)***.

The fastest and thereby most commonly applied method to generate multicopy clones is still the screening of transformants for increased resistance on high antibiotic concentrations using Geneticin/G418 ***(55, 79)***, Zeocin™ ***(53)***, or hygromycin ***(58)***. It is also possible to screen clones for enhanced resistance to other substances such as formaldehyde ***(57)*** or 3-amino-1,2,4 triazol (3-AT) ***(85)***.

3.4.1 Generation of Multicopy strains using dominant markers (Zeocin™, G418 and Blasticidin)

The most widely used method to screen for multicopy transformants is based on selection of transformants on increasing concentrations of Zeocin™ antibiotic. Resistance to Zeocin™ is conferred by the *Sh ble* gene product, which sequesters the Zeocin™ glycopeptide by stoichiometrically binding it instead of catalysing its degradation ***(86)***. Therefore, increased resistance of transformants to the drug would most probably result from increased expression of *Sh ble* gene product. In most of the *P. pastoris* expression plasmids, expression of the *Sh ble* gene is controlled by a strong constitutive promoter such as P_{TEF1} or P_{ILV5} ***(9, 53)***, which puts a heavy burden on the cells. This could be a possible reason for existence of low copy transformants even when they are selected on highest Zeocin™ concentrations, e.g. 2000 µg/ml ***(81, 87, 88)***. With the aim to render selection conditions more stringent, we have recently constructed expression vectors (*Pichia* Pool 2, TU Graz), using the weaker *P. pastoris ARG4* promoter to drive expression of the Zeocin™ resistance gene *Sh ble*. The basal expression levels from this promoter ensure that transformants bearing single to multiple copies can be selected on a range of 25 – 400 µg/mL of Zeocin™. Only a few colonies are formed on higher antibiotic concentrations, with a high chance of being multicopy clones, making the screening process easier and more efficient.

To generate multicopy expression strains based on increasing resistance to Zeocin™, G418/Geneticin® or Blasticidin, the linearized expression vector should be transformed into electrocompetent or spheroplasted cells. Immediately after electroporation the cells are re-suspended in 1 mL of BYPD and 1 M sorbitol (1:1) and are regenerated at 28°C, 200 rpm for 2 hours[1]. Selection then can be performed on BYPD plates containing different concentrations of antibiotic (for Zeocin™ 100-2000 µg/mL, for G418/Geneticin® 500 - 1000 µg/ml, and for Blasticidin 50 – 500 µg/ml).

It is described, and we also observed in our lab, that chances of generating multiple integrations are increased, if the expression vector is linearized with *SacI* for insertions at the *AOX1* locus, followed by transformation into Mut^S strains, e.g. KM71 or CBS7435 Mut^S ***(89)***.

3.4.2 Generation of Multicopy strains using Pichia Pink

Recently, a new colour based method was developed by Du et al. employing an attenuated *ADE2* gene and its complementary expression using native truncated promoters to compensate adenine auxotrophy ***(11)***. The gene product of *ADE2*, phosphoribosylaminoimidazole carboxylase, is required to catalyse the sixth step in the formation of purine nucleotides. Inefficient expression of the marker gene from a truncated promoter results in a build-up of purine precursors inside the cell, giving it a reddish colour. Based on this principle, two vectors namely pPink-LC and pPink-HC are available from Life Technologies^{TM} for selection of single copy and high copy clones, respectively. In pPink-HC, the expression of *ADE2* is controlled by a truncated and thereby weaker promoter. Hence, only clones having multiple copies integrated are able to produce enough protein to supplement *ADE2* auxotrophy. Consequently, transformants having multiple integration of expression cassette can be readily identified based on the white colour of the colonies formed.

Notes

1. In order to have more stringent conditions for selection of multicopy transformants and to eliminate single copy transformants from plates with higher Zeocin^{TM} concentrations, regeneration time should be kept low and only bigger colonies appearing after three days of incubation should be selected.
2. Screening on G418 is sensitive to cell density.
3. Cloning of multiple expression cassettes into a single vector may lead to rearrangements in *E. coli*.

References

1. C.P. Kurtzman (2009) Biotechnological strains of *Komagataella* (*Pichia*) *pastoris* are *Komagataella phaffii* as determined from multigene sequence analysis., Journal of industrial microbiology & biotechnology. 36, 1435–8.

2. A. Küberl, J. Schneider, G.G. Thallinger, et al. (2011) High-quality genome sequence of *Pichia pastoris* CBS7435., Journal of biotechnology. 154, 312–320.

3. K. De Schutter, Y.-C. Lin, P. Tiels, et al. (2009) Genome sequence of the recombinant protein production host *Pichia pastoris*., Nature biotechnology. 27, 561–566.

4. D. Mattanovich, A. Graf, J. Stadlmann, et al. (2009) Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*., Microbial cell factories. 8, 29.

5. D. Mattanovich, N. Callewaert, P. Rouzé, et al. (2009) Open access to sequence: browsing the *Pichia pastoris* genome., Microbial cell factories. 8, 53.

6. G.P. Lin, J. Lin, A. Jay, et al. (2001) New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of *Pichia pastoris*., Gene. 263, 159–69.

7. D. Thor, S. Xiong, C.C. Orazem, et al. (2005) Cloning and characterization of the *Pichia pastoris MET2* gene as a selectable marker., FEMS yeast research. 5, 935–942.

8. J.M.M. Cregg, K.J.J. Barringer, A.Y. Hessler, et al. (1985) *Pichia pastoris* as a host system for transformations., Molecular and Cellular …. 5, 3376–3385.

9. L. Näätsaari, B. Mistlberger, C. Ruth, et al. (2012) Deletion of the *Pichia pastoris ku70* homologue facilitates platform strain generation for gene expression and synthetic biology, PLoS ONE. 7,.

10. Keck Graduate Institute, Faculty and Research, James M. Cregg. http://www.kgi.edu/faculty-and-research/profiles/james-m-cregg.html. Last accessed on September 9, 2014.

11. M. Du, M.B. Battles, and J.H. Nett (2012) A color-based stable multi-copy integrant selection system for *Pichia pastoris* using the attenuated *ADE1* and *ADE2* genes as auxotrophic markers., Bioengineered bugs. 3, 32–7.

12. J.J. Clare, M.A. Romanes, F.B. Rayment, et al. (1991) Production of mouse epidermal growth factor in yeast: high-level secretion using *Pichia pastoris* strains containing multiple gene copies, Gene. 105, 205–212.

13. H.L. Zhao, C. Xue, Y. Wang, et al. (2008) Increasing the cell viability and heterologous protein expression of *Pichia pastoris* mutant deficient in *PMR1* gene by culture condition optimization., Applied microbiology and biotechnology. 81, 235–41.

14. R.A. Brierley (1998) Secretion of recombinant human insulin-like growth factor I (*IGF*-I)., Methods in molecular biology (Clifton, N.J.). 103, 149–77.

15. H.M. Weiss, W. Haase, H. Michel, et al. (1995) Expression of functional mouse 5-HT5A serotonin receptor in the methylotrophic yeast *Pichia pastoris*: pharmacological characterization and localization., FEBS letters. 377, 451–456.

16. A. Henry, C.L. Masters, K. Beyreuther, et al. (1997) Expression of human amyloid precursor protein ectodomains in *Pichia pastoris*: analysis of culture conditions, purification, and characterization., Protein expression and purification. 10, 283–291.

17. M.C. Colao, S. Lupino, A.M. Garzillo, et al. (2006) Heterologous expression of *lcc1* gene from Trametes trogii in *Pichia pastoris* and characterization of the recombinant enzyme., Microbial cell factories. 5, 31.

18. R. Vad, E. Nafstad, L.A. Dahl, et al. (2005) Engineering of a *Pichia pastoris* expression system for secretion of high amounts of intact human parathyroid hormone., Journal of biotechnology. 116, 251–260.

19. M. Wu, Q. Shen, Y. Yang, et al. (2013) Disruption of *YPS1* and *PEP4* genes reduces proteolytic degradation of secreted *HSA/PTH* in *Pichia pastoris* GS115., Journal of industrial microbiology & biotechnology. 40, 589–599.

20. X.Q. Yao, H.L. Zhao, C. Xue, et al. (2009) Degradation of *HSA-AX15(R13K)* when expressed in *Pichia pastoris* can be reduced via the disruption of *YPS1* gene in this yeast., Journal of biotechnology. 139, 131–136.

21. M.W.T. Werten and F. a. De Wolf (2005) Reduced proteolysis of secreted gelatin and *Yps1*-mediated alpha-factor leader processing in a *Pichia pastoris kex2* disruptant., Applied and environmental microbiology. 71, 2310–2317.

22. Z. Ni, X. Zhou, X. Sun, et al. (2008) Decrease of hirudin degradation by deleting the *KEX1* gene in recombinant *Pichia pastoris*., Yeast (Chichester, England). 25, 1–8.

23. T. Boehm, S. Pirie-Shepherd, L.B. Trinh, et al. (1999) Disruption of the *KEX1* gene in *Pichia pastoris* allows expression of full-length murine and human endostatin., Yeast (Chichester, England). 15, 563–572.

24. C.E. Ballou (1990), Isolation, characterization, and properties of *Saccharomyces cerevisiae mnn* mutants with nonconditional protein glycosylation defects, http://www.sciencedirect.com/science/article/pii/007668799085038P\nhttp://www.ncbi.nlm.nih.gov/pubmed/2199792.

25. S.R. Hamilton, R.C. Davidson, N. Sethuraman, et al. (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins., Science (New York, N.Y.). 313, 1441–1443.

26. W. Vervecken, V. Kaigorodov, N. Callewaert, et al. (2004) In vivo synthesis of mammalian-like, hybrid-type N-glycans in *Pichia pastoris*, Applied and Environmental Microbiology. 70, 2639–2646.

27. J.H. Nett, T.A. Stadheim, H. Li, et al. (2011) A combinatorial genetic library approach to target heterologous glycosylation enzymes to the endoplasmic reticulum or the Golgi apparatus of *Pichia pastoris*., Yeast (Chichester, England). 28, 237–252.

28. P.P. Jacobs, S. Geysens, W. Vervecken, et al. (2009) Engineering complex-type N-glycosylation in *Pichia pastoris* using GlycoSwitch technology., Nature protocols. 4, 58–70.

29. F.W. Krainer, C. Gmeiner, L. Neutsch, et al. (2013) Knockout of an endogenous mannosyltransferase increases the homogeneity of glycoproteins produced in *Pichia pastoris*., Scientific reports. 3, 3279.

30. M. Inan, D. Aryasomayajula, J. Sinha, et al. (2006) Enhancement of protein secretion in *Pichia pastoris* by overexpression of protein disulfide isomerase., Biotechnology and bioengineering. 93, 771–778.

31. C. Sha, X. Yu, N. Lin, et al. (2013) Enhancement of lipase r27RCL production in *Pichia pastoris* by regulating gene dosage and co-expression with chaperone protein disulfide isomerase., Enzyme and microbial technology. 53, 438–43.

32. C.W. Tsai, P.F. Duggan, R.L. Shimp, et al. (2006) 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. 121, 458–70.

33. L.M. Damasceno, K. a. Anderson, G. Ritter, et al. (2007) Cooverexpression of chaperones for enhanced secretion of a single-chain antibody fragment in *Pichia pastoris*., Applied microbiology and biotechnology. 74, 381–389.

34. M. Hirz, G. Richter, E. Leitner, et al. (2013) A novel cholesterol-producing *Pichia pastoris* strain is an ideal host for functional expression of human *Na,K-ATPase α3β1* isoform., Applied microbiology and biotechnology. 97, 9465–78.

35. J.M. Cregg, S. Shen, M. Johnson, et al. (1998) Classical genetic manipulation., Methods in molecular biology (Clifton, N.J.). 103, 17–26.

36. J.H. Nett and T.U. Gerngross (2003) Cloning and disruption of the *PpURA5* gene and construction of a set of integration vectors for the stable genetic modification of *Pichia pastoris*., Yeast (Chichester, England). 20, 1279–1290.

37. M. a Gleeson, C.E. White, D.P. Meininger, et al. (1998) Generation of protease-deficient strains and their use in heterologous protein expression., Methods in molecular biology (Clifton, N.J.). 103, 81–94.

38. S. Shen, G. Sulter, T.W. Jeffries, et al. (1998) A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast Pichia pastoris., Gene. 216, 93–102.

39. J.M. Cregg, K.R. Madden, K.J. Barringer, et al. (1989) Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*., Molecular and cellular biology. 9, 1316–1323.

40. G.P.L. Cereghino, J.L. Cereghino, C. Ilgen, et al. (2002) Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*, 329–332.

41. J.F. Tschopp, P.F. Brust, J.M. Cregg, et al. (1987) Expression of the lacZ gene from two methanol-regulated promoters in *Pichia pastoris*, Nucleic Acids Research. 15, 3859–3876.

42. H. Hohenblum, B. Gasser, M. Maurer, et al. (2004) Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*., Biotechnology and bioengineering. 85, 367–75.

43. J.L. Cereghino and J.M. Cregg (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*., FEMS microbiology reviews. 24, 45–66.

44. F.S. HartnerF.S., C. Ruth, D. Langenegger, et al. (2008) Promoter library designed for fine-tuned gene expression in *Pichia pastoris*., Nucleic acids research. 36, e76.

45. A. Mellitzer, C. Ruth, C. Gustafsson, et al. (2014) Synergistic modular promoter and gene optimization to push cellulase secretion by *Pichia pastoris* beyond existing benchmarks., Journal of biotechnology.

46. C. Ruth, T. Zuellig, a Mellitzer, et al. (2010) Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in *Pichia pastoris*., Systems and synthetic biology. 4, 181–191.

47. J. Sinha, B. a. Plantz, M. Inan, et al. (2005) Causes of proteolytic degradation of secreted recombinant proteins produced in methylotrophic yeast *Pichia pastoris*: Case study with recombinant ovine interferon-??, Biotechnology and Bioengineering. 89, 102–112.

48. H.R. Waterham, M.E. Digan, P.J. Koutz, et al. (1997) Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter., Gene. 186, 37–44.

49. G. Stadlmayr, A. Mecklenbräuker, M. Rothmüller, et al. (2010) Identification and characterisation of novel *Pichia pastoris* promoters for heterologous protein production., Journal of biotechnology. 150, 519–529.

50. I.B. Sears, J. O'Connor, O.W. Rossanese, et al. (1998) A versatile set of vectors for constitutive and regulated gene expression in *Pichia pastoris*., Yeast (Chichester, England). 14, 783–790.

51. T. Vogl and A. Glieder (2013) Regulation of *Pichia pastoris* promoters and its consequences for protein production., New biotechnology. 30, 385–404.

52. J.H. Nett, N. Hodel, S. Rausch, et al. (2005) Cloning and disruption of the *Pichia pastoris* ARG1, ARG2, ARG3, HIS1, HIS2, HIS5, HIS6 genes and their use as auxotrophic markers., Yeast (Chichester, England). 22, 295–304.

53. D.R. Higgins, K. Busser, J. Comiskey, et al. (1998) Small vectors for expression based on dominant drug resistance with direct multicopy selection., Methods in molecular biology (Clifton, N.J.). 103, 41–53.

54. C.A. Scorer, J.J. Clare, W.R. McCombie, et al. (1994) Rapid selection using G418 of high copy number transformants of *Pichia pastoris* for high-level foreign gene expression., Bio/technology (Nature Publishing Company). 12, 181–184.

55. J. Lin-Cereghino, M.D.M. Hashimoto, A. Moy, et al. (2008) Direct selection of *Pichia pastoris* expression strains using new G418 resistance vectors., Yeast (Chichester, England). 25, 293–299.

56. M. Kimura, A. Takatsuki, and I. Yamaguchi (1994) Blasticidin S deaminase gene from Aspergillus terreus (BSD): a new drug resistance gene for transfection of mammalian cells, Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. 1219, 653–659.

57. A.J. Sunga and J.M. Cregg (2004) The *Pichia pastoris* formaldehyde dehydrogenase gene (FLD1) as a marker for selection of multicopy expression strains of P. pastoris., Gene. 330, 39–47.

58. J. Yang, L. Nie, B. Chen, et al. (2014) Hygromycin-resistance vectors for gene expression in *Pichia pastoris*, Yeast. 31, 115–125.

59. M.-T. Chen, S. Lin, I. Shandil, et al. (2012) Generation of diploid *Pichia pastoris* strains by mating and their application for recombinant protein production., Microbial cell factories. 11, 91.

60. J. Lin-Cereghino and G.P. Lin-Cereghino (2007) Vectors and strains for expression., Methods in molecular biology (Clifton, N.J.). 389, 11–26.

61. A. Eiden-Plach, T. Zagorc, T. Heintel, et al. (2004) Viral preprotoxin signal sequence allows efficient secretion of green fluorescent protein by Candida glabrata, *Pichia pastoris*, Saccharomyces cerevisiae, and Schizosaccharomyces pombe., Applied and environmental microbiology. 70, 961–966.

62. X.-L. Shi, M.-Q. Feng, J. Shi, et al. (2007) High-level expression and purification of recombinant human catalase in *Pichia pastoris*., Protein expression and purification. 54, 24–29.

63. M. Ahmad, M. Hirz, H. Pichler, et al. (2014), Protein expression in *Pichia pastoris*: Recent achievements and perspectives for heterologous protein production, http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4047484&tool=pmcentrez&rendertype=abstract.

64. D.R. Higgins and J. Cregg (1998) Pichia Protocols, Humana Press, New Jersey.

65. E. Paifer, E. Margolles, J. Cremata, et al. (1994) Efficient expression and secretion of recombinant alpha amylase in *Pichia pastoris* using two different signal sequences., Yeast (Chichester, England). 10, 1415–1419.

66. W.E. Payne, P.M. Gannon, and C. a Kaiser (1995) An inducible acid phosphatase from the yeast *Pichia pastoris*: characterization of the gene and its product., Gene. 163, 19–26.

67. C.N. Chang, M. Matteucci, L.J. Perry, et al. (1986) Saccharomyces cerevisiae secretes and correctly processes human interferon hybrid proteins containing yeast invertase signal peptides., Molecular and cellular biology. 6, 1812–1819.

68. R.J.M. Raemaekers, L. de Muro, J.A. Gatehouse, et al. (1999) Functional phytohemagglutinin (PHA) and Galanthus nivalis agglutinin (GNA) expressed in *Pichia pastoris* correct N-terminal processing and secretion of heterologous proteins expressed using the PHA-E signal peptide., European journal of biochemistry / FEBS. 265, 394–403.

69. M. Tokunaga, N. Wada, and F. Hishinuma (1987) A novel yeast secretion vector utilizing secretion signal of killer toxin encoded on the yeast linear DNA plasmid pGKL1., Biochemical and biophysical research communications. 144, 613–9.

70. S. Yang, Y. Kuang, H. Li, et al. (2013) Enhanced production of recombinant secretory proteins in *Pichia pastoris* by optimizing Kex2 P1' site., PloS one. 8, e75347.

71. a J. Brake, J.P. Merryweather, D.G. Coit, et al. (1984) Alpha-factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae., Proceedings of the National Academy of Sciences of the United States of America. 81, 4642–4646.

72. C. a Staley, A. Huang, M. Nattestad, et al. (2012) Analysis of the 5' untranslated region (5'UTR) of the alcohol oxidase 1 (AOX1) gene in recombinant protein expression in *Pichia pastoris*., Gene. 496, 118–127.

73. K. Sreekrishna, R.G. Brankamp, K.E. Kropp, et al. (1997) Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*., Gene. 190, 55–62.

74. S.B. Ellis, P.F. Brust, P.J. Koutz, et al. (1985) Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*., Molecular and cellular biology. 5, 1111–1121.

75. R. Couderc and J. Baratti (1980) Oxidation of methanol by the yeast, *Pichia pastoris*: Purification and properties of the alcohol oxidase., Agricultural and biological chemistry. 44, 2279–2289.

76. F.W. Krainer, C. Dietzsch, T. Hajek, et al. (2012) Recombinant protein expression in *Pichia pastoris* strains with an engineered methanol utilization pathway., Microbial cell factories. 11, 22.

77. I.A. Pla, L.M. Damasceno, T. Vannelli, et al. (2006) Evaluation of Mut+ and MutS *Pichia pastoris* phenotypes for high level extracellular scFv expression under feedback control of the methanol concentration., Biotechnology progress. 22, 881–888.

78. S.-J. Kim (2009) Optimization of the Functional Expression of Coprinus cinereus Peroxidase in *Pichia pastoris* by Varying the Host and Promoter, Journal of Microbiology and Biotechnology. 19, 966–971.

79. J.J. Clare, F.B. Rayment, S.P. Ballantine, et al. (1991) High-Level Expression of Tetanus Toxin Fragment C in *Pichia pastoris* Strains Containing Multiple Tandem Integrations of the Gene, Nature biotechnology. 9, 455–460.

80. O. Cos, A. Serrano, J.L. Montesinos, et al. (2005) Combined effect of the methanol utilization (Mut) phenotype and gene dosage on recombinant protein production in *Pichia pastoris* fed-batch cultures., Journal of biotechnology. 116, 321–335.

81. R. Aw and K.M. Polizzi (2013) Can too many copies spoil the broth?, Microbial cell factories. 12, 128.

82. K. Sreekrishna (2010) Pichia, optimization of protein expression, Encyclopedia of Bioprocess Technology. 190, 695–701.

83. A.J. Sunga, I. Tolstorukov, and J.M. Cregg (2008) Posttransformational vector amplification in the yeast *Pichia pastoris*., FEMS yeast research. 8, 870–6.

84. H. Marx, A. Mecklenbräuker, B. Gasser, et al. (2009) Directed gene copy number amplification in *Pichia pastoris* by vector integration into the ribosomal DNA locus, FEMS Yeast Research. 9, 1260–1270.

85. J. Menéndez, B. García, and Y. Hidalgo (2000) New method for the selection of multicopy transformants of *Pichia pastoris*, using 3-amino-1,2,4 triazol., BioTechniques. 29, 1094–9.

86. A. Gatignol, H. Durand, and G. Tiraby (1988) Bleomycin resistance conferred by a drug-binding protein., FEBS letters. 230, 171–5.

87. H.J.M. van Blokland, T.H.J. Kwaks, R.G.A.B. Sewalt, et al. (2007) A novel, high stringency selection system allows screening of few clones for high protein expression., Journal of biotechnology. 128, 237–45.

88. K. Nordén, M. Agemark, J.Å.H. Danielson, et al. (2011) Increasing gene dosage greatly enhances recombinant expression of aquaporins in *Pichia pastoris*., BMC biotechnology. 11, 47.

89. M. Romanos, C. Scorer, K. Sreekrishna, et al. (1998) The generation of multicopy recombinant strains., Methods in molecular biology (Clifton, N.J.). 103, 55–72.

Table 2. Vectors for Intracellular Expression				
Vector Name	**Selection in *P. pastoris***	**Promoter**	**General Features**	**Reference**
pHIL-D2	*HIS4*	P_{AOX1}	*NotI* linearization site for *AOX1* replacement, *EcoRI* for cloning, *Sal*I or *Stu*I linearization for *HIS4* insertion	Life Technologies™
pAO815	*HIS4*	P_{AOX1}	*EcoRI* for cloning, *BglII* and *BamHI* sites for *in vitro* multimerization, *Bgl*II linearization for *AOX1* replacement, *Sal*I or *Stu*I linearization for *HIS4* insertion	Life Technologies™
pPIC3.5K	*HIS4*/G418	P_{AOX1}	MCS, *Sac*I linearization for *AOX1* insertion, *Sal*I linearization for *HIS4* insertion, G418 selection for multicopy strains	Life Technologies™
pPICZ (A, B, C)	Zeo^r	P_{AOX1}	different MCS (A, B and C), C-terminal 6XHis-tag, c-*myc* epitope, *BglII* and *BamHI* sites for *in vitro* multimerization, Zeocin™ selection for multicopy strains	Life Technologies™
pPIC6 (A, B, C)	Bsd^r	P_{AOX1}	Similar to pPICZ, except for Blasticidin selection for multicopy strains	Life Technologies™
pGAPZ (A, B, C)	Zeo^r	P_{GAP}	different MCS (A, B and C), C-terminal 6XHis-tag, c-*myc* epitope, Zeocin™ selection for multicopy strains	Life Technologies™
pFLD	Zeo^r	P_{FLD}	MCS, C-terminal 6XHis-tag, V5 epitope, targets integration into *FLD1* locus, induction with methanol or methylamine	Life Technologies™
PichiaPink™ (pPINK-HC, pPINK-LC)	*ADE2*	P_{AOX1}	Colour-based selection of strains, high-copy and low-copy plasmids, MCS, truncated promoter for marker gene, integration into *TRP2* or *AOX1* locus possible	Life Technologies™ ***(11)***
pJL-IX	*FLD1*	P_{AOX1}	*NotI* linearization site for *AOX1* replacement, *EcoRI* for cloning, formaldehyde selection for multicopy expression strains, transformed strain has to be *FLD1* deficient	***(57)***
pBLHIS-IX	*HIS4*	P_{AOX1}	Different combinations of MCS/auxotrophic selection markers available, different restrictions sites for *in vitro* multimerization, linearization site located in the marker gene	Keck Graduate Insitute ***(6, 7)***
pBLARG-IX	*ARG4*			
pBLADE-IX	*ADE1*			
pBLURA-IX	*URA3*			
pBLMET-IX	*MET2*			
pKAN B	$Tn903kan^r$	P_{AOX1}	MCS, resistance marker under control of P_{GAP} for direct selection of transformants using Kanamycin in *E. coli* and G418 in *P. pastoris*	***(55)***
pJAN/pJAZ/pJAG	*NAT1*/Zeo^r /G418	P_{AOX1}	Seamless cloning based on Type IIS restriction enzymes	Biogrammatics

pD902/pD905	Zeo^r	P_{AOX1}/P_{GAP}	Electra cloning system (seamless with Type IIS restriction enzyme *Sap*I), IP-Free©	DNA 2.0
pRSFC plasmid family (18 variants)	Zeo^r/*HIS4*	P_{AOX1}/P_{GAP}	Seamless cloning of a PCR product using Type IIS restriction enzymes, multiple combinations of N- or C-terminal tags (6xHIS, FLAG, MYC, Strep, MBP and eGFP)[a], blunt end ligation requires confirmation of GOI orientation	TU Graz, Vogl et al. 2014 (manuskript under revision)
pXYZ plasmid family	*HIS4*/*ARG4*/Zeo^r/ G418	P_{AOX1}/P_{GAP}	*Bgl*II/*Sph*I/*Swa*I linearization sites for *AOX1* gene replacement, *ARG4* promoter drives expression of marker gene, Unique restriction sites are provide to replace the marker CDS or complete marker cassette	TU Graz ***(63)***

[a] HIS-, MBP- and Strep-tag fusion plasmids are also available with a TEV-protease cleavage site.

Table 3. Vectors for Secretory Expression

Vector Name	Selection in *P. pastoris*	Promoter	General Features	Reference
pHIL-S1	*HIS4*	P_{AOX1}	*PHO1* secretion signal, MCS for in-frame fusion of the GOI, *Bgl*II linearization site for *AOX1* replacement, *Sal*I or *Stu*I linearization for *HIS4* insertion	Life Technologies™
pPIC9K	*HIS4*/G418	P_{AOX1}	α-mating factor secretion signal, MCS for in-frame fusion of the GOI, *Bgl*II linearization for *AOX1* replacement, *Sal*I linearization for *HIS4* insertion, G418 selection for multicopy strains	Life Technologies™
pPICZα (A, B, C)	Zeo^r	P_{AOX1}	α-mating factor secretion signal, different MCS (A, B and C), C-terminal 6XHis-tag, c-*myc* epitope, *BglII* and *BamHI* sites for *in vitro* multimerization, Zeocin™ selection for multicopy strains	Life Technologies™
pPIC6α (A, B, C)	*Bsd*^{r}	P_{AOX1}	Similar to pPICZα except for Blasticidin selection for multicopy strains	Life Technologies™
pGAPZα (A, B, C)	Zeo^r	P_{GAP}	α-mating factor secretion signal, different MCS (A, B and C), C-terminal 6XHis-tag, c-*myc* epitope, Zeocin™ selection for multicopy strains	Life Technologies™
pFLDα	Zeo^r	P_{FLD}	α-mating factor secretion signal, MCS, C-terminal 6XHis-tag, V5 epitope, targets integration into *FLD1* locus, induction with methanol or methylamine	Life Technologies™
PichiaPink™ (pPINKα-HC)	*ADE2*	P_{AOX1}	α-mating factor secretion signal, colour-based selection of strains, truncated promoter for marker gene, low-copy and high-copy plasmids, 7 other secretion signals are available and can be cloned via three-way ligation	Life Technologies™ ***(11)***

pJL1-IX	*FLD1*	P_{AOX1}	α-mating factor secretion signal, MCS, *NotI* linearization site for *AOX1* replacement, formaldehyde selection for multicopy expression strains, transformed strain has to be *FLD1* deficient	***(57)***
pBLHIS-SX	*HIS4*	P_{AOX1}	α-mating factor secretion signal, different combinations of MCS/auxotrophic selection markers available, different restrictions sites for *in vitro* multimerization, linearization site located in the marker gene	KGI ***(6, 7)***
pBLARG-SX	*ARG4*			
pBLADE-SX	*ADE1*			
pBLURA-SX	*URA3*			
pBLMET-SX	*MET2*			
pKANα B	*Tn903kan*^{r}	P_{AOX1}	α-mating factor secretion signal, MCS, resistance marker under control of P_{GAP} for direct selection of transformants using Kanamycin in *E. coli* and G418 in *P. pastoris*	***(55)***
pJAN-s1/pJAZ-s1/pJAG-s1	*NAT1*/Zeo^r/G418	P_{AOX1}	α-mating factor secretion signal, seamless cloning based on Type IIS restriction enzymes	Biogrammatics
pRSFC plasmid family (22 variants)	Zeo^r /*HIS4*	P_{AOX1}/P_{GAP}	Seamless cloning of a PCR product using Type IIS restriction enzymes, multiple combinations of N- or C-terminal tags (6xHIS, FLAG, MYC, Strep, MBP and eGFP)[b] and signal sequences, blunt end ligation requires confirmation of GOI orientation	TU Graz Vogl et al. 2014 (manuskript under revision)
pXYZ plasmid family	*HIS4*/*ARG4*/Zeo^r/ G418	P_{AOX1}/P_{GAP}	α-mating factor secretion signal, *Bgl*II/*Sph*I/*Swa*I linearization sites for *AOX1* gene replacement, *ARG4* promoter drives expression of marker gene	TuGraz
pD912/915 vector family	*Zeo*^{r}	P_{AOX1}/P_{GAP}	Electra cloning system (seamless with Type IIS restriction enzyme *Sap*I), 11 different secretion signals available[a], IP-Free©	DNA 2.0

[a] The α-MF secretion signal is provided once with Kex2p (KR) and Ste13p cleavage sites (EAEA), once lacking EA repeats, and once as truncated version (pre-region only).

[b] HIS-, MBP- and Strep-tag fusion plasmids are also available with a TEV-protease cleavage site.

Chapter 3

The effect of different variants of *S.cerevisiae* and *P.pastoris* alpha mating factor on secretion of Levanase and Horseradish peroxidase

Mudassar Ahmad[1], Ingund Rosales Rodriguez[1], Harald Pichler[1, 2] and Helmut Schwab[1, 2]

[1]*Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, A-8010 Graz, Austria,*

[2]*Austrian Centre of Industrial Biotechnology (ACIB), Petersgasse 14, A-8010 Graz, Austria*

Contribution:

MA designed the experiments, analyzed and interpreted the data. MA and IR discussed the results. MA performed the experiments and wrote the manuscript. HS and HP supervised the study. All Authors read and approved the final manuscript.

The effect of different variants of *S.cerevisiae* and *P.pastoris* alpha mating factor on secretion of Levanase and Horseradish peroxidase

Mudassar Ahmad[1], Ingund Rosales Rodriguez[1], Harald Pichler[1, 2] and Helmut Schwab[1, 2]

[1]*Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, A-8010 Graz, Austria,*

[2]*Austrian Centre of Industrial Biotechnology (ACIB), Petersgasse 14, A-8010 Graz, Austria*

Corresponding author:

Prof. Helmut Schwab

Institute of Molecular Biotechnology

Graz University of Technology

Petersgasse 14/5

8010 Graz

Austria

Tel.: +43 (0)316 873 – 4070

Fax: +43 (0)316 873 – 4071

E-mail: helmut.schwab@tugraz.at

Abstract:

Pichia pastoris, a methylotrophic yeast, has been used to express hundreds of recombinant proteins for industrial, pharmaceutical and research applications. Heterologous protein secretion in this yeast can be achieved by using either the native secretion signal of the protein of interest or by fusing it in frame with a foreign secretion signal. We have recently identified the putative *Pichia pastoris* alpha mating factor. In the present study, we have evaluated and compared the secretory potential of putative *P.p*astoris alpha mating factor secretion signal (*Pp*_αMF) to the *Saccharomyces cerevisiae* alpha mating factor secretion signal (*Sc*_αMF) using levanase and horseradish peroxidase (HRP) as reporter proteins. Furthermore, we studied the effect of the Glu-Ala spacer sequence on secretion in context of both secretion signals. *Pp*_αMF turned out to be inferior in directing both reporter proteins to the extracellular medium compared to *Sc*_αMF. Surprisingly, the Glu-Ala spacer sequence did not show any effect in combination with *Sc*_αMF. However, in case of *Pp*_αMF, the secretion of levanase increased with increasing number of Glu-Ala repeats and maximum secretion from this signal sequence was achieved with 5 Glu-Ala repeats. In order to determine the effect of *KEX2* deletion on HRP secretion mediated by different variants of secretion signals, expression landscapes of *P. pastoris* wild type and *Δkex2* strains were generated. Deletion of Kex2p endo-protease did not have any effect on HRP secretion directed by *Sc*_αMF. Secretion was reduced by approximately 50% in case of *Pp*_αMF, indicating that efficient cleavage of the pro-region is more dependent on Kex2p protease activity in the latter case. Additionally, we have designed a whole new series of expression vectors for intracellular and secretory expression of recombinant proteins based on CBS 7435 *P. pastoris* strain.

Keywords: Yeast, *P. pastoris* alpha mating factor, secretion signal, protein secretion, *S. cerevisiae* alpha mating factor, Glu-Ala spacer, Levanase, horseradish peroxidase.

Introduction

Pichia pastoris, a methylotrophic yeast, has been extensively used, over the years, to express hundreds of heterologous proteins for industrial, pharmaceutical and research applications. *Pichia pastoris* features extremely desirable characteristics such as ability to grow to high cell densities on simple media, ease of handling and genetic manipulations, availability of tightly regulated and inducible promoters, commercially available expression systems and ability to carry out complex posttranslational modifications of expressed proteins ***(1, 2)***. High expression levels of heterologous proteins –intracellular and secretory – can be achieved in *P. pastoris* by using commercially available expression vectors. *P. pastoris* only secretes few endogenous proteins into the culture medium, therefore, the majority of total protein present in the supernatant is of heterologous origin, resulting in simple and efficient downstream processing ***(3)***. The protein of interest can be targeted to the secretory pathway either by using its native secretion signal or by fusing it in frame with a heterologous signal sequence. Several heterologous secretion signals such as *S. cerevisiae* alpha mating factor signal (*Sc*_αMF) ***(4)***, acid phosphatase signal ***(5)***, invertase signal ***(6)***, *Phaseolus vulgaris* agglutinin signal ***(7)***, and killer toxin signal ***(8)*** have been used to direct protein secretion in *P. pastoris* with varying success ***(9, 10)***.

To date, *Sc*_αMF has been used most extensively and successfully to direct protein secretion in *P.pastoris* as compared to other secretion signals. *Sc*_αMF is composed of a pre-region (19 amino acids), which is cleaved by signal peptidase ***(11)*** in the endoplasmic reticulum and a pro-region (67 amino acids) ending in dibasic amino acids (KR), which are recognized and cleaved by the Kex2p endo-protease in trans-Golgi network ***(4, 12–14)***. In the native situation, this site is followed by a spacer peptide composed of Glu-Ala-Glu-Ala repeats. This spacer is cleaved by the dipeptidyl aminopeptidase Ste13p in trans-Golgi or in the secretory vesicles leaving from trans-Golgi to the

plasma membrane ***(13)***. The pro-region is glycosylated at three distinct N-glycosylation sites and elimination of these sites through mutagenesis does not abolish secretion but reduces the overall secretion efficiency ***(15, 16)***. The striking ability of this secretion signal is that it can guide proteins into the secretory pathway either co-translationally or post-translationally depending upon the protein of interest ***(17)***. This property of *S. cerevisiae* alpha mating factor has led to propose that pro-region may act as a chaperonin to keep the heterologous protein in unfolded state during post-translational translocation into the ER lumen ***(18)***. Furthermore, it is presumed that the pro-region plays a critical role in protein trafficking from the ER lumen to the Golgi compartment because removal of this region either eliminates protein secretion or reduces it to a great extent ***(19, 20)***.

One of the most common problems encountered while using this secretion signal is either imprecise or incomplete cleavage of the pro-region from the fusion protein, leaving a number of amino acids attached to the N-terminus of the heterologous protein. One possible cause could be that the structure of the fusion protein masks the Kex2p recognition site. The introduction of Glu-Ala spacer between the pro-region and the fusion protein extends the dibasic Kex2p processing site away from the N-terminus of the heterologous protein and, thus, facilitates proper cleavage ***(9)***. It has also been shown in previous studies that the *Sc*_αMF pro-region can be processed by multiple proteases of the yapsin family such as *YPS1/YAP3* and *YPS2/MKC7* ***(21–23)***. Nevertheless, there are a number of reports available, where *Sc*_αMF either failed to secrete the heterologous protein or secretion levels were minimal. One promising strategy to overcome this limitation is to look for secretion signals with higher secretory capacity as well as better processing efficiency than *Sc*_αMF ***(8, 24)***. Therefore, we wanted to test if the recently identified putative *Pp*_αMF could possess a higher processing efficiency resulting in better protein secretion from *P. pastoris*.

In the present study, we have constructed a new set of expression vectors for intracellular and secretory protein expression in *P. pastoris*. We have used these newly developed expression vectors to compare the secretory potential of *Sc*_αMF and *Pp*_αMF secretory leader sequences using *B. subtilis* levanase and horseradish peroxidase as reporter genes. Our studies showed that on the one hand, secretion from *Pp*_αMF is inferior for the tested proteins as compared to *Sc*_αMF On the other hand, we can conclude that Glu-Ala repeats are of higher importance for processing the *Pp*_αMF.

Material and Methods

Strains, Media and reagents:

For recombinant DNA work, *Escherichia coli* Top 10F´ (Life Technologies, Carlsbad, CA) was used. *P. pastoris* CBS7435 wild type (NRRL-Y11430, ATCC 76273), CBS7435 *his4* ***(25)*** and CBS7435 Δ*his4* Δ*kex2* (Ahmad et. al. unpublished results) strains were used for genetic transformations. Plasmid DNA isolation kit, Phusion polymerase and restriction enzymes were purchased from Thermo scientific (Bremen, Germany). T4 Wizard® SV Gel PCR Clean-Up System and DNA Ligase were purchased from Promega (Madison, WI). Zeocin^{TM} was purchased from InvivoGen (Eubio, Vienna, Austria). All other chemicals used in this study were purchased from Lactan (Graz, Austria). *E. coli* media components were purchased from AppliChem (VWR International GmbH, Vienna, Austria). *P. pastoris* media components were from BD Biosciences (Becton Dickinson GmbH, Vienna, Austria). *E. coli* cells were cultivated in Luria-Bertani medium (1% tryptone, 0.5% yeast extract, 0.5% NaCl and 2% agar) supplemented with 100 µg/ml of ampicillin for plasmid maintenance. *P. pastoris* was grown in BYPD (2% peptone, 1% yeast extract, 2% glucose, 200 mM potassium phosphate buffer, pH 7.0), ½ BYPD (1% peptone, 0.5% yeast extract, 1% glucose, 200 mM potassium phosphate buffer, pH 7.0), BMD (1.34% yeast

nitrogen base w/o amino acids; 4 x 10^{-5}% biotin, 2% dextrose, 200 mM potassium phosphate buffer, pH 7.0) and BMM (1.34% yeast nitrogen base w/o amino acids; 4 x 10^{-5}% biotin, 1% methanol, 200 mM potassium phosphate buffer, pH 7.0) For horseradish peroxidase expression studies, cells were grown to high cell densities in BMD and shifted to BMM induction media as essentially described by Weis et al. ***(26)*** and Vogl et al ***(27)***.

Vector constructions:

The primers used in this study are given in the supplementary information (Table S1). The origin and function of different elements used to construct *Pichia* pool expression vectors is given in supplementary information (Table S2). The sequence files are provided in supplementary genbank files. The origin of replication and ampicillin resistance gene were obtained from pUC8 vector backbone. alcohol oxidase 1 (*AOX1*) promoter, *AOX1* terminator, 3`region of *AOX1*, argininosuccinate lyase (*ARG4*) promoter, Arg4 terminator, Arg4 structural gene, and the His4 gene (phosphoribosyl-ATP pyrophosphatase; phosphoribosyl-AMP cyclohydrolase and histidinol dehydrogenase) were obtained from *P.pastoris* CBS 7435 wild type strain. Kanamycin and Zeocin^{TM} resistance genes were amplified from pPpKan_S and pPpT4 vectors respectively ***(25)***. *Saccharomyces cerevisiae* alpha mating factor pre-pro secretion signal (*Sc*_αMF) was obtained as a synthetic DNA. All parts were joined using either overlap extension PCR or by conventional cloning techniques. Levanase and horseradish genes were obtained from culture collection of Institute of Molecular biotechnology, Technical University of Graz, Austria.

P. pastoris **transformations:**

The condensed protocol ***(28)*** was used to prepare *P. pastoris* competent cells. Briefly, 2 µg of plasmid DNA restricted with *SwaI*/*BglII*/*SphI* (depending upon the plasmid used) was used to

transform freshly prepared competent cells using electroporation. Immediately after electroporation, 1 ml of premixed solution of 1 M sorbitol and BYPD (1:1) was added and cells were allowed to regenerate for 2 h at 28°C and 120 rpm. BMD agar plates without amino acids were used for the selection of His and Arg prototrophy. Transformants of the Zeocin^{TM} marker were selected on BYPD-Zeo plates supplemented with 25 µg/ml Zeocin^{TM}. For selection of *KanMX6* marker transformants, the concentration of G418 in the BYPD media was 300 mg/l.

Characterization of expression strains:

For screening of methanol utilization phenotype, 96 well deep well plates were filled with 250 µl of BMD media and inoculated with transformants. The cells were allowed to grow on standard cultivations condition (28^{0}C, 340 rpm and 80% humidity) for *P.pastoris* ***(26)***. After 48h of growth period, cells were stamped on BMM agar plates and again allowed to grow for 72h at 28^{0}C. The mut phenotype was scored by comparing the growth of transformants against CBS 7435 wild type and CBS 7435 Mut^S strains ***(25)***. Later on, colony PCR was performed to confirm screening results at genomic level. For levanase expression, transformants showing activities close to the mean values/construct were subjected to the real time PCR to determine gene copy number. The method used to determine gene copy number has been described by Abad et al ***(29)***.

Fermentation conditions for reporter proteins:

Fermentations for levanase as well as horseradish peroxidase expression were carried out in 96 well deep well plates at standard *P.pastoris* cultivation conditions i.e., 28^{0}C, 340 rpm and 80% humidity. For levanase fermentations, 5-10 µl of glycerol stocks were used to inoculate 250 µl of ½ BYPD and cultivated for 24h to generate pre-cultures. Ten microliter of these pre-cultures was used to inoculate main culture in 250 µl of ½ BYPD. After 24h of cultivations main cultures were

induced with 250 µl of BMM 1% (i.e., buffered minimal methanol 1%). Second methanol induction was carried out after 8h with 50 µl of BMM 10%. The subsequent methanol inductions were carried out by adding 50 µl of BMM 5% per well in the morning and 50 µl of BMM 10% in evening (approximately, after 8 hours of morning induction). Sample of 50 µl were taken after 24h, 48h and 72h for levanase activity. For horseradish peroxidase, fermentations were carried out according to the protocol described by Weis et al. The only exception was that we used BMD 2% instead of BMD 1% for initial growth ***(26, 27)***. Supernatant were separated from cells by centrifuging samples at 4000 rpm and 4^0C for 10 minutes in Eppendorf 5810R centrifuge. Supernatant from each well was assayed for levanase and HRP activity in UV-microplates (Greiner Bio-one GmbH, Kremsmunster, Austria). The cells were resuspended in 50 µl of ddH2O and OD600 was measured using plate reader (Spectramax 384plus, Molecular Devices, Sunnyvale, CA, USA).

Determination of Levanase activity in the supernatants:

Sucrose degrading activity of levanase can be measured by calculating concentration of glucose in enzyme reaction mixture. Glucose generated due to levanase activity was measured with "Glucose UV Kit" (Dipromed, Weigelsdorf, Austria), which is based on hexokinase method. Briefly, 20 µl of supernatant was mixed with 20 µl of 50 mg/ml of sucrose (Sigma) followed by incubation at 37^0C for 10 min. Levanase was deactivated by incubating the reaction mixture at 95^0C for 5 minutes in PCR machine. Ten microliter of the reaction mixture was mixed with 190 µl of Glucose UV reagent and sample were incubated at 37^0C for 10 minutes. The absorbance was measure at 340 nm with plate reader (Spectramax 384plus, Molecular Devices, Sunnyvale, CA, USA). Volumetric activities were calculated from standard curve generated for known concentration of glucose. Data given in figure 3B was obtained from 12 biological repeats of single

copy Mut^S transformants for each construct. One unit of levanase was defined as the amount of enzyme required to liberate one micromole of glucose per min ***(30)***.

Western Blots:

For western blot analysis of secreted and intracellular levanase fractions, 13 µl of supernatant and 20 µg of total cell protein was separated by loading on 12.5% SDS-Page gels respectively [Laemmli 1970]. Intracellular protein fractions were prepared by incubating the cell pellet in Y-PER TM Yeast Protein Extraction Reagent (Thermo Scientific, MA, USA) according to manufacturer's protocol. The protein concentration was measured by Lowry's method using bovine serum albumin as standard [LOWRY, ROSEBROUGH, FARR, and RANDALL 1951]. Subsequently, western blot analysis was carried out using standard protocols as described by Haid and Suissa, 1983 ***(31)***. Rabbit polyclonal antibody (1:5,000) raised against purified levanase expressed in *E.coli* was used as a primary antibody (Culture collection of Institute of Molecular Biotechnology, Technical University of Graz, Austria) [Wanker, Huber, and Schwab 1995]. Anti-rabbit IgG-alkaline phosphatase (1:25,000) produced in goat was used as a secondary antibody (Sigma-Aldrich, St. Louis, MO, USA). The colorimetric detections of immune-reactive protein bands was done using BCIP/NBT solution (BioRad, Hercules, CA, USA). Deglycosylation of proteins was carried out using EndoH (New England Biolabs) according to manufacturer's protocol.

Measurement of HRP activity in supernatants:

HRP activity was measured by using an ABTS assay as described by ***(26)***. The transformants were induced for 72h followed by centrifugation 4,000 rpm for 10 min at 4^0C. Fifteen microliter of the cultivation supernatant was mixed with 140 µl assay solution (0.5 mM ABTS in 50 mM NaOAc, pH 4.5, 2.9 mM H_2O_2) in a 96-well UV-microplates (Greiner Bio-one GmbH, Kremsmunster,

Austria). The increase in absorbance at 405 was measured in the plate reader at 405 nm (Spectramax 384plus, Molecular Devices, Sunnyvale, CA, USA) for 5 min at room temperature.

Results and Discussion

Construction of "*Pichia* pool" expression vectors:

In this study, we have constructed a set of 27 expression vectors (termed *Pichia* pool pXYZ vector family). For a detailed overview see Table 1 and a schematic representation is given in Figure 1. For sequence information see supplementary data. The vectors constructed during this study are *E.coli/P.pastoris* shuttle vectors i.e., cloning and amplification is carried out in *E.coli* followed by linearization of the expression cassette and transformation into *P. pastoris* to generate stable expression strains. These vectors have the benefit that no *E. coli* sequences are present in the integrated DNA. Two sets of expression vectors have been constructed namely, pXYZ for intracellular and pXaYZ for secretory expression. For intracellular expression of Gene of Interest (GOI) cloning can be performed using *EcoRI/NotI* restriction sites. The Kozak sequence for yeast (i.e., CGAACG) has to be restored for optimal translation initiation. The *EcoRI* site has been engineered by single point mutation in the P_{AOX1} (for details see Figure 1) without changing the promoter strength. Therefore, the gene of interest can be cloned without the interfering nucleotides between the promoter and start codon resulting in higher expression levels from this promoter ***(32, 33)***. Similarly, for secretory expression of GOI, we have fused the *S.cerevisiae* alpha mating factor secretion signal (*Sc*_αMF) with P_{AOX1} without any intervening nucleotides between the promoter sequence and start codon of the secretion signal. The gene of interest can be cloned by using *XhoI/NotI* sites, however, the Kex2p protease processing site (i.e., AAAAGA) has to be restored by including the nucleotide sequence of the processing site onto the primers (Figure 1) ***(3)***.

The vectors were designed to be integrated into the *AOX1* locus by homologous recombination. In case of successful targeting event the *AOX1* gene is deleted and the knockout strain has to rely on weaker expression of the *AOX2* gene to convert methanol into formaldehyde resulting in methanol utilization slow phenotype (Mut^S) ***(34)***. Therefore different restriction sites have been implemented allowing the use of different restriction enzymes to linearize the expression cassettes. However, due to lower homologous recombination frequencies in *P.pastoris* a higher portion of expression cassettes are integrated at random loci within the genome resulting in Mut^S and methanol utilization plus phenotype (Mut^+) ***(10, 25)***. The Mut phenotype can be easily be validated by growth of transformants on methanol (as a sole carbon source) plates. The targeting efficiency for *AOX1* locus with these vectors was measured to be ≥ 60%, which is significantly higher than previously reported targeting efficiencies of 5-25% ***(10, 35)***. We did not observe any effect of either sticky ends, blunt ends or expression cassette size on homologous recombination efficiencies.

The expression of the selection marker is uniformly controlled by the rather weak *ARG4* promoter and terminator sequence. The basal level of expression from this promoter is sufficient for selection of single copy transformants for all the selection markers used ***(36)***. One of the most commonly used strategy to increase heterologous protein expression is by increasing the copy number of the GOI. There are several methods available to achieve this end ***(37, 38)***. However, the most commonly used strategy relies on the selection of high copy transformants against increasing concentrations of Zoecin™ ***(39)***. The *Sh_ble* gene product binds Zeocin™ stoichiometrically rather than catalyzing its degradation ***(40)***. Thus increasing the expression of the *Sh_ble* gene product will also increase the resistance of transformants against Zeocin™ antibiotic. Most of commercially available vector systems use strong promoters i.e., P_{TEF1} and P_{ILV5}

to drive the expression of *Sh ble* gene product, which results in decreased probability of selecting higher copy transformants on increasing concentration of Zeocin™. Single copy transformants can exist on up to 2000 µg/ml of Zeocin™ antibiotic concentration ***(37, 41, 42)***. In order to make selection conditions more stringent, we have employed P_{ARG4} promoter, which is considerably weaker than P_{TEF1} or P_{ILV5} promoter. The basal level of expression from this promoter is sufficient to select single copy transformants on 25 µg/ml of Zeocin™ antibiotic. Due to the weaker promoter strength of the ARG4 promoter used in the vectors of this study only few colonies were able to survive on higher concentrations of Zeocin™, with a higher chance of being multicopy transformants, making the overall process of selection less laborious and efficient. We were able to select multicopy transformants having 20 ± 5 copies against a range of 400-500 µg/ml of Zeocin™ antibiotic (data not shown).

Comparison of *S. cerevisiae* and *P. pastoris* alpha mating factors secretion signals using Levanase as reporter:

S. cerevisiae alpha mating factor pre-pro secretion signal (*Sc*_αMF) is up to now the signal of choice when secretion of heterologous proteins from *P. pastoris* is desired. Recently, a putative homolog of *Sc*_αMF was identified from the newly sequenced genome of *P. pastoris CBS7435* ***(43)***. The nucleotide and protein sequences of *Sc*_αMF and *Pp*_αMF are given in Figure 2A & 2B respectively. *Pp*_αMF possesses a more complex structure as compared to *Sc*_αMF. The size of the pre-region of both secretion signals appears to be similar. However, *Pp*_αMF has a longer pro-region in addition to extended Glu-Ala repeats (marked red in figure 2A & 2B). Due to the successful use of *Sc*_αMF, we were interested whether the putative *Pp*_αMF might possess even higher secretory potential because of its homologous nature. In order to test this assumption, we chose to secrete *Bacillus subtilis* levanase from *P. pastoris* using different variants of *Sc*_αMF and

*Pp*_αMF (see figure 3B). These variants differ in the number of Glu-Ala repeats between the Kex2p cleavage site and the reporter proteins. We chose these reporter proteins because they are secreted efficiently from *P. pastoris* during small scale cultivation in deep well plates and activities can be measured in the supernatants using a simple enzymatic assay.

In total, seven expression constructs for levanase, under the control of P_{AOX1} promoter, were generated with varying number of Glu-Ala repeats (For details see Figure 2B). All expression constructs were linearized with *BglII* restriction enzyme to target the *AOX1* locus for gene replacement resulting in Mut^S phenotype which can easily be identified by growth on methanol as a sole carbon source ***(10, 27)***. It has been shown previously that gene dosage and locus of integration can affect overall expression and secretion levels of heterologous proteins in *P. pastoris* ***(29, 37, 42)***. In order to eliminate any background effects resulting from these factors, 80 transformants per construct were first screened for the Mut phenotype (data not shown). Transformants showing Mut^S phenotype and mean secretion levels for the respective construct were further analyzed by colony PCR, to confirm integration into *AOX1* locus. The copy number of integrated expression cassettes was determined by qPCR. Transformants having a single copy of the expression cassette integrated at the *AOX1* locus were compared to each other for levanase secretion levels.

The levanase activity was measured in the supernatants after 24 h, 48 h and 72 h of induction with methanol (Figure 3). The highest levanase activity of 2.62 ± 0.42 units per ml was achieved with *Sc*_αMF as a secretion signal irrespective of the number of Glu-Ala repeats. In case of *Pp*_αMF, highest activity of 1.96 ± 0.09 was achieved for secreted levanase measured in the supernatant after 72 h of methanol induction. Strikingly, Glu-Ala repeats did not show any significant effect on levanase secretion in case of *Sc*_αMF. However, in the case of *Pp*_αMF, increasing Glu-Ala

repeats had a direct positive effect on levanase secretion and highest levels of secretion were reached with the construct having five Glu-Ala repeats (Figure 3B). To further investigate the distribution of secreted and intracellularly retained levanase, supernatants and cell lysates (glycosylated and deglycosylated) were separated by SDS-PAGE. For accurate comparison, supernatants were taken from expression strains that showed similar activity to the mean values. For intracellular fractions, data has been normalized with total intracellular protein. Western Blot results showed that both secreted and intracellular fractions of levanase are processed by Kex2*p* endoprotease and were moderately glycosylated. We were not able to detect any unprocessed form (pre-Kex2p cleavage) of levanase in the cell lysate. The deglycosylated secreted and intracellular fractions were in close agreement for theoretically calculated molecular weight of 76 kDa of levanase ***(44)***. The western blot data were consistent with measured levanase activities i.e., strains with higher activity in the supernatant showed also higher protein levels of secreted levanase in the supernatant and lower protein levels in the cell lysate and vice versa.

It has been suggested previously that inclusion of Ste13p cleavage sites (Glu-Ala-Glu-Ala) between the Kex2p recognition site and the protein of interest can increase the overall Kex2p processing efficiency resulting in enhanced secretion levels ***(9, 27)***. The general understanding is that the secreted protein folds into a structure that masks the processing site, making it inaccessible to the processing enzymes. Therefore, inclusion of Glu-Ala repeats extends the processing site away from the folded protein making it more accessible to processing enzymes ***(9)***. Levanase activities and Western Blot results suggest that the pro-region of *Pp*_αMF folds into a structure that hinders Kex2p protease to cleave the pro-region from levanase, which results in increased retention time for levanase inside the secretory pathway and reduced secretion levels. However, this does not appear to be the case for *Sc*_αMF as increasing the Glu-Ala repeats does not affect

secretion levels ***(20)***. It has been reported previously that there are several other aspartic yapsin proteases such as Yap1 (previously known as Yap3) and yapsin 2 (previous known as Mkc7) whose substrate specificities overlap with Kex2p protease ***(21, 22, 45)***. Therefore, it could be possible that *Pp*_αMF is more dependent on efficient Kex2p protease activity than *Sc*_αMF

In order to further test this assumption and whether *Pp*_αMF can direct secretion of other proteins in addition to levanase, we chose horseradish peroxidase as a second reporter. We constructed three expression constructs namely, pAaHSwa2EAHRP, pAaHSwa5EAHRP and pAHSwaPp5EAHRP. The expression constructs were linearized with *SwaI* restriction enzyme and electroporated into CBS 7435 Δhis4 and CBS 7435 Δ*kex2* strains. Subsequently, the transformants were subjected to Mut phenotype selection on methanol plates and 24 Mut^S independent transformants per construct were cultivated in deep well plates and landscapes were generated to show average volumetric activities of secreted HRP (Figure 4A & 4B). The landscapes were generated to emphasize the average activities of constructs rather than focusing on highest activities as the latter usually represents only outliers ***(26)***. The data suggests that efficient cleavage of HRP from the pro-region of *Pp*_αMF is strongly dependent on the presence of a functional Kex2p protease. For the *Sc*_αMF, no significant difference was observed.

Conclusion

In the present study, we have used newly developed "*Pichia* pool" expression vectors to determine and compare the secretory potential of *Pp*_αMF and *Sc*_αMF leader sequence. Secretion efficiency was compared for *Bacillus subtilis* levanase and horseradish peroxidase as model proteins. Our data clearly shows that *Pp*_αMF is inferior to *Sc*_αMF leader sequence in targeting recombinant proteins to the exterior of cell, at least for the tested proteins. Additionally, it appears that cleavage

of *Pp*_αMF pro-region is more dependent on Kex2p protease activity than in the case of *Sc*_αMF. Further experimentation is needed to gain more insight into the role of Glu-Ala spacer sequences and the dependence of *Pp*_αMF on Kex2p protease activity.

Acknowledgement

MA, HS, IR, HP designed the study. MA carried out experimental work and wrote the manuscript. We would like to thank Melanie Hirz for critical reading of the manuscript. The authors gratefully acknowledge support from NAWI Graz. Stipend for MA was provided by Higher Education Commission of Pakistan.

References

1. J.L. Cereghino and J.M. Cregg (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*., FEMS microbiology reviews. 24, 45–66.

2. S. Macauley-Patrick, M.L. Fazenda, B. McNeil, et al. (2005) Heterologous protein production using the *Pichia pastoris* expression system., Yeast (Chichester, England). 22, 249–270.

3. J. Lin-Cereghino and G.P. Lin-Cereghino (2007) Vectors and strains for expression., Methods in molecular biology (Clifton, N.J.). 389, 11–26.

4. J. Kurjan and I. Herskowitz (1982) Structure of a yeast pheromone gene (*MFα*): A putative α-factor precursor contains four tandem copies of mature α-factor, Cell. 30, 933–943.

5. W.E. Payne, P.M. Gannon, and C. a Kaiser (1995) An inducible acid phosphatase from the yeast *Pichia pastoris*: characterization of the gene and its product., Gene. 163, 19–26.

6. E. Paifer, E. Margolles, J. Cremata, et al. (1994) Efficient expression and secretion of recombinant alpha amylase in *Pichia pastoris* using two different signal sequences., Yeast (Chichester, England). 10, 1415–1419.

7. R.J.M. Raemaekers, L. de Muro, J.A. Gatehouse, et al. (1999) Functional phytohemagglutinin (PHA) and *Galanthus nivalis* agglutinin (GNA) expressed in *Pichia pastoris* correct N-terminal processing and secretion of heterologous proteins expressed using the *PHA-E* signal peptide., European journal of biochemistry / FEBS. 265, 394–403.

8. A. Eiden-Plach, T. Zagorc, T. Heintel, et al. (2004) Viral preprotoxin signal sequence allows efficient secretion of green fluorescent protein by *Candida glabrata*, *Pichia pastoris*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*., Applied and environmental microbiology. 70, 961–966.

9. J.M. Cregg (2007) Introduction: distinctions between *Pichia pastoris* and other expression systems., Methods in molecular biology (Clifton, NJ). 389, 1–10.

10. R. Daly and M.T.W. Hearn (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production., Journal of molecular recognition : JMR. 18, 119–138.

11. M.G. Waters, E.A. Evans, and G. Blobel (1988) Prepro-alpha-factor has a cleavable signal sequence., The Journal of biological chemistry. 263, 6209–6214.

12. J. Brake, D.J. Julius, and J. Thorner (1983) A functional prepro-alpha-factor gene in *Saccharomyces* yeasts can contain three, four, or five repeats of the mature pheromone sequence., Molecular and cellular biology. 3, 1440–1450.

13. D. Julius, L. Blair, a Brake, et al. (1983) Yeast alpha factor is processed from a larger precursor polypeptide: the essential role of a membrane-bound dipeptidyl aminopeptidase., Cell. 32, 839–852.

14. D. Julius, A. Brake, L. Blair, et al. (1984) Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-alpha-factor., Cell. 37, 1075–1089.

15. J. Kurjan (1985) Alpha-factor structural gene mutations in *Saccharomyces cerevisiae*: effects on alpha-factor production and mating., Molecular and cellular biology. 5, 787–96.

16. S. Caplan, R. Green, J. Rocco, et al. (1991) Glycosylation and structure of the yeast MF alpha 1 alpha-factor precursor is important for efficient transport through the secretory pathway., Journal of bacteriology. 173, 627–635.

17. J. a Rothblatt, J.R. Webb, G. Ammerer, et al. (1987) Secretion in yeast: structural features influencing the post-translational translocation of prepro-alpha-factor in vitro., The EMBO journal. 6, 3455–3463.

18. E. Fabre, J.M. Nicaud, M.C. Lopez, et al. (1991) Role of the proregion in the production and secretion of the Yarrowia lipolytica alkaline extracellular protease, Journal of Biological Chemistry. 266, 3782–3790.

19. B. Chaudhuri, K. Steube, and C. Stephan (1992) The pro-region of the yeast prepro-alpha-factor is essential for membrane translocation of human insulin-like growth factor 1 in vivo., European journal of biochemistry / FEBS. 206, 793–800.

20. C. Oka, M. Tanaka, M. Muraki, et al. (1999) Human lysozyme secretion increased by alpha-factor pro-sequence in *Pichia pastoris*., Bioscience, biotechnology, and biochemistry. 63, 1977–83.

21. M. Egel-Mitani, H.P. Flygenring, and M.T. Hansen (1989) A novel aspartyl protease allowing *KEX2*-independent MF alpha propheromone processing in yeast., Yeast (Chichester, England). 6, 127–137.

22. H. Komano and R.S. Fuller (1995) Shared functions in vivo of a glycosyl-phosphatidylinositol-linked aspartyl protease, *Mkc7*, and the proprotein processing protease *Kex2* in yeast., Proceedings of the National Academy of Sciences of the United States of America. 92, 10752–10756.

23. V. Olsen, N.X. Cawley, J. Brandt, et al. (1999) Identification and characterization of *Saccharomyces cerevisiae* yapsin 3, a new member of the yapsin family of aspartic proteases encoded by the *YPS3* gene., Journal Biochemical Society. 339, 407–411.

24. K. Kottmeier, K. Ostermann, T. Bley, et al. (2011) Hydrophobin signal sequence mediates efficient secretion of recombinant proteins in *Pichia pastoris*., Applied microbiology and biotechnology. 91, 133–141.

25. L. Näätsaari, B. Mistlberger, C. Ruth, et al. (2012) Deletion of the *pichia pastoris* ku70 homologue facilitates platform strain generation for gene expression and synthetic biology, PLoS ONE. 7,.

26. R. Weis, R. Luiten, W. Skranc, et al. (2004) Reliable high-throughput screening with *Pichia pastoris* by limiting yeast cell death phenomena., FEMS yeast research. 5, 179–189.

27. T. Vogl, M. Ahmad, F.W. Krainer, et al. (2015) Restriction site free cloning (RSFC) plasmid family for seamless, sequence independent cloning in *Pichia pastoris*., Microbial cell factories. 14, 103.

28. J. Lin-Cereghino, W.W. Wong, S. Xiong, et al. (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*., BioTechniques. 38, 44,46,48.

29. S. Abad, K. Kitz, A. Hörmann, et al. (2010) Real-time PCR-based determination of gene copy numbers in *Pichia pastoris*., Biotechnology journal. 5, 413–420.

30. E. Wanker, A. Huber, and H. Schwab (1995) Purification and characterization of the *Bacillus subtilis* levanase produced in *Escherichia coli.*, Applied and environmental microbiology. 61, 1953–1958.

31. A. Haid and M. Suissa (1983) Immunochemical identification of membrane proteins after sodium dodecyl sulfate—polyacrylamide gel electrophoresis, Elsevier.

32. N.C. Crook, E.S. Freeman, and H.S. Alper (2011) Re-engineering multicloning sites for function and convenience., Nucleic acids research. 39, e92.

33. C. a Staley, A. Huang, M. Nattestad, et al. (2012) Analysis of the 5’ untranslated region (5'UTR) of the alcohol oxidase 1 (*AOX1*) gene in recombinant protein expression in *Pichia pastoris*., Gene. 496, 118–127.

34. J.M. Cregg, K.R. Madden, K.J. Barringer, et al. (1989) Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*., Molecular and cellular biology. 9, 1316–1323.

35. M. Romanos (1995) Advances in the use of *Pichia pastoris* for high-level gene expression, Current Opinion in biotechnology. 1–7.

36. M. Ahmad, M. Hirz, H. Pichler, et al. (2014), Protein expression in *Pichia pastoris*: Recent achievements and perspectives for heterologous protein production,

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4047484&tool=pmcentrez&rendertype=abstract.

37. R. Aw and K.M. Polizzi (2013) Can too many copies spoil the broth?, Microbial cell factories. 12, 128.

38. K. Sreekrishna (2010) *Pichia*, optimization of protein expression, Encyclopedia of Bioprocess Technology. 190, 695–701.

39. D.R. Higgins, K. Busser, J. Comiskey, et al. (1998) Small vectors for expression based on dominant drug resistance with direct multicopy selection., Methods in molecular biology (Clifton, N.J.). 103, 41–53.

40. A. Gatignol, H. Durand, and G. Tiraby (1988) Bleomycin resistance conferred by a drug-binding protein., FEBS letters. 230, 171–5.

41. H.J.M. van Blokland, T.H.J. Kwaks, R.G.A.B. Sewalt, et al. (2007) A novel, high stringency selection system allows screening of few clones for high protein expression., Journal of biotechnology. 128, 237–45.

42. K. Nordén, M. Agemark, J.Å.H. Danielson, et al. (2011) Increasing gene dosage greatly enhances recombinant expression of aquaporins in *Pichia pastoris*., BMC biotechnology. 11, 47.

43. A. Küberl, J. Schneider, G.G. Thallinger, et al. (2011) High-quality genome sequence of *Pichia pastoris* CBS7435., Journal of biotechnology. 154, 312–320.

44. E. Wanker, E. Klingsbichel, and H. Schwab (1995) Efficient secretion of *Bacillus subtilis* levanase by *Saccharomyces cerevisiae*., Gene. 161, 45–49.

45. V. Olsen, N.X. Cawley, J. Brandt, et al. (1999) Identification and characterization of *Saccharomyces cerevisiae* yapsin 3, a new member of the yapsin family of aspartic proteases encoded by the *YPS3* gene., The Biochemical journal. 339 (Pt 2, 407–411.

Tables

Tab. 1: '*Pichia* Pool' expression vectors family constructed during this study.

#	Name	Promoter	Mode of expression [a]	Restriction Site [b]	Selection marker [c]	Plasmid size (bp)
1	pAHBgl	*AOX1*	Intracellular	*BglII*	*P.pastoris HIS4* gene (wild type)	7111
2	pAHSph	*AOX1*	Intracellular	*SphI*	*P.pastoris HIS4* gene (wild type)	7111
3	pAHSwa	*AOX1*	Intracellular	*SwaI*	*P.pastoris HIS4* gene (wild type)	7115
4	pAABgl	*AOX1*	Intracellular	*BglII*	*P.pastoris ARG4* gene (wild type)	5977
5	pAASph	*AOX1*	Intracellular	*SphI*	*P.pastoris ARG4* gene (wild type)	5977
6	pAASwa	*AOX1*	Intracellular	*SwaI*	*P.pastoris ARG4* gene (wild type)	5981
7	pAKBgl	*AOX1*	Intracellular	*BglII*	Kanamycin^R/Geneticin^R	5389
8	pAKSph	*AOX1*	Intracellular	*SphI*	Kanamycin^R/Geneticin^R	5389
9	pAKSwa	*AOX1*	Intracellular	*SwaI*	Kanamycin^R/Geneticin^R	5392
10	pAZBgl	*AOX1*	Intracellular	*BglII*	*Sh ble*/Zeocin ™	4954
11	pAZSph	*AOX1*	Intracellular	*SphI*	*Sh ble*/Zeocin ™	4954
12	pAZSwa	*AOX1*	Intracellular	*SwaI*	*Sh ble*/Zeocin ™	4958
13	pAaHBgl	*AOX1*	Secretory	*BglII*	*P.pastoris HIS4* gene (wild type)	7359
14	pAaHSph	*AOX1*	Secretory	*SphI*	*P.pastoris HIS4* gene (wild type)	7359
15	pAaHSwa	*AOX1*	Secretory	*SwaI*	*P.pastoris HIS4* gene (wild type)	7363
16	pAaABgl	*AOX1*	Secretory	*BglII*	*P.pastoris ARG4* gene (wild type)	6225
17	pAaASph	*AOX1*	Secretory	*SphI*	*P.pastoris ARG4* gene (wild type)	6225
18	pAaASwa	*AOX1*	Secretory	*SwaI*	*P.pastoris ARG4* gene (wild type)	6229
19	pAaKBgl	*AOX1*	Secretory	*BglII*	Kanamycin^R/Geneticin^R	5635
20	pAaKSph	*AOX1*	Secretory	*SphI*	Kanamycin^R/Geneticin^R	5636
21	pAaKSwa	*AOX1*	Secretory	*SwaI*	Kanamycin^R/Geneticin^R	5640
22	pAaZBgl	*AOX1*	Secretory	*BglII*	*Sh ble*/Zeocin ™	5202
23	pAaZSph	*AOX1*	Secretory	*SphI*	*Sh ble*/Zeocin ™	5202
24	pAaZSwa	*AOX1*	Secretory	*SwaI*	*Sh ble*/Zeocin ™	5206
25	pGaHBgl	*GAP*	Secretory	*BglII*	*P.pastoris HIS4* gene (wild type)	7491
26	pGaHSwa	*GAP*	Secretory	*SwaI*	*P.pastoris HIS4* gene (wild type)	7495
27	pGaZSwa	*GAP*	Secretory	*SwaI*	*Sh ble*/Zeocin ™	5338

[a] = alpha mating factor pre-pro secretion signal containing *XhoI* site in front of the Kex2p processing site; synthetic DNA based on published sequence.
[b] = Designed for integration by gene replacement into the AOX1 locus.
[c] = Ampicillin resistance selection is applicable for *E. coli*.

Tab. 2: Expression constructs generate during this study.

#	Name	Mode of expression	Secretory leader	Selection marker [c]
1	pAHBgl_Levanase	Intracellular	--	*His4*
2	pAaHbgl_Levanase	Secretory	*Sc*_αMF *without EA repeats*	*His4*
3	pAaHbgl_2EALevanase	Secretory	*Sc*_αMF *with 2EA repeats*	*His4*
4	pAaHbgl_5EALevanase	Secretory	*Sc*_αMF *with 5EA repeats*	*His4*
5	pAHbgl_PpaLevanase	Secretory	*Pp*_αMF *without EA repeats*	*His4*
6	pAHbgl_PpaEALevanase	Secretory	*Pp*_αMF *with EA repeats*	*His4*
7	pAHbgl_Ppa5EALevanase	Secretory	*Pp*_αMF *with 5EA repeats*	*His4*
8	pAaHBgl_2EAHRP0	Secretory	*Sc*_αMF *with 2EA repeats*	*His4*
9	pAaHBgl_5EAHRP0	Secretory	*Sc*_αMF *with 5EA repeats*	*His4*
10	pAHBgl_Ppa5EAHRP0	Secretory	*Pp*_αMF *with 5EA repeats*	*His4*

[c] = Ampicillin resistance selection is applicable for *E. coli*.

Figures

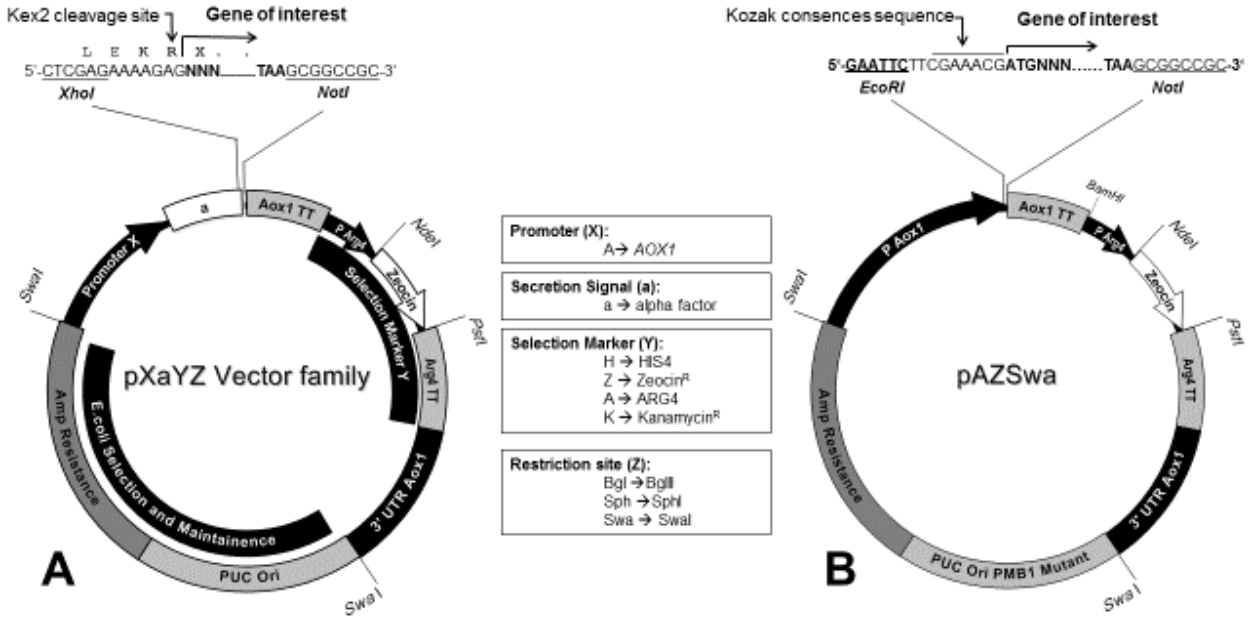


Figure 1: Schematic representation of '*Pichia* Pool' expression vectors (adopted from Ahmad, Hirz, Pichler, & Schwab, 2014). A, represents the general diagram of pXYZ vectors series for secretory and intracellular expression of gene of interest in *P. pastoris*. X represents promoter, a represents alpha mating factor secretion signal for secretory expression of gene of interest, Y represents selection marker for *Pichia* transformants and Z represents the restriction sites to linearize the plasmids to target the expression cassette for *AOX1* locus. Heterologous gene expression can be carried out by either methanol inducible promoter *AOX1* or constitutive promoter Gap. Selection marker expression in uniformly driven by Arg4 promoter and terminator. *Pichia* transformants can be selected for His4 or Arg4 prototrophy or against antibiotic resistance i.e., Zeocin^R or Kanamycin^R. Ampicillin resistance and PUC Origin of replication is provide for plasmid selection and maintenance in *E.coli* respectively. For intracellular expression of gene of interest (GOI) can be cloned using EcoRI-NotI. The Kozak consensus sequence i.e., CGAAACG

has to be restored for optimal translation initiation of GOI. For secretory expression of GOI can be cloned using XhoI-NotI sites. The Kex2p recognition site (*KR*) has to be restored for optimal protein processing through the secretory pathway. B, expression vector pAZSwa is shown as an example for intracellular expression vector. Unique restriction sites i.e., *BamHI*, *NdeI* and *PstI* are provided for effortless exchange of Arg4 promoter and selections markers with desired sequences. 3' UTR of *AOX1* is provided to facilitate the homologous recombination event at *AOX1* locus.

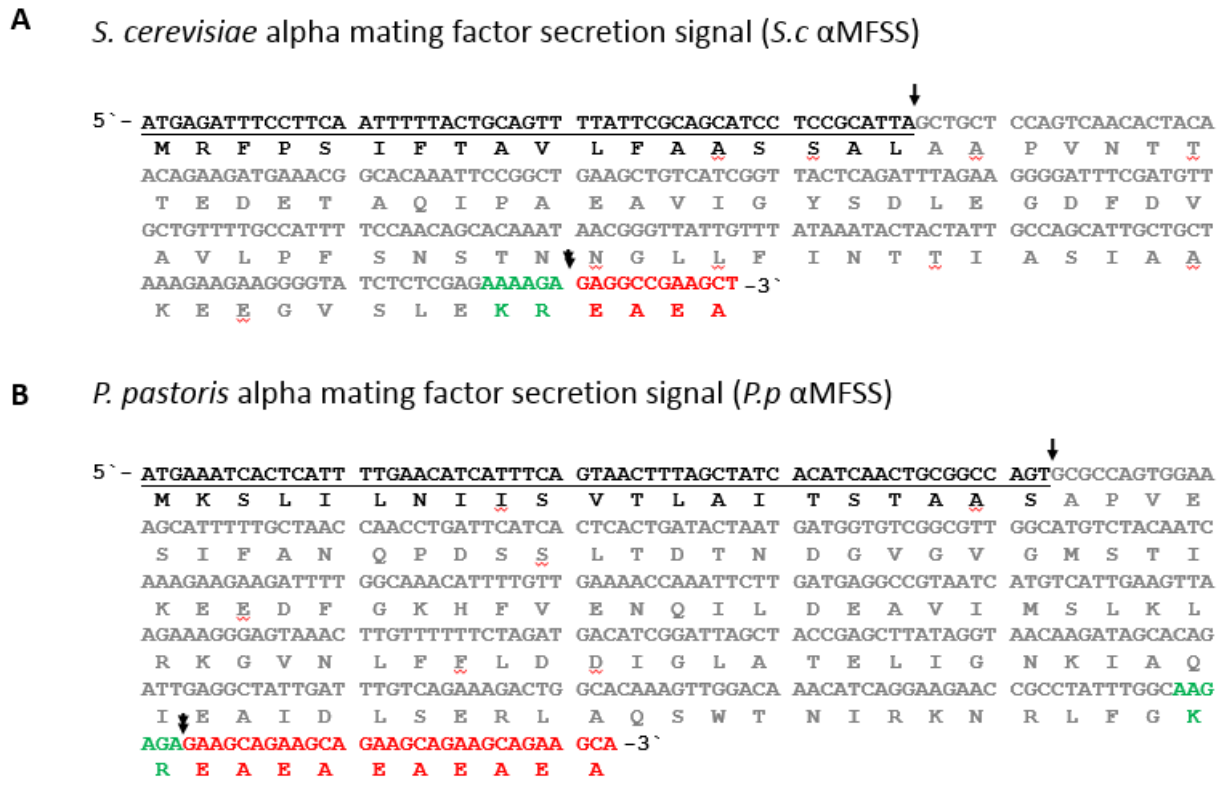


Figure 2: Comparison of *S.cerevisiae* and *P.pastoris* alpha mating factor secretion signal (Adopted from Küberl et al., 2011). Nucleotide sequence of *S.cerevisiae* (A) and *P.pastoris* alpha mating factor signal sequence with EA *repeats* (B). The deduced amino acid sequences are given in single letter code. Single arrows indicate the cleavage site for Pre regions (underlined and black). Double arrows indicate the Kex2p processing site whereby pro-region (grey) is removed from heterologous protein. Kex2p recognition sites are marked green and *Glu-Ala* repeats are marked red.

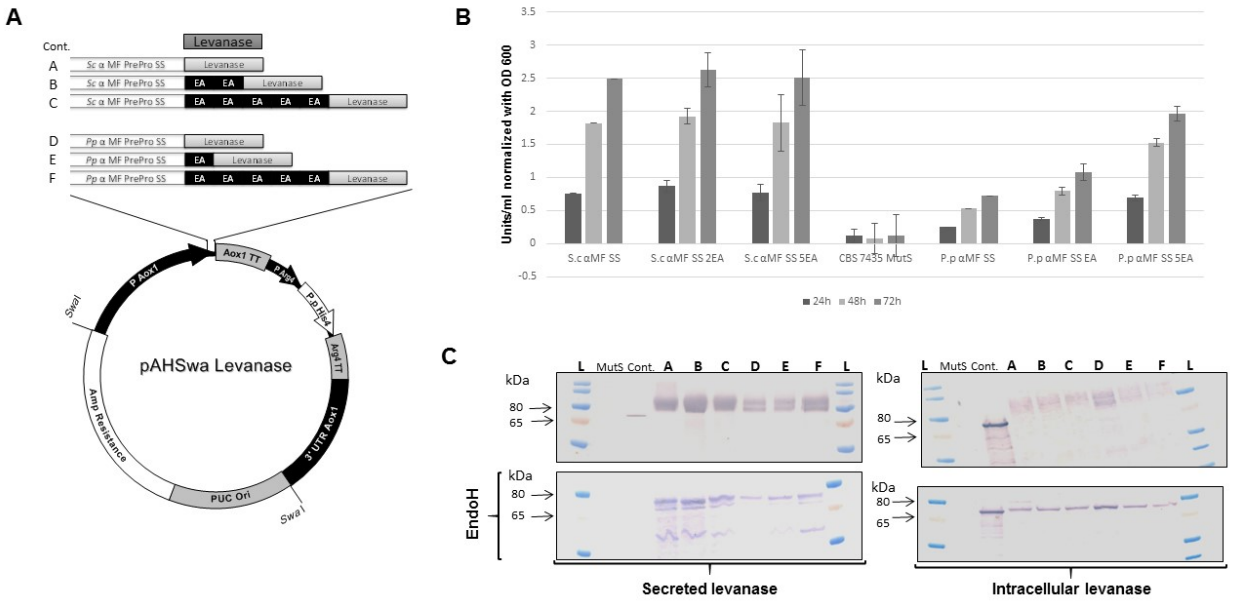


Figure 3: Figure 3: A: Schematic representation of different expression constructs generated during this study for secreted (A-F) and intracellular expression (control) of levanase. The vector elements are not drawn to scale. All cassettes were linearized with SwaI restriction enzyme and integrated into AOX1 locus (verified by PCR and methanol utilization slow phenotype mutS). B: Effect of different secretion signal variants on levanase secretion in P. pastoris. Levanase activity was measured in culture supernatants as described in material and methods. The mean values and standard deviation were calculated from 12 biological replicates and normalized by OD600. C: Western Blot detection of secreted and intracellularly expressed levanase (76 kDa) after 72 hours of methanol induction. Thirteen micro liter of supernatant or 17 microliter of cell lysate normalized with total cell protein (glycosylated and EndoH deglycosylated) were separated by SDS-PAGE, transferred to a nitrocellulose membrane and probed with rabbit polyclonal anti-bodies raised against purified levanase expressed in E. coli (Wanker et al., 1995). L, Page Ruler™ Plus Pre-stained Protein Ladder, MutS, CBS7435 methanol utilization slow strains (empty control strains),

cont., MutS strains expressing levanase intracellularly. Lane letters correspond to same strains as described in panel B.

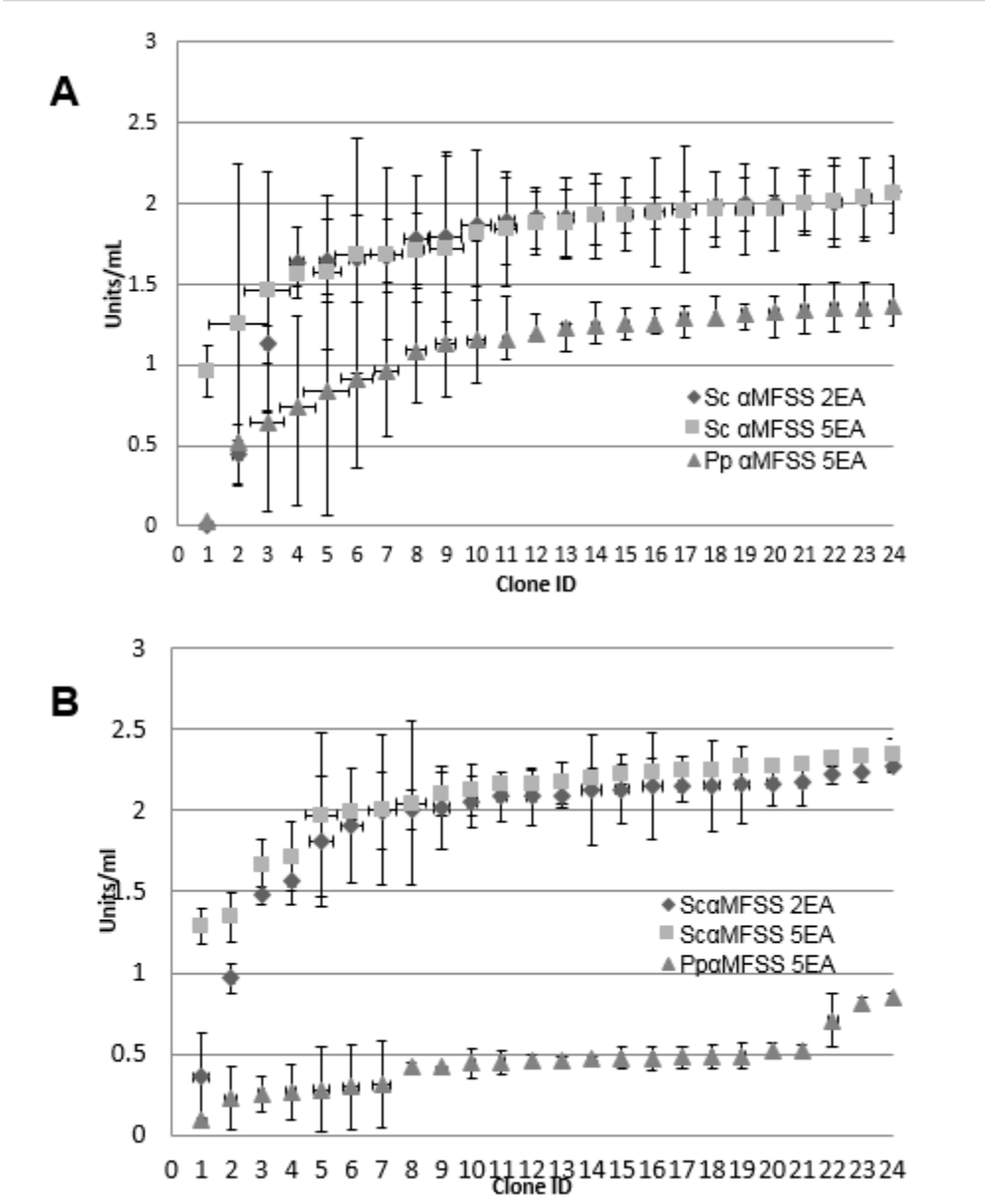


Figure 4: Horseradish peroxidase (Variants 0) secretion using different variants of *S.cerevisiae* and *P.pastoris* alpha mating factor secretion signal in deep well plates. The clones were pre-screened for methanol utilization slow phenotype. Mean values and standard deviation was calculated from three separate cultivations. A, expression in CBS 7435 Δ*aox1*. B, expression in CBS 7435 Δ*aox1* Δ*kex2* strain.

#	Primer name	Sequence(5'-3')
1	5'UTRAox1F	CTT TGA TGC CTG AAA TCC CAG CGC CTA CAA TGA TGA CA
2	3'UTRAox1R	CCG TTC GGT ATT AGA ATT TGT GAC TAA CAG TGT TCT TAC
3	Arg4TTF	GAT CTC CTG AGA CAA AGT TCA CGG GTA TCT AG
4	AlphaFSSR	GGC AAA ACA GCA ACA TCG AAA TCC CCT TC
5	PpAlphaF_Rev	CTTGAATCGG-CTCTCTTGCCAAATAGGCGG
6	EcoRI_LevFor	ACGAATTCTTCGAAACG-ATGAAAAAGGCCGATTCAAGCTAC
7	P2 Levrev	ACGCGGCCGCTTAAGACTCCTTCGTTACATTCTG
8	P1 Levfwd	CTCTCGAGAAAAGAGCCGATTCAAGCTACTATGATGAGG
9	α2EALevXhoIF	CTCTCGAGAAAAGAGAGGCCGAAGCTGCCGATTCAAGCTACTATGATGAGG
10	5EA_XhoIF	CTCTCGAGAAAAGA-GAAGCAGAAGCAGAAGCAGAAGC
11	PpAlphaF_EcoRIF	ACGAATTCGAAACG-ATGAAATCACTCATTTTGAACATC
12	PpAlphaF_Rev	CTTGAATCGG-CTCTCTTGCCAAATAGGCGG
13	PpAlphaFLevFor	GGCAAGAGA-GCCGATTCAAGCTACTATGAT
14	PpAlphaF2Rev	TCGGC-TGCTTCTCTCTTGCCAAATAGGCGG
15	PpAlphaF2LevFor	GGCAAGAGAGAAGCA-GCCGATTCAAGCTACTATGA
16	PpAlphaF3Rev	TGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTCTCTTGCC
17	PpAlphaF3LevFor	CAGAAGCAGAAGCAGAAGCAGCCGATTCAAGCTACTATGA
18	Seq1 fwd	CAGTCTCTCTATCGCTTCTGAAC
19	Seq1 rev	CCC AAT AAC TGG GCT GGT T
20	Seq2 fwd	CCTGATCAGCCTATCTCG
21	Seq2 rev	CCG AAG AAC GTT TTC CAA TG
22	Seq3 fwd	ACTAACTGACTGTCGTACGG
23	Seq3 rev	CGG TAT CAT TGC AGC ACT
24	Seq4 fwd	ATGCTTACCTTCTGGACC
25	Seq4 rev	GCT CTG CTA ATC CTG TTA CCA
26	Seq5 fwd	CACAGAATCAGGGGATAACG
27	Seq5 rev	CCT TTG AGT GAG CTG ATA CC
28	Seq6 fwd	AGAACAGTATTTGGTATCTGC
29	Seq6 rev	CGC ATC AGA CGA AGG ATG T
30	Seq7 fwd	AGCTAGAGTAAGTAGTTCG
31	Seq7 rev	CGC ACA ACC ATG CTA AGA TA
32	Seq8 fwd	GTTATTGTCTCATGAGCGG
33	HIS4seq1741for	CACAAGGTTGGTGCTAAGTG

34	HIS4seq1059for	TAGATGTGCCAAGTACGGTG
35	HIS4seq332for	CCGCTAATGCTAGTATCGCT
36	P1 Bgl2Exfwd	TCCTCCGATCGTTGTCAGAAGTAAGTTGG
37	P2 Bgl2Exrev	AATCAAAAGCTTGTCAATTGGAACCAGTCG
38	P3 Bgl2Mutfwd	GCCCTTTCGTCGCATGCAACATCCAAAGACGAAAG
39	P4 Bgl2Mutrev	TCTTTGGATGTTGCATGCGACGAAAGGGCC
40	P1 Sph1Exfwd	ATTTAACTGCAGTATACTGAGTTTGTTAATGATACAATAAACTG
41	P2 Sph1Exrev	TTCTGACAACGATCGGAGGACCGAAGG
42	P3 Sph1Mutfw	TTTATCTCAAGATCTTCACTGACTCGCTGCGCTC
43	P4 Sph1Mutrev	GCAGCGAGTCAGTGAAGATCTTGAGATAAATTTCACG
44	SeqLfwd1	GGGGAATATCACTTGTTCTATCAATACCATC
45	SeqLrev1	CGGTCACCGGCCGCAAGCACCATCAC
46	SeqLfwd2	CTGTATCCTGGTCTGATATTCCATCCACAG
47	SeqLrev2	CTTACTTTAAATTCTGCATTTATTTCATAG
48	SeqLfwd3	CTGGACGACTGTAAATGGCACGTGGGC
49	SeqLrev3	AGGTCATGCTTCGCATCCACATTGGC
50	kanforNde	CATCAT-ATGGGTAAGGAAAAGACTCACG
51	kanrevPst	CTGCTGCAG-TTAGAAAAACTCATCGAGCATC
52	zeoforNde	CATCAT-ATGGCTAAACTCACCTCTGC
53	zeorevPst	CTGCTGCAG-TTAGTCCTGCTCTTCTGCGACG
54	pEHNdeIRev	ATGTCATATGTAGCTGGTAATAAGTTTAG
55	pEHPstIFar	GATGAGATTTCCTTCAATTTTTACTGCTGTTTTATTCGCA
56	pEHPstIRev	CGGAGGATGCTGCGAATAAAACAGCAGTAAAAATTGAAGG
57	pEHSacIFar	ATTGGAGCTCGCTCATTCCAATTCC
58	pEVABglIIFar	ATTTTAAACGTGAAATTTATCTCAGCAGATCTCACTGACT
59	pEVABglIIRev	CCGAACGACCGAGCGCAGCGAGTCAGTGAGATCTGCTGAG
60	pEHSacIFar	ATTGGAGCTCGCTCATTCCAATTCC
61	pEHNdeIRev	ATGTCATATGTAGCTGGTAATAAGTTTAG
62	pEHPstIFar	GATGAGATTTCCTTCAATTTTTACTGCTGTTTTATTCGCA
63	pEHPstIRev	CGGAGGATGCTGCGAATAAAACAGCAGTAAAAATTGAAGG
64	HRP_5EAF	GCAGAAGCAGAAGCAGAAGC-CAACTTACTCCAACCTTCTAC
65	HRP_NotIR	AAGCGGCCGC-ATTATGAGTTAGAGTTGACAAC
66	Lev2EA_XhoIF	CACTCGAGAAAAGAGAGGCT-GAAGCTGCCGATTCAAGCTACTATG

Supplementary Information File 2

Table: Elements of *'Pichia Pool' expression* vectors and their function.

Elements	Origin	Function
P_Aox1	*Pichia pastoris* CBS7435	*P. pastoris AOX1* promoter for GOI expression
P_Gap	*Pichia pastoris*	*P.pastoris Gap* promoter for GOI Expression
Aox1_TT	*Pichia pastoris* CBS7435	Transcription terminator of *AOX1* gene in *P. pastoris* for GOI transcription termination
P_Arg4	*Pichia pastoris* CBS7435	*ARG4* promoter for expression of selection marker genes in *P. pastoris*
Arg4_TT	*Pichia pastoris* CBS7435	Transcription terminator for expression of selection marker genes
Sh_ble	Synthetic gene, amplified from pPpT4 [Näätsaari, Mistlberger, Ruth, Hajek, Hartner, and Glieder 2012a]	Confers resistance to antibiotic Zeocin
*Arg4*_CDS	*Pichia pastoris* CBS7435	*P. pastoris* wild type gene coding for argininosuccinate lyase; selection marker
KanMX6	*KanMX6* amplified from pPpT4_Kan [Näätsaari, Mistlberger, Ruth, Hajek, Hartner, and Glieder 2012b]	*KanMX6* gene; selection marker
*His4*_CDS	*Pichia pastoris* CBS7435	*P. pastoris HIS4* wild type gene; selection marker
pUC Ori	pUC8 (Genbank Acc. Nr. L08959)	pUC origin of replication for plasmid maintenance in *E. coli*
*bla*_cds	*β-lactamase* gene from pUC8 (Genbank Acc. Nr. L08959)	Ampicillin resistance in *E. coli;* selection marker

Chapter 4

Protease deficient and Biosynthetic gene knockout strains generated by a novel, user - friendly *P.pastoris* vector system for gene deletion

Mudassar Ahmad[1], Christine M. Winkler[1], Markus Kolmbauer[1], Harald Pichler[1, 2] and Helmut Schwab[1,2, §]

[1]*Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, A-8010 Graz, Austria*

[2]*Austrian Centre of Industrial Biotechnology (ACIB), Petersgasse 14, A-8010 Graz, Austria*

[§] Corresponding author

Contribution:

MA (85%), CW (15%) designed the experiments, analysed and interpreted the data. Majority of the wet lab work (approximately 65%) was carried out by MA. MK worked on the project as a master student under the supervision of MA and did approximately 35% of the total wet lab work. MA wrote the Materials and Methods, results and discussion and prepared figures and tables. CW wrote abstract, introduction and conclusion. CW and MA corrected the manuscript. HS and HP supervised the study. All Authors read and approved the final manuscript

Protease deficient and Biosynthetic gene knockout strains generated by a novel, user - friendly *P.pastoris* vector system for gene deletion

Mudassar Ahmad[1], Christine M. Winkler[1], Markus Kolmbauer[1], Harald Pichler[1, 2] and Helmut Schwab[1,2, §]

[1]*Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, A-8010 Graz, Austria*

[2]*Austrian Centre of Industrial Biotechnology (ACIB), Petersgasse 14, A-8010 Graz, Austria*

[§] Corresponding author

Email addresses:

MA: mudassar.ahmad@student.tugraz.at

CMW: christine.winkler@tugraz.at

MK: kolm_max@hotmail.com

HP: harald.pichler@tugraz.at

HS: helmut.schwab@tugraz.at

Abstract

Background:

Targeted gene knockouts play an important role in the study of gene function. For the generation of knockouts in the industrially important yeast *Pichia pastoris*, several protocols have been published to date. Nevertheless, creating a targeted knockout in *P. pastoris* still is a painful process, as the existing protocols are labour-intensive and/or prone to accumulate nucleotide mutations. In this study we aimed to make gene targeting in *P. pastoris* faster and more efficient.

Results:

We introduce a novel vector-based system for the generation of targeted knockouts in *P. pastoris*. The knockout vectors can easily be adapted to the gene of interest and strain background by efficient exchange of target homology regions and selection markers in single cloning steps. The respective selection marker can be recycled after the successful gene knockout. Excision of the marker is mediated by Flp recombinase and occurs at high frequency of ≥95%. We have validated our knockout system by deleting several genes involved in biosynthetic pathways and 26 protease genes. Surprisingly, deletion of the protease encoding gene *PEP4* and *KEX2* could only be achieved when using the *HIS4* gene as selection marker. On the contrary, knockout attempts employing a Zeocin™ resistance marker were not successful. To the best of our knowledge, we describe for the first time the knockout of *PRO3* and *PHA2* in *P. pastoris*. Knockout strains of *PHA2* did not display the anticipated auxotrophy for phenylalanine, but rather showed leaky growth on minimal medium. Additionally, we suggest a fast pooling method to identify rare homologous recombination events for multiple gene knockouts in parallel.

Conclusions:

The knockout vector system presented in this study was shown to be a versatile tool for gene targeting in *P. pastoris* with subsequent marker recycling. Gene targeting efficiencies depended on the targeted locus and the selection marker used. We propose that our method will accelerate the study of cellular and molecular processes in *P. pastoris*.

Keywords

Yeast, *Pichia pastoris*, gene targeting, gene knockout, knockout vectors, Flp/FRT recombinase, auxotrophic strains, protease deficient strains, marker recycling

Background

Gene targeting provides one of the most valuable molecular tools in the study of gene function and has been extensively used to elucidate cellular and molecular processes in yeasts. Genes are targeted by linear DNA cassettes that replace the targeted locus *in vivo* by homologous recombination. Sequence information about the target locus is a prerequisite for gene replacement by homologous recombination.

During the last years, the genome sequences of the important *P. pastoris* strains GS115, DSMZ 70382 and CBS7435, have become available [1–3]. These and related strains have been used successfully for the expression of over 400 proteins from various kingdoms of life [http://www.kgi.edu/faculty-and-research/profiles/james-m-cregg.html]. Nevertheless, to expand the range of heterologous products further, more profound knowledge about this yeast's metabolic and regulatory pathways would be beneficial. Techniques to selectively replace or disrupt genes can drive this process forward and enlarge the available molecular toolbox. Unfortunately, the targeting of genes in *P. pastoris* has proven to be problematic. The efficiency of gene replacement was reported to be extremely low, with homologous targeting sequences of <500 bp leading to <0.1% of positive targeting events. Using extended homology regions of >1 kb at each side, this rate could be increased to >50% [4].

Transformed DNA fragments can integrate into the genome via two distinct DNA repair mechanisms that play overlapping roles in yeasts: homologous recombination (HR) and non-homologous end joining (NHEJ). HR is mediated through proteins encoded by genes in the Rad52 epistasis group and is generally known to be an accurate repair mechanism, as it involves base-pairing of long stretches of matched base pairs [4]. On the other hand, NHEJ requires little to no

sequence homology to operate [5]. Free DNA ends are first bound by the heterodimer Ku70/80, which in turn recruits the catalytic subunit of DNA protein kinase (DNA-PKcs) [4, 6]. Accurate HR represents the dominant repair mechanism in the model yeast *S. cerevisiae*, and targeted knockouts can be achieved with short flanking homology regions of only 40 bp [7]. This property of *S. cerevisiae* allows construction of knockout cassettes by one-step PCR, which integrate with routinely 70% efficiency at the correct locus [8]. In order to foster HR in *P. pastoris*, a key player of NHEJ, Ku70p, was deleted by Näätsaari et al. [9]. They reported an efficiency of 97% when targeting the *HIS4* locus in a *ku70* strain with 250 bp of homologous sequence flanking the integration cassette on both sides. However, the growth rate of *ku70* strains compared to wild type is reduced by 11% [10].

Different strategies for the construction of *P. pastoris* gene targeting cassettes are described. Homologous flanking regions of ~1 kb are commonly used for the specific targeting of a locus. Combined with a selection marker, this requirement results in targeting cassettes of several thousand base pairs in length, which can be assembled either by cloning [10–14] or overlap-extension PCR (OE-PCR) [9, 15]. The published cloning methods usually require several sub-cloning steps and careful selection of appropriate restriction endonucleases. This process is complicated by low restriction efficiencies and incompatibility of some enzymes. The main disadvantage of fusing long DNA fragments by OE-PCR is the risk of accumulating nucleotide mutations during the amplification process. Secondly, this approach requires exceptionally long primers for sufficient overlaps.

Aside from targeting efficiency, an adequate set of selection markers represents an important factor in gene targeting. Various auxotrophic and antibiotic resistance markers have already been described in *P. pastoris* [10, 12, 16–20]. Nevertheless, the need for marker recycling is stressed by

extensive genetic engineering projects, such as the manipulation of the yeast's glycosylation pathway [22]. Nett and co-workers [10, 17] adapted the Ura-blaster system [23, 24] for *P. pastoris*. This protocol for marker recycling makes use of auxotrophy for uracil and resistance to 5-fluoroorotic acid (5-FOA) in *ura3* and *ura5* strains. Unfortunately, uracil auxotrophic strains suffer from severe growth retardation, even when grown in media supplemented with uracil [20]. Other methods for counter-selection make use of toxic genes. Examples are the *T-urf13* gene from the mitochondrial genome of male-sterile maize [25] and the *E. coli*-derived toxin gene *mazF* [26]. Expression of the toxins exerts strong selection pressure on the transformed cells, stimulating recombination and subsequent loss of the marker cassette. The significant selection pressure, however, causes cells to be less viable and might lead to conditional lethality for some gene deletions, as Nett and co-workers reported for *T-urf13* [10]. The stressful effects of toxins can be avoided by employing site-specific recombinase enzymes for marker recycling. These enzymes trigger the excision of sequences placed between two recombinase target sequences. Näätsaari and colleagues [9] placed Flp recombinase under control of the inducible *AOX1* promoter, and flanked the marker cassette with 34 bp FRT recombination sites. Methanol induction of the *AOX1* promoter resulted in excision of the marker cassette together with the Flp recombinase gene itself. A similar approach using the *Cre-loxP* system of phage P1 [27] was shown to be likewise applicable in *P. pastoris* [28].

In this study, we describe a simple and potent system to create knockout cassettes for gene targeting in *P. pastoris*. PCR-amplified homology sequences are integrated into a vector in a single cloning step. The cloning is made highly efficient by the specific properties of the employed *Sfi*I restriction endonuclease. The method introduced here allows effortless exchange of selection markers within the targeting vector, while obviating the need for amplification of long DNA fragments by PCR, a notoriously laborious and error-prone process. In combination with the Flp recombinase system

for marker recycling described above, our system is applicable for repeated gene deletions. We demonstrate the efficiency of our approach by reproducing already described gene deletions of *P. pastoris LYS2* [29], *MET2* [18], *TYR1* [12], *SUB2* [28], *PEP4, PRB1* [29] *PRC1* [30] *YPS1*, *YPS2*, *YPS7* [31], *KEX1* [32] and *KEX2* [33]. To our knowledge, we describe for the first time the targeting of putative proteases *PrtP*, *CTSE*, *KPX1-KPX9 (Knockout Protease X)* in addition to biosynthetic genes *PHA2* and *PRO3*, in the latter case creating a *P. pastoris* strain auxotrophic for proline. Owing to our fruitless attempts to create a phenylalanine-auxotrophic strain by targeting *PHA2*, the gene encoding the key enzyme for phenylalanine biosynthesis in *S. cerevisiae* [34], we argue that alternative biosynthesis routes for phenylalanine must exist in *P. pastoris*. Moreover, we introduce a fast and cleverly devised pooling method to identify rare homologous recombination events for multiple gene deletions in parallel. Taken together, our approach combines the advantages of previously known techniques, and, by adding innovative details, develops them further to make gene targeting in *P. pastoris* a fast and easy experience.

Results and discussion

Construction of knockout vector backbones

The strategy to recycle selection markers based on the Flp/FRT recombinase system was first described by Wirsching et al. [35] and later optimized by Näätsaari et al. [10] for use in *P. pastoris*. In both protocols, the knockout cassette was assembled and amplified by OE-PCR, a process prone to mutations. In the present study, we aimed at constructing knockout vectors that can be linearized at a unique internal restriction site to give the final knockout cassette containing the Flp/FRT marker recycling system. To achieve this goal, the Flp recombinase expression cassette, Zeocin™ resistance cassette and *E. coli* origin of replication were flanked by two 34 bp FRT repeats. We cloned a stuffer fragment, flanked by two *Sfi*I restriction sites (GGCCNNNN/NGGCC), in between of these FRT repeats to construct the knockout plasmid pPpKC1 (Figure 1A). The single-stranded overhangs generated by the *Sfi*I restriction enzyme were designed to be incompatible to each other to prevent religation of restricted backbone and to facilitate directional cloning of the insert. We hence termed these sites *Sfi*I *1* and *Sfi*I *2*. The special feature of *Sfi*I restriction endonuclease, a type IIF restriction enzyme, is that it interacts with two restriction sites simultaneously and cleaves them in a concerted manner, guaranteeing high restriction efficiencies [36]. We observed exceptionally high ligation efficiencies of more than 95% with *Sfi*I-cut vectors and inserts (data not shown). Furthermore, we exchanged the Zeocin™ marker cassette of pPpKC1 for the alternative *P. pastoris* markers *KanMX6*, *HIS4* and *ARG4* to expand the versatility of the system. The latter three yeast markers were combined with an ampicillin resistance marker for selection in *E. coli*. These modifications yielded the knockout vectors pPpKC2, pPpKC3 and pPpKC4, respectively (Figure 1B). We included the same *Sfi*I *1* and *Sfi*I *2* recognition sequences in all the constructed knockout vectors, thereby promoting effortless exchange of target homology regions between them.

The marker cassette of each plasmid was tested for functionality by transforming adequate *P. pastoris* strains and selecting transformants on respective media. Arg4 promoter/terminator sequences were used to drive expression of all marker genes in these knockout plasmids due to their short size and extensive testing during routine protein expression experiments carried out in our lab using "*Pichia* pool" plasmid [37]. During the course of these experiments we never observed reduced transformation efficiencies or malformed colonies compared to T4 plasmid, which uses a strong ILV5 promoter to drive marker gene expression [9]. The presence of homologous sequences i.e. Ar4 promoter/terminator did not result in reduced homologous recombination efficiencies when targeting Aox1 locus. On the contrary, we observed higher homologous recombination when targeting AOX1 locus using this marker cassette in our routine protein expression experiments. The gene replacement efficiencies for AOX1 locus were always more than 60 – 70 % compared to reported 5-25 % by using A. gossypii TEF promoter/terminator sequences [38, 39]. Furthermore, we examined selective concentrations of Zeocin^{TM} (25, 50 and 100 µg/ml) and Geneticin (200 and 300 µg/ml). A concentration of 25 µg/ml Zeocin^{TM} was found to be sufficient for identifying single copy transformants on YPD media. Higher concentrations of Zeocin^{TM} resulted in a reduced number of transformants and also led to undesired multicopy integrations. However, in BMD media a higher concentration of 100 µg/ml Zeocin^{TM} was needed to select positive transformants. When using the *KanMX6* selection marker, we found 300 µg/ml of Geneticin to be the optimum concentration to select positive transformants on YPD.

The clear advantage of our vector-based approach to construct knockout cassettes is the possibility to amplify the construct *in vivo* in *E. coli* prior to transformation. This strategy reduces the risk of nucleotide mutations that are likely to accumulate during extensive rounds of PCR amplification. Moreover, the selection marker of the cassette can be varied by a simple cloning procedure.

Tailoring of knockout vectors

We chose to verify the applicability of our knockout vector approach by targeting five biosynthetic genes (*MET2, LYS2, PRO3, TYR1* and *PHA2*) and twenty six protease genes. Table 1 provides detailed information on the disrupted genes. To target the knockout cassettes to these loci, we amplified approximately 1000 bp of the respective 5'- and 3'- regions from gDNA of wild-type CBS7435. During this PCR step the restriction sites *Sfi*I *1* and *Sfi*I *2* were added on primers. These restriction sites were later used for cloning of the target homology regions into the knockout vector backbones. The two amplified 5'- and 3'- homology fragments were joined by OE-PCR, thereby introducing a unique blunt end restriction site, e.g. *Sma*I, between the fragments that could later be used for linearization of the vector. We generated this unique *Swa*I restriction site by choosing the binding position of the outermost primers on the genome sequence, P1 and P4, in a way that they reconstitute the recognition sequence for the blunt end restriction enzyme after fusion in the OE-PCR (Figure 2A). Following restriction with *Sfi*I, the product of OE-PCR was cloned into the vector backbone pPpKC1 (Figure 2B).

Construction and characterization of auxotrophic knockout strains

Knockout plasmids based on pPpKC1 and harboring 5'- and 3'- homology regions to target *MET2, LYS2, PRO3, TYR1* and *PHA2* were linearized at the unique *SwaI* site. The resulting linear knockout cassettes were transformed into CBS7435 wild type cells to create strains auxotrophic for a single amino acid. Alternatively, the knockout cassettes were transformed into CBS7435 *his4* or *arg4* [10], to create double auxotrophic strains. In summary, we created nine single or double auxotrophic strains, namely *met2*, *met2arg4*, *met2his4*, *lys2*, *lys2arg4*, *lys2his4*, *pro3*, *tyr1* and *pha2*. An advantage of targeting these genes is the simple and reliable detection of the knockout based on the growth phenotype on minimal medium. Transformants of pPpKC1_*MET2*- and *LYS2*-

knockout cassettes were selected on YPD + Zeocin™. Whittaker and Whittaker [12] reported the inability of *P. pastoris tyr1* to grow on rich complex media, i.e. YPD. The same phenotype was observed for *S. cerevisiae pro3* by Brandriss [40]. Accordingly, we selected for *tyr1*, *pro3* and *pha2* transformants on BMD + Zeocin™. The efficiency of gene targeting was assessed by pinning the transformants on selective and non-selective media in parallel. A transformant was classified to be a successful knockout if it showed growth on BMD supplemented with the respective amino acid, but not on BMD alone. The calculated average gene targeting efficiencies ranged between 4 and 88% and are listed in Table 1. As the length of the homology regions was fairly similar for all targeted genes, the strong variation in targeting efficiency between the different loci must result from another unidentified factor.

Following phenotypic analysis, we confirmed that the observed amino acid auxotrophy was indeed caused by disruption of the targeted gene. We isolated gDNA of the transformants to verify integration of the knockout cassette by PCR (Figure 2C). Primer pairs P5+PAox1SeqR (PCR I) and PucSeqF+P6 (PCR II) were used to confirm the correct integration on the 5'- and 3'- side, respectively. To trigger marker recycling, cells were shifted to methanol as the sole carbon source, which induced expression of Flp recombinase from P_{AOX1}. Subsequently, Flp recombinase looped out the vector elements residing between the two FRT elements. One FRT element remained at the rearranged locus, flanked by the two *Sfi*I recognition sites. Marker recycling efficiencies for Flp-mediated recombination after 24 and 48 h of induction in buffered minimal methanol (BMM) media were determined by testing single colonies for their resistance to Zeocin™ and were found to be 50% and ≥95%, respectively. We further verified this rearrangement by performing control PCRs with primer pairs P5+P6 (PCR III) and P7+P8 (PCR IV), and by sequencing the products of PCR III. Representative results of PCR III are shown in Figure 3. All constructed strains and their genotypes after marker recycling are given in Table 2. Growth phenotypes of all biosynthetic gene

knockout strains on selective media are shown in Figure 4. As expected, only *pro3* and *tyr1* knockout strains did not grow on BYPD. All knockout strains grew on minimal medium supplemented with the respective amino acids. The growth phenotypes of *met2* and *lys2* knockout strains had already been described in earlier publications [18, 27]. We recorded growth curves for the *pro3* knockout strain on BMD and BYPD, both supplemented with proline (Figure 5A). The *pro3* knockout strain grew to high cell densities, but showed a longer lag phase than the wild type strain.

We were surprised to find that the *pha2* strain, which we expected to be deficient in phenylalanine biosynthesis, grew on minimal media lacking amino acids (Figure 4). From different kingdoms of life, two pathways for the synthesis of phenylalanine are known, starting either from arogenate or from phenylpyruvate. In *S. cerevisiae*, the only known route to phenylalanine starts from phenylpyruvate, which is produced from prephenate through the action of prephenate dehydratase [34]. We attempted to generate strains auxotrophic for phenylalanine by deleting *PHA2*, the gene encoding prephenate dehydratase. Unexpectedly, we observed a leaky and retarded growth phenotype of the *pha2* knockout strain on minimal medium (Figures 4 & 5B). Colonies turned pink after approximately 10 days on plate, which was not the case if supplemented with phenylalanine (Figure 6). These findings hint at the existence of more than one route for the biosynthesis of this aromatic amino acid in *P. pastoris*.

Construction and characterization of protease-deficient strains employing novel pooling method

In contrast to *S. cerevisiae*, homologous recombination events occur at rather low frequency in *P. pastoris* [5]. Consequently, if a gene is targeted by homologous recombination, only a small number of transformants carry the correct gene knockout. For genes that act in biosynthetic

pathways, the knockout results in a clear phenotype and, thus, can easily be identified. No clear knockout phenotype has been described for knockouts of the protease encoding genes. As a consequence, we expected to screen a large number of transformants until identifying a correct gene knockout. When several knockout experiments are carried out in parallel, the screening process can be significantly accelerated by using our novel pooling method for the isolation of gDNA described in Figure 7. Briefly, each locus was targeted in separate transformation reaction by a specific knockout cassette. Single colonies from selection plates were used to inoculate 96 DWPs. Transformants were allowed to grow for 24h and were pinned on non-selective plates to generate backup library. Cells from identical wells from different DWPs were pooled together e.g. cells from well A1 of different DWPS were pooled together etc. Thus, gDNA from mixed/pooled transformants can be isolated in single step. Isolated gDNA is used as template to check for correct integration of the knockout cassettes by PCR as sketched in Figure 2C. A PCR product is only produced in case of legitimate homologous recombination.

We identified targeting events of all the protease genes with the efficiencies listed in Table 1. As described for the auxotrophic knockout strains, we hence induced marker recycling and confirmed the successful excision by PCR and sequencing (data not shown).

Unexpectedly, we were not able to delete *PEP4 and KEX2* using a Zeocin^{TM} resistance marker on the knockout cassette. Extensive screening of ≥400 transformants identified six clones for *Pex4* with the cassette integrated at the target locus. However, all of the six transformants turned out to have the coding sequence of *PEP4* reintegrated at another position in the genome (data not shown). We also observed similar results for *KEX2* knockout. We assume that gene targeting was complicated by the important role of *PEP4* and *KEX2* as major proteases. *PEP4* was described to activate itself as well as other proteases, such as proteinase B (Prb1) and carboxypeptidase Y (Prc1)

[41] and *KEX2* is involved in processing of proportions in secretory pathway. Deletion of these proteases could therefore have a detrimental effect on cell viability. Additionally, the strong antibiotic Zeocin^{TM} might put too much pressure on the weakened cells. In order to omit any negative effect of Zeocin^{TM} in the selection process, we decided to change the marker in the knockout cassette to *HIS4*. Consequently, five out of 24 screened transformants showed correct integration of the knockout cassette for *PEP4* knockout. We were also able to construct *kex2* knockout strains with minimal effort with His4 as a selection marker. In order to further investigate negative effects of Zeocin^{TM} antibiotic, serial dilution of knockout strains were plated on 5 µg/ml of Zeocin^{TM} antibiotic. Both of the knockout strains showed increase sensitivity to Zeocin^{TM} antibiotic compared with wild type strains (data not shown). Contrary to our observations, Pan et al. [15] reported the knockout of *PEP4* and Werten & de Wolf [33] reported knockout of *KEX2* using P_{TEF1}/ Zeocin^{TM} as a selection marker. This promoter is significantly stronger than the P_{ARG4} promoter we used for marker expression [42]. Lower expression levels of the resistance gene may explain why we did not succeed in obtaining *pep4 and kex2* knockouts with Zeocin^{TM} as a marker. However, we were not able to transform pep4 and kex2 knockout strains using commercially available expression vectors, which utilize stronger promoters P_{TEF1}/P_{ILV5} [9, 43] to express Zeocin^{TM} resistant gene. Beside the single knockout strains for 26 proteases, we also generated double and triple knockout strains summarized in the table2.

Conclusions

We trust that the method and tools presented here will contribute to the investigation of gene function in *P. pastoris* by making the creation of gene knockout strains more efficient and effective. Our knockout vector system allows straightforward tailoring to the gene of interest and the *P. pastoris* strain used. The target homology regions can be easily exchanged in a single cloning step.

Likewise, the selection marker of the vector can be varied as required. On top of that, the marker can be efficiently recycled later on, thereby enabling repeated rounds of gene targeting.

In this study, we confirmed the feasibility of our knockout vector system by targeting five amino acid biosynthesis and 26 known and putative protease genes. The observed knockout efficiencies varied significantly (4 – 88%) between the targeted genes. Also, knockout success appeared to depend on the marker that was used to select transformants. The fact that we could only achieve deletion of *PEP4 and KEX2* when we used *HIS4* as a marker instead of Zeocin™ highlights the advantage of biosynthetic marker genes, especially when the gene knockout decreases viability.

With minor modifications, the presented vector system could be exploited for targeted integration of protein expression cassettes at a defined locus. Moreover, the possibility to recycle the selection marker allows a cascade of expression cassettes to be integrated into the genome. This quality makes our vector system a convenient tool in metabolic engineering projects.

Methods

Strains and media

Escherichia coli Top 10F´ (Life Technologies, Carlsbad, CA) was used for recombinant DNA work. *P. pastoris* CBS7435 wild type (NRRL-Y11430, ATCC 76273), CBS7435 *his4* and CBS7435 *arg4* [9] strains were used as hosts for genetic modifications. Phusion polymerase, DNA modifying enzymes, DNA ladder and plasmid DNA isolation kit were purchased from Thermo Scientific (Bremen, Germany). T4 DNA Ligase and Wizard® SV Gel PCR Clean-Up System were obtained from Promega (Madison, WI). L-Lysine–HCl, L-Phenylalanine and L-Proline were purchased from SERVA Electrophoresis (Heidelberg, Germany). L-Arginine-HCl, L-Histidine, L-Methionine, and L-Tyrosine were purchased from Carl ROTH GmbH (Karlsruhe, Germany). Zeocin^{TM} was from InvivoGen (Eubio, Vienna, Austria). All other chemical reagents used in this study were purchased from Lactan (Graz, Austria). *E. coli* media components were obtained from AppliChem (VWR International GmbH, Vienna, Austria). *E. coli* was cultivated in Luria-Bertani medium (1% tryptone, 0.5% yeast extract, 0.5% NaCl and 2% agar) supplemented with 100 µg/ml of ampicillin or 25 µg/ml Zeocin^{TM}. *P. pastoris* media components were from BD Biosciences (Becton Dickinson GmbH, Vienna, Austria). *P. pastoris* was grown in BYPD (2% peptone, 1% yeast extract, 2% glucose, 200 mM potassium phosphate buffer, pH 7.0) or BMD (1.34% yeast nitrogen base w/o amino acids; 4 x 10^{-5}% biotin, 2% dextrose, 200 mM potassium phosphate buffer, pH 7.0, supplemented with or without respective amino acids). Auxotrophic knockouts *tyr1*, *pro3* and *pha2* were grown on BMD media without or with respective amino acids as these knockout strains were, except for the latter, not able to grow in rich media [12, 40]. To recycle the selection marker, transformants were cultivated in BMM (1.34% yeast nitrogen base w/o amino acids; 4 x

10^{-5}% biotin, 0.5% methanol, 200 mM potassium phosphate buffer, pH 7.0, with or without amino acid supplementation).

Construction of knockout vector backbone

All primers used in the current study are given in the supplementary information (Additional File 1). The four basic knockout plasmids (pPpKC 1-4) harbouring different selection markers were constructed during this study (Figures 1A and 1B). The plasmid pPpT4 (JQ519689) [9] was used as initial backbone to construct the pPpKC1 knockout plasmid. The origin and function of different components used to construct these basic knockout plasmids are given in the supplementary information (Additional File 2). A synthetic DNA fragment, denoted as "stuffer", was amplified by PCR from plasmid pAaHBglHRP0 with primers PciIFRTSfiI1F/BglIIFRTSfiI2R (HPLC purified), digested with *Pci*I *and Bgl*II and cloned into the pPpT4 vector. Different components of the knockout plasmid pPpKC1 were amplified and joined by OE-PCR followed by classical restriction enzyme cloning using strategically placed restriction sites (*Pci*I, *Bgl*II, *and Nco*I). Equimolar ratio of different PCR products preferentially of similar size was used for OE-PCR. The vector backbone pPpKC1 was completely sequenced. Plasmids derived from pPpKC1 were only sequenced for exchanged parts. The marker cassette *KanMX6* (consisting of Argininosuccinate lyase (*ARG4*) promoter, *ARG4* terminator and synthetic *KanMX6* coding sequence) was amplified from pAKBgl expression plasmids (Ahmad et al., unpublished results), and *Pvu*II-*Avr*II-cloned into pPpKC1 to generate pPpKC2. The *HIS4* (phosphoribosyl-ATP pyrophosphatase; phosphoribosyl-AMP cyclohydrolase and histidinol dehydrogenase, X56180) and *ARG4* coding sequences were obtained by restricting pAHBgl and pAABgl expression plasmids with *Nde*I-*Pst*I and were cloned into pPpKC2 using the same restriction enzymes to generate pPpKC3 and pPpKC4, respectively. Details are provided in Additional File 3.

Construction of knockout cassettes

To construct the knockout cassettes, 5'- and 3'- homology regions were amplified from gDNA of wild type CBS7435 in two separate PCR reactions and joined in an OE-PCR (Figure 2A). The exact lengths of the amplified homology regions are given in Table 1. Two sets of primers (P1/P2 for 5'- homology and P3/P4 for 3'- homology) were used for amplification of homology regions for each target gene. Apart from a sequence complementary to the target locus, the primers were designed to have the following features: the primers P2 and P3 contained the "*Sfi*I 2" (5'-GGCCGATCAGGCC-3') and "*Sfi*I 1" (5'-GGCCACTAGGGCC-3') recognition sequences, respectively. The forward primer (P1) for 5'- homology and reverse primer (P4) for 3'-homology contained sequences complementary to each other (~ 20 nucleotides) for OE-PCR. Their binding sites on the genome sequence were selected in such a way that when the two fragments are combined they generate a unique blunt end restriction enzyme site, e.g. *Sma*I, for subsequent linearization of the knockout cassette (Figure 2A). In principle, any blunt end restriction enzyme site, which is not present in the backbone, can be generated for this purpose. The fragment resulting from OE-PCR was digested with *Sfi*I and ligated into the knockout vector backbone (Figure 2B). Ligation was confirmed by colony-PCR using primers PucSeqF and PAox1SeqR, in addition to restriction analysis with *Sfi*I. The insert was also sequenced by LGC Genomics (Berlin, Germany).

***Pichia* transformations**

P. pastoris competent cells were prepared using the condensed protocol [44]. Approximately 2 µg of linear DNA cassettes were transformed into competent cells using electroporation. Immediately after electroporation, 500 µl of 1 M sorbitol and 500 µl of YPD or BMD-AA (*pro3*, *tyr1* and *pha2* knockouts) were added and cells were allowed to regenerate for 2 h at 28°C and 120 rpm. Transformants of the Zeocin™ marker were selected on YPD plates supplemented with 25 µg/ml

Zeocin™ or BMD-AA plates supplemented with 100 µg/ml Zeocin™. For selection of *KanMX6* marker transformants, the concentration of G418 in the media was 300 mg/l. Amino acids were generally supplemented to a concentration of 150 mg/l, except for Histidine, which was added to 40 mg/l.

Characterization of knockout strains

For analysis of gene knockouts resulting in auxotrophies (*pha2, met2, lys2, pro3, tyr1*), single colonies of transformants were inoculated in 250 µl of BMD-AA in 96-well deep well plates (DWP) and grown for 24 h at 28°C and 320 rpm. The cultures were pinned onto BMD, BMD-AA and YPD plates to calculate the targeting efficiencies for each locus based on fast/slow growth (*pha2*) or growth/no growth phenotypes (*met2, lys2, pro3, tyr1*). For confirmation of the site-specific integration, two independent PCR reactions, namely PCR I and PCR II, were performed. As shown in Figure 2C, the outer primers P5 and P6 bind ~100 bp outside of the 5'- and 3'-homology regions selected for homologous recombination, whereas the inner primers PAox1SeqR and PucSeqF bind in *AOX1* promoter and *pUC* origin of replication, respectively. A PCR product is obtained only if integration has occurred at the right locus. In a first step, transformants were screened for the 5'- homology region (PCR I). Clones, which showed correct integration, were examined in a second PCR by using primers for the 3'- homology region (PCR II). Transformants showing correct integration on both side of the target locus were retrieved from the backup library; gDNA of the respective strain was isolated and reconfirmed by PCR reactions I, II, III and IV (Figure 2C). Gene knockouts lacking an easily identifiable phenotype (*known and putative proteases*) were confirmed by PCR only. Genomic DNA of multiple clones was isolated in one step using our pooling method to speed up the screening process (Figure 7). The concentration and

quality of isolated gDNA was verified using Nano-Drop (Thermo Scientific) and approximately 10-20 ng of gDNA were used per PCR reaction.

Isolation of genomic DNA

P. pastoris gDNA was isolated using a modification of the protocol by Hoffman and Winston [45]. All centrifugations were carried out using table top centrifuges at full speed. The cultures were grown in 96-well DWP in 600 µl of YPD or BMD-AA. The DWPs were incubated for 24-36 h at 28°C, 320 rpm and 80% humidity. Cultures were pipetted into Eppendorf tubes followed by centrifugation for 1 min. The supernatants were decanted and approximately 0.3 g of acid washed glass beads (Art-Nr. A553.1, Carl Roth GmbH), 150 µl yeast lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris, pH 8.0, 1 mM EDTA) and 150 µl of phenol:chloroform:isoamylalcohol (Art-Nr. A156, Carl Roth GmbH) were added to each tube. The tubes were vortexed for at least 8 min using Disruptor Genie SI-D238 (Scientific Industries, Inc.). One hundred and fifty microliter of TE buffer, pH 8.0, (Life Technologies) was added, followed by centrifugation for 5 min. The separated aqueous phase was transferred to a new Eppendorf tube and mixed with 1 ml of 100% ice cold ethanol followed by incubation at -20°C for 30 min to increase the overall yield of gDNA. After incubation, tubes were centrifuged for 1 min to pellet gDNA and the supernatant was discarded. The pellets were dried at 60°C and resuspended in 100-200 µl of sterile deionized water.

Marker Recycling

To start expression of Flp recombinase from P_{AOX1}, and thereby recycling of the selection marker, transformants were cultivated in 50 ml of BMM media at 28°C and 120 rpm. After 24 h and 48 h of induction, cultures were streaked on non-selective media to generate single colonies. Cells

arising from single colonies were cultivated in 96-well DWPs and screened for removal of the marker by pinning on selective and non-selective agar medium. The marker recycling efficiencies were calculated as percentage of the colonies that had lost the marker cassettes.

Growth rate studies

The growth rate of *P. pastoris* wild type and knockout strains *met2, lys2, pro3, tyr1* and *pha2* was analyzed by measuring the optical density (OD_{600}) in triplicate of cultures grown in 50 ml of BYPD or BMD media with our without supplementation of respective amino acids in 300 ml baffled flasks.

List of abbreviations

Buffered minimal dextrose supplemented with respective amino acids - BMD-AA

Buffered minimal methanol – BMM

Buffered YPD - BYPD

Deep well plate – DWP

Flippase - Flp

Flippase recombination target – FRT

Genomic DNA – gDNA

Homologous recombination – HR

Non-homologous end joining – NHEJ

Overlap-extension polymerase chain reaction – OE-PCR

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MA, CW, HP and HS conceived the study. MA and MK carried out the wet lab work and MA, CW, HP and HS wrote the manuscript. All authors have read and approved the manuscript.

Acknowledgements

We would like to acknowledge the Higher Education Commission of Pakistan (MA) and DK Molecular Enzymology (CW) for funding. We would also like to thank Manuela Killinger for excellent technical assistance during construction of the biosynthetic gene knockouts.

References

1. De Schutter K, Lin Y-C, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouzé P, Van de Peer Y, Callewaert N: **Genome sequence of the recombinant protein production host *Pichia pastoris*.** *Nat Biotechnol* 2009, **27**:561–566.

2. Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, Jaenicke S, Brinkrolf K, Goesmann A, Szczepanowski R, Pühler A, Schwab H, Glieder A, Pichler H: **High-quality genome sequence of *Pichia pastoris* CBS7435.** *J Biotechnol* 2011, **154**:312–320.

3. Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, Kleinheinz M, Sauer M, Altmann F, Gasser B: **Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*.** *Microb Cell Fact* 2009, **8**:29.

4. Pastwa E, Blasiak J: **Non-homologous DNA end joining**. *ACTA Biochim Pol Ed* 2003, **50**.

5. Krejci L, Altmannova V, Spirek M, Zhao X: **Homologous recombination and its regulation.** *Nucleic Acids Res* 2012, **40**:5795–5818.

6. Dudásová Z, Dudás A, Chovanec M, Dudášová Z, Dudáš A, Chovanec M: **Non-homologous end-joining factors of *Saccharomyces cerevisiae*.** *FEMS Microbiol Rev* 2004, **28**:581–601.

7. Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, Boeke JD: **Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications.** *Yeast* 1998, **14**:115–32.

8. Gueldener U, Heinisch J, Koehler GJ, Voss D, Hegemann JH: **A second set of loxP marker cassettes for *Cre*-mediated multiple gene knockouts in budding yeast.** *Nucleic Acids Res* 2002, **30**:e23.

9. Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A: **Deletion of the *Pichia pastoris ku70* homologue facilitates platform strain generation for gene expression and synthetic biology**. *PLoS One* 2012, **7**.

10. Nett JH, Gerngross TU: **Cloning and disruption of the *PpURA5* gene and construction of a set of integration vectors for the stable genetic modification of *Pichia pastoris*.** *Yeast* 2003, **20**:1279–1290.

11. Lin-Cereghino GP, Godfrey L, de la Cruz BJ, Johnson S, Khuongsathiene S, Tolstorukov I, Yan M, Lin-Cereghino J, Veenhuis M, Subramani S, Cregg JM: **Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in *Pichia pastoris*.** *Mol Cell Biol* 2006, **26**:883–97.

12. Whittaker MM, Whittaker JW: **Construction and characterization of *Pichia pastoris* strains for labeling aromatic amino acids in recombinant proteins.** *Protein Expr Purif* 2005, **41**:266–74.

13. Heiss S, Maurer M, Hahn R, Mattanovich D, Gasser B: **Identification and deletion of the major secreted protein of *Pichia pastoris*.** *Appl Microbiol Biotechnol* 2013, **97**:1241–9.

14. Wu M, Shen Q, Yang Y, Zhang S, Qu W, Chen J, Sun H, Chen S: **Disruption of *YPS1* and *PEP4* genes reduces proteolytic degradation of secreted HSA/PTH in *Pichia pastoris* GS115.** *J Ind Microbiol Biotechnol* 2013, **40**:589–599.

15. Pan R, Zhang J, Shen W-LL, Tao Z-QQ, Li S-PP, Yan X: **Sequential deletion of *Pichia pastoris* genes by a self-excisable cassette**. *FEMS Yeast Res* 2011, **11**:292–298.

16. Cereghino JL, Cregg JM: **Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*.** *FEMS Microbiol Rev* 2000, **24**:45–66.

17. Nett JH, Hodel N, Rausch S, Wildt S: **Cloning and disruption of the *Pichia pastoris ARG1, ARG2, ARG3, HIS1, HIS2, HIS5, HIS6* genes and their use as auxotrophic markers.** *Yeast* 2005, **22**:295–304.

18. Thor D, Xiong S, Orazem CC, Kwan A-CC, Cregg JM, Lin-Cereghino J, Lin-Cereghino GP: **Cloning and characterization of the *Pichia pastoris MET2* gene as a selectable marker.** *FEMS Yeast Res* 2005, **5**:935–942.

19. Lin-Cereghino J, Hashimoto MDM, Moy A, Castelo J, Orazem CC, Kuo P, Xiong S, Gandhi V, Hatae CT, Chan A, Lin-Cereghino GP, Lin-Cereghino J, Hashimoto MDM, Moy A: **Direct selection of *Pichia pastoris* expression strains using new G418 resistance vectors.** *Yeast* 2008, **25**:293–299.

20. Lin GP, Lin J, Jay A, Johnson MA, Lim M, Gleeson M a G, Cregg JM, Lin Cereghino GP, Lin Cereghino J, Jay Sunga A, Johnson MA, Lim M, Gleeson M a G, Cregg JM, Sunga AJ, Johnson MA, Lim M, Gleeson M a G, Cregg JM: **New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of *Pichia pastoris*.** *Gene* 2001, **263**:159–69.

21. Alani E, Cao L, Kleckner N: **A method for gene disruption that allows repeated use of *URA3* selection in the construction of multiply disrupted yeast strains.** *Genetics* 1987, **116**:541–5.

22. Lu P, Davis BP, Hendrick J, Jeffries TW: **Cloning and disruption of the beta-isopropylmalate dehydrogenase gene *(LEU2)* of *Pichia stipitis* with *URA3* and recovery of the double auxotroph.** *Appl Microbiol Biotechnol* 1998, **49**:141–6.

23. Soderholm J, Bevis BJ, Glick BS: **Vector for pop-in/pop-out gene replacement in *Pichia pastoris***. *Biotechniques* 2001, **31**:306–312.

24. Yang J, Jiang W, Yang S: ***MazF* as a counter-selectable marker for unmarked genetic modification of *Pichia pastoris***. *FEMS Yeast Res* 2009, **9**:600–609.

25. Güldener U, Heck S, Fiedler T, Beinhauer J, Hegemann JH, Fielder T, Beinhauer J, Hegemann JH: **A new efficient gene disruption cassette for repeated use in budding yeast.** *Nucleic Acids Res* 1996, **24**:2519–24.

26. Marx H, Mattanovich D, Sauer M: **Overexpression of the riboflavin biosynthetic pathway in *Pichia pastoris*.** *Microb Cell Fact* 2008, **7**:23.

27. Austin RJ, Kuestner RE, Chang DK, Madden KR, Martin DB: **SILAC compatible strain of *Pichia pastoris* for expression of isotopically labeled protein standards and quantitative proteomics**. *J Proteome Res* 2011, **10**:5251–5259.

28. Salamin K, Sriranganadane D, Léchenne B, Jousson O, Monod M: **Secretion of an endogenous subtilisin by *Pichia pastoris* strains GS115 and KM71**. *Appl Environ Microbiol* 2010, **76**:4269–4276.

29. Gleeson M a, White CE, Meininger DP, Komives E a: **Generation of protease-deficient strains and their use in heterologous protein expression.** *Methods Mol Biol* 1998, **103**:81–94.

30. Ohi H, Ohtani W, Okazaki N, Furuhata N, Ohmura T: **Cloning and characterization of the *Pichia pastoris PRC1* gene encoding carboxypeptidase Y**. *Yeast* 1996, **12**:31–40.

31. Guan B, Lei J, Su S, Chen F, Duan Z, Chen Y, Gong X, Li H, Jin J: **Absence of Yps7p, a putative glycosylphosphatidylinositol-linked aspartyl protease in *Pichia pastoris*, results in aberrant cell wall composition and increased osmotic stress resistance.** *FEMS Yeast Res* 2012, **12**:969–979.

32. Boehm T, Pirie-Shepherd S, Trinh LB, Shiloach J, Folkman J: **Disruption of the *KEX1* gene in *Pichia pastoris* allows expression of full-length murine and human endostatin.** *Yeast* 1999, **15**:563–572.

33. Werten MWT, De Wolf F a.: **Reduced proteolysis of secreted gelatin and *Yps1*-mediated alpha-factor leader processing in a *Pichia pastoris kex2* disruptant.** *Appl Environ Microbiol* 2005, **71**:2310–2317.

34. Braus GH: **Aromatic amino acid biosynthesis in the yeast *Saccharomyces cerevisiae*: a model system for the regulation of a eukaryotic biosynthetic pathway.** *Microbiol Rev* 1991, **55**:349–70.

35. Wirsching S, Michel S, Morschhäuser J: **Targeted gene disruption in *Candida albicans* wild-type strains: the role of the *MDR1* gene in fluconazole resistance of clinical *Candida albicans* isolates.** *Mol Microbiol* 2000, **36**:856–65.

36. Wentzell LM, Nobbs TJ, Halford SE: **The *SfiI* restriction endonuclease makes a four-strand DNA break at two copies of its recognition sequence.** *J Mol Biol* 1995, **248**:581–595.

37. Ahmad M, Hirz M, Pichler H, Schwab H: **Protein expression in *Pichia pastoris*: Recent achievements and perspectives for heterologous protein production**. *Applied Microbiology and Biotechnology* 2014:5301–5317.

38. Romanos M: **Advances in the use of *Pichia pastoris* for high-level gene expression**. *Curr Opin Biotechnol* 1995:1–7.

39. Daly R, Hearn MTW: **Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production.** *J Mol Recognit* 2005, **18**:119–138.

40. Brandriss MC: **Isolation and preliminary characterization of *Saccharomyces cerevisiae* proline auxotrophs.** *J Bacteriol* 1979, **138**:816–22.

41. Rothman JH, Hunter CP, Valls L a, Stevens TH: **Overproduction-induced mislocalization of a yeast vacuolar protein allows isolation of its structural gene.** *Proc Natl Acad Sci U S A* 1986, **83**:3248–52.

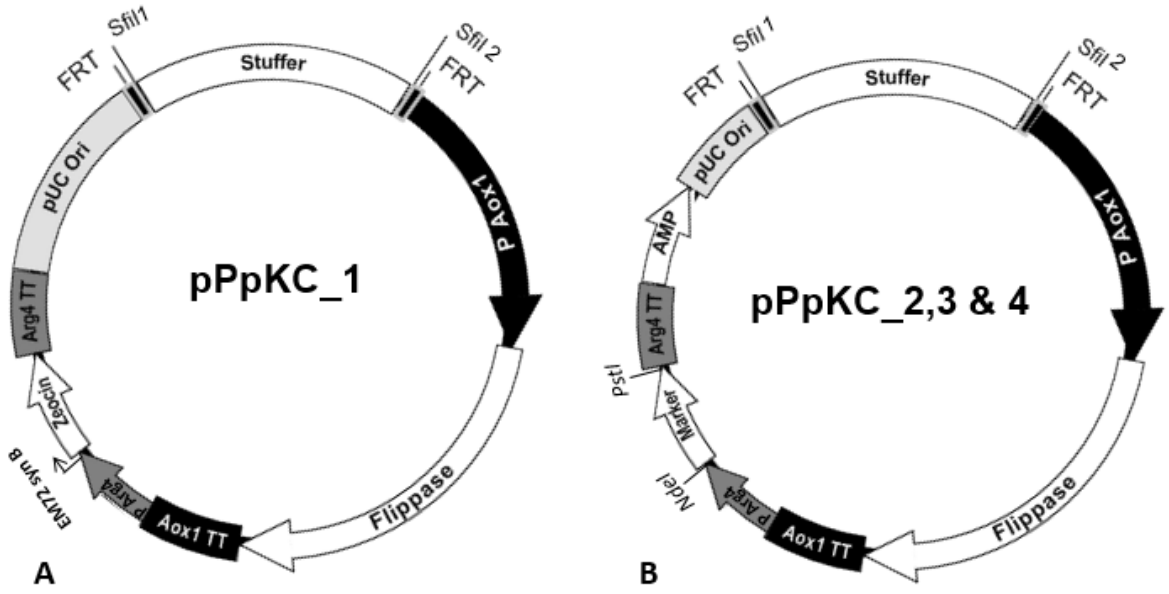
42. Schirmaier F, Philippsen P: **Identification of two genes coding for the translation elongation factor *EF-1* alpha of *S. cerevisiae*.** *EMBO J* 1984, **3**:3311–5.

43. Lin-Cereghino J, Lin-Cereghino GP: **Vectors and strains for expression.** *Methods Mol Biol* 2007, **389**:11–26.

44. 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.

45. Hoffman CS, Winston F: **A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*.** *Gene* 1987, **57**:267–72.

Figures



Name	Size (bp)	Selection marker *P. pastoris*	Selection marker *E. coli*	Accession No.
pPpKC1	5229	Zeo^R	Zeo^R	
pPpKC2	6809	G418^R	Amp^R	
pPpKC3	8532	*HIS4*	Amp^R	
pPpKC4	7398	*ARG4*	Amp^R	

Figure 1: The basic knockout plasmids harboring different *P. pastoris* selection markers constructed during this study. (A) pPpKC1. (B) pPpKC2, 3 and 4. Indicated are the unique restriction sites *Nde*I and *Pst*I to exchange the marker

.

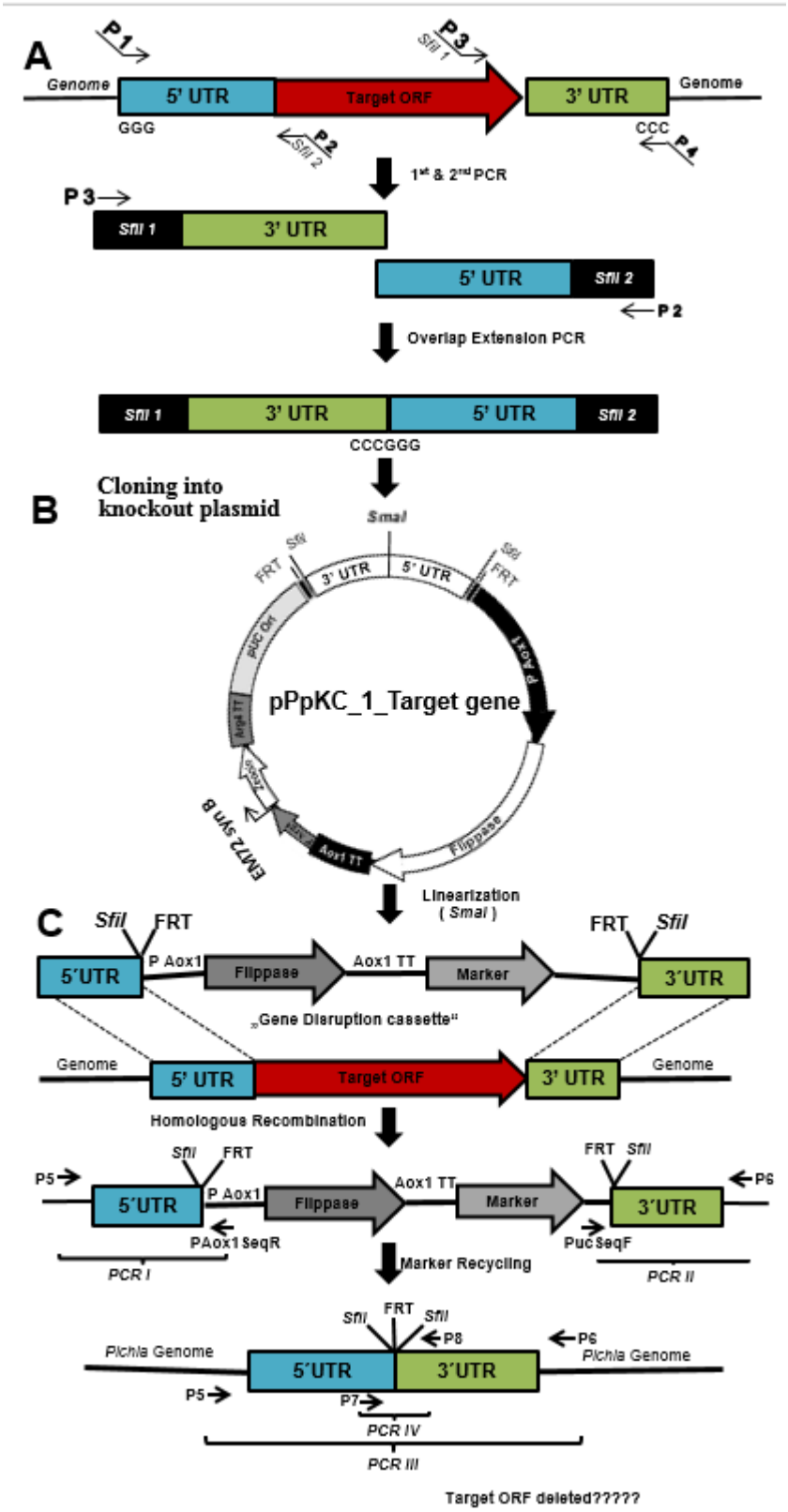

A
P1
P3
SfiI 1
Genome
5' UTR
Target ORF
3' UTR
Genome
GGG
SfiI 2
P2
CCC
P4
1st & 2nd PCR
P3
SfiI 1
3' UTR
5' UTR
SfiI 2
P2
Overlap Extension PCR
SfiI 1
3' UTR
5' UTR
SfiI 2
CCCGGG
B
Cloning into knockout plasmid
SmaI
FRT
SfiI
3' UTR
5' UTR
SfiI
FRT
pPpKC_1_Target gene
Linearization (SmaI)
C
SfiI
FRT
5'UTR
P Aox1
Flippase
Aox1 TT
Marker
FRT
SfiI
3'UTR
„Gene Disruption cassette"
Genome
5' UTR
Target ORF
3' UTR
Genome
Homologous Recombination
P5
SfiI
FRT
5'UTR
P Aox1
Flippase
Aox1 TT
Marker
FRT
SfiI
3'UTR
P6
PAox1 SeqR
PucSeqF
PCR I
PCR II
Marker Recycling
SfiI
FRT
SfiI
P8
P6
PIchIa Genome
5'UTR
3'UTR
PIchIa Genome
P5
P7
PCR IV
PCR III
Target ORF deleted?????

Figure 2: Schematic representation of the experimental procedure for gene deletion and its confirmation. (A) By performing two PCR reactions, the 3'- and 5'- homology regions of the respective target gene were amplified separately. The two PCR products were joined by OE-PCR, creating a unique restriction site (*Sma*I) for subsequent linearization. (B) The *Sfi*I-restricted 3'- and 5'- homology regions were cloned into the knockout vector. The final knockout vector was linearized using *Sma*I prior to transformation into *P. pastoris*. (C) Homologous recombination replaced the target ORF with the linear knockout cassette. The correct integration was verified by amplifying region I) containing the 5'- homology (primer pair P5/PAox1SeqR) and region II) containing the 3'- homology (primer pair PucSeqF/P6). Clones with positive results for both PCRs were selected for marker recycling. The removal of the integrated marker cassette was verified by amplification of region III) and region IV). P1 – P8: Primers; FRT: Flippase recombination target.

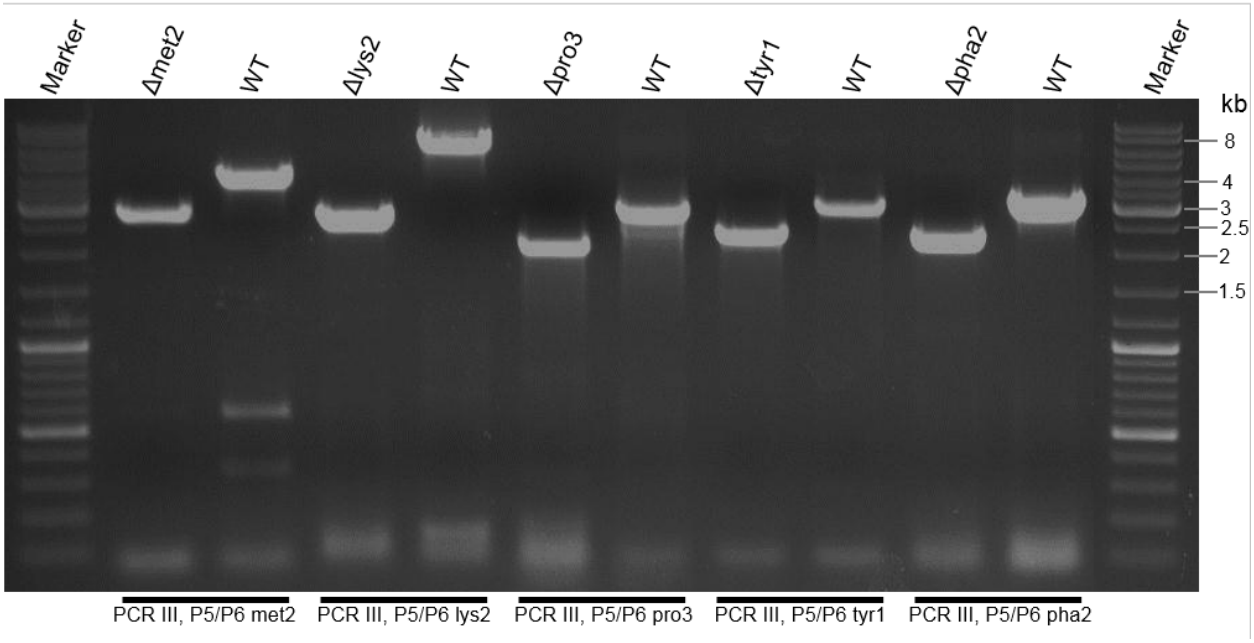


Figure 3: Verification of successful gene knockout by PCR analysis using knockout specific primer pairs P5/P6 (PCR III of Figure 2). The results for wild type and knockout strains are shown. Marker: GeneRuler DNA Ladder Mix (Thermo Scientific); WT: wild type CBS7435.

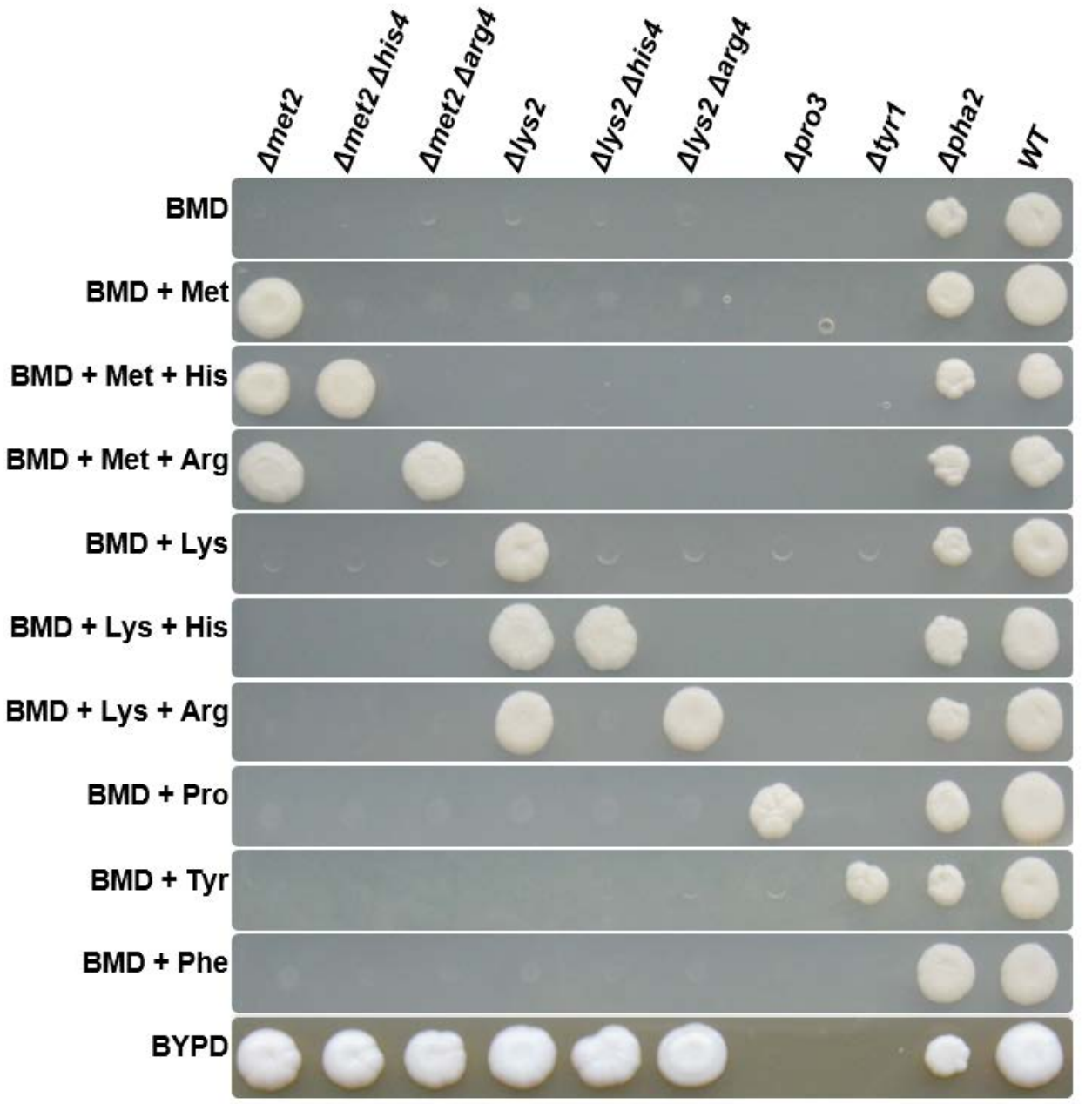


Figure 4: Growth behaviour of *P. pastoris* biosynthetic gene knockout strains. Upon cultivation in 96-well deep-well plates containing 250 µl BMD media supplemented with the respective amino acids for 24 h at 28°C, 320 rpm and 80% humidity, approximately equal number of cells (OD_{600} = 0.5) were pinned onto BMD/BYPD plates (supplemented with or without respective amino acids) and incubated for 3-4 days at 28°C.

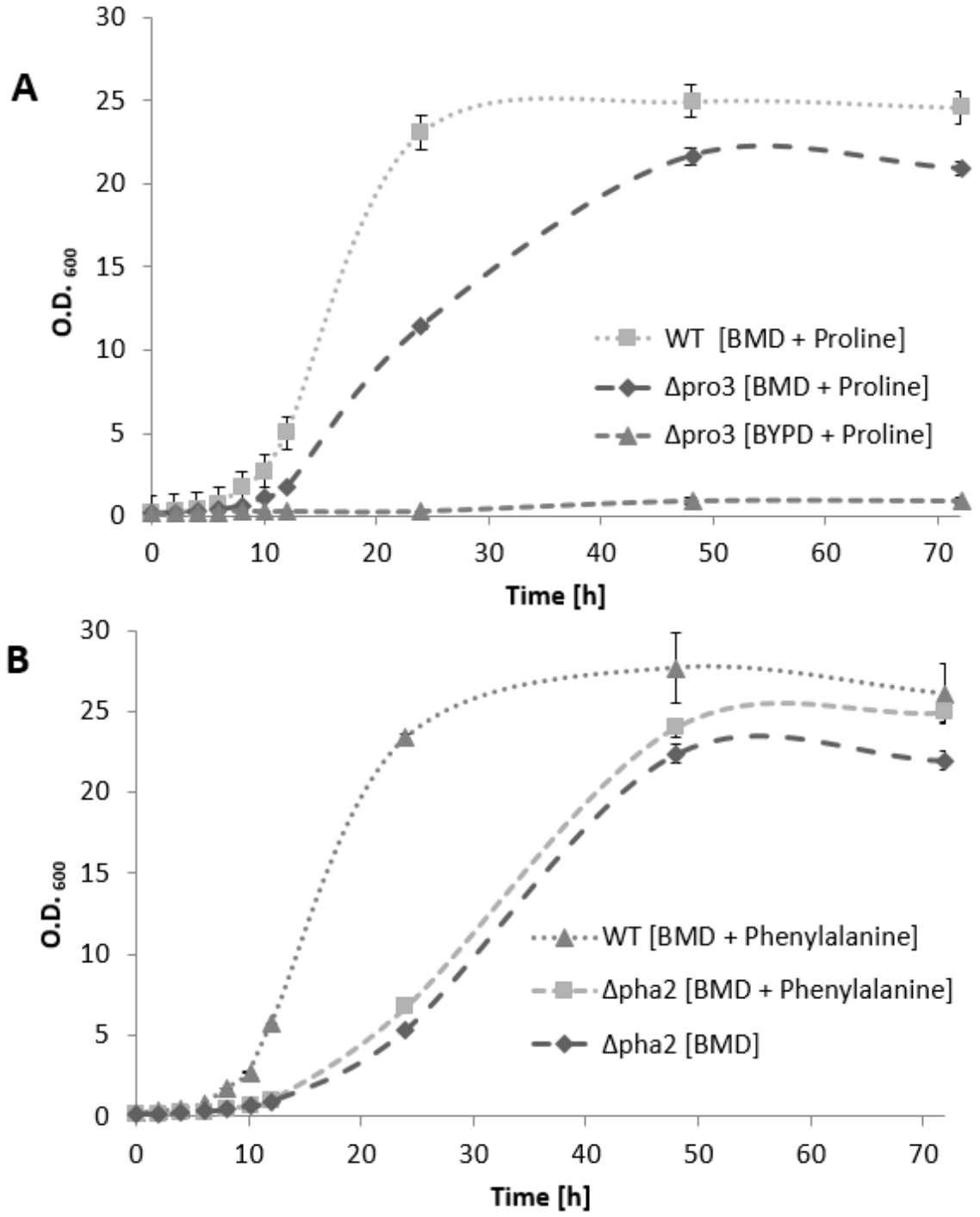


Figure 5: Growth rate analysis of *P. pastoris* wild type, *pro3* and *pha2* strains. The strains were cultivated in 300 ml baffled shake flasks at 28°C and 120 rpm. Experiments were performed in triplicate.

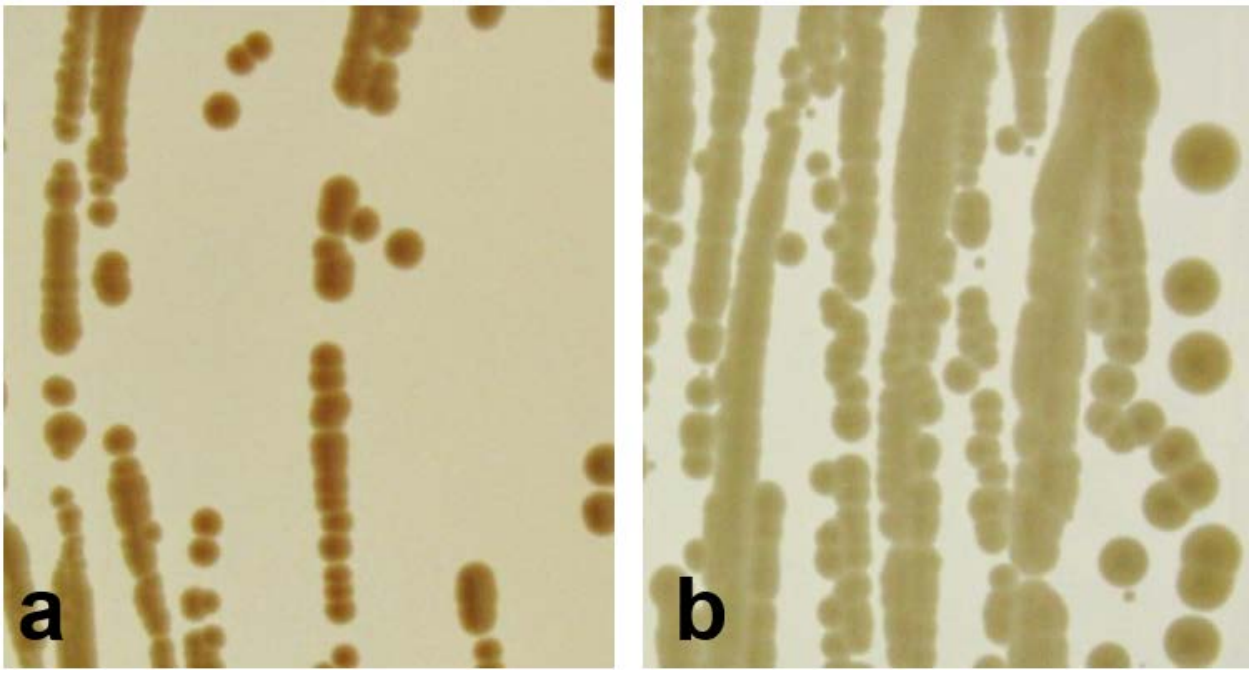


Figure 6: Phenotype of the *pha2* strain on (a) buffered minimal media, (b) buffered minimal media supplemented with phenylalanine (150 µg/ml) after ~10 days of incubation at 28°C

.

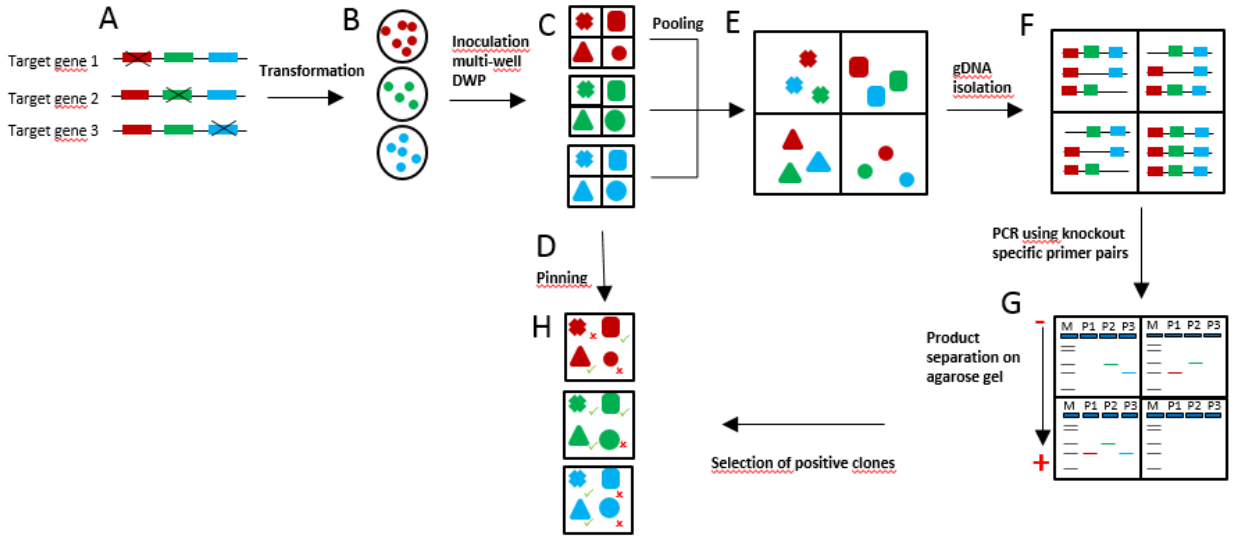


Figure 7: Strategy for fast identification of positive gene targeting events in multi-well format. (A) In separate DNA transformation reactions different genes are targeted with specific gene targeting cassettes. (B) After transformation, cells are plated on selective medium. (C) Single colonies are used to inoculate wells in DWPs. (D) Cells are pinned onto agar plates, generating the transformant library. (E) Cells from well A1 of different DWPs are pooled, cells from well A2, B1 etc. (F) Isolation of gDNA producing mixed template DNA. (G) For each targeted gene a PCR reaction with primer pairs P5/PAox1SeqR or PucSeqF/P6 is performed. A PCR product is obtained only in case of legitimate homologous recombination. (H) Clones with positive PCR result are selected and retrieved from the transformant library.

Table 1: Knockout efficiencies for biosynthetic and protease genes

#	Deleted gene	Protein ID CBS 7435	Protein Size (a.a.)	Signal peptide	Selection Marker	% Targeting efficiencies	*S.cerevisiae* Homolog	Blastp e-value[d]
Biosynthetic gene knockouts – phenotypic growth analysis on selective media								
1	Met2[a]	CCA40261.1	475	--	Zeocin™	14.9	Met2p	1.4e-134
2	Lys2[a]	CCA37057.1	1400	--	Zeocin™	4.2	Lys2p	0
3	Pro3	CCA40748.1	274	--	Zeocin™	33.7	Pro3p	1.1e-70
4	Tyr1	CCA38031.1	431	--	Zeocin™	8.5	Tyr1p	1.1e-141
5	Pha2[b]	CCA40709.1	299	--	Zeocin™	47.2	Pha2p	1.8e-51
Protease gene knockouts – confirmation by PCR using gDNA isolated by pooling method								
1	sub2	CCA37470.1	477	YES	Zeocin™	20	Prb1p	4.2e-96
2	pep4[c]	CCA39046.1	410	YES	*HIS4*	72	Pep4p	1.5e-157
3	prb1	CCA36690.1	559	YES	Zeocin™	68	Prb1p	7.3e-147
4	prc1	CCA36928.1	523	YES	Zeocin™	8	Prc1p	1.1e-173
5	yps1	CCA40555.1	599	NO	Zeocin™	88	Yps1p	1.3e-95
6	yps2	CCA39867.1	527	YES	Zeocin™	33.6	Yps1p	5.8e-44
7	yps3	CCA39863.1	473	YES	Zeocin™	ND	Yps1p	4.5e-39
8	yps7	CCA39772.1	582	YES	Zeocin™	36.7	Yps7p	1.8e-23
9	kex1	CCA38812.1	624	YES	Zeocin™	36.7	Kex1p	2.8e-90
10	kex2[c]	CCA38676.1	777	YES	*HIS4*	56.3	Kex2p	6.8e-176
11	prtP	CCA38447.1	1810	YES	Zeocin™	26.7	Flo10p	8.2e-08
12	ctse	CCA36842.1	536	NO	Zeocin™	32	Mkc7p	5e-41
13	kpx1	CCA40794.1	327	YES	Zeocin™	8	Ent2p	0.99
14	kpx2	CCA37536.1	509	YES	Zeocin™	ND	Ape3p	5e-116
15	kpx4	CCA39283.1	612	YES	Zeocin™	6	Mkc7p	1.9e-30
16	kpx6	CCA37160.1	444	YES	Zeocin™	ND	RRT12p	5.3e-70
17	kpx8	CCA40011.1	593	YES	Zeocin™	48	Yps1p	6.9e-48
18	kpx9	CCA36656.1	410	YES	Zeocin™	ND	YDR415Cp	1.6e-78
19	kpx10	CCA38814.1	1610	YES	Zeocin™	72	Flo10p	8.9e-08
20	kpx12	CCA40896.1	381	YES	Zeocin™	ND	GPI8p	4.2e-128
21	kpx13	CCA39526.1	534	YES	Zeocin™	ND	YBR139Wp	7.7e-152
22	kpx17	CCA39747.1	578	YES	Zeocin™	76	ECM14	5.7e-108
23	kpx20	CCA40153.1	587	YES	Zeocin™	56.7	CWP1p	0.034
24	kpx21	CCA40152.1	1474	YES	Zeocin™	42.7	FLO10p	8.5e-10
25	kpx24	CCA36885.1	276	NO	Zeocin™	26.7	SRT1p	1.4e-47
26	kpx25	CCA39190.1	990	NO	Zeocin™	63.3	PFF1p	1.9e-145

a = The targeting efficiencies for *MET2* and *LYS2* loci were averaged from transformations into different strain backgrounds, e.g. wild type, *his4* and *arg4*.

b = 83 of totally 176 transformants showed leaky and retarded growth on BMD media. 24 of these 83 growth-retarded transformants were screened for integration of the knockout cassette into the correct locus and all of them were positive. Therefore, we assumed that all 83 clones with retarded growth were successful *pha2* knockouts.

c = Knockout was not successful with Zeocin™ marker

d = e-values represent homology to the closest *S. cerevisiae* homolog performed at SGD (http://www.yeastgenome.org/), ND = No deletion.

Table 2:	Strains used and constructed during this study			
Strain	**Genotype**	**Knockout Vector Used**	**Specific Growth Rate**	**Reference**
CBS7435	WT	--	0.29 ± 0.00	[9]
Pp3520	*his4*	--	--	[9]
Pp3521	*arg4*	--	--	[9]
Pp3445	*aox1*	--	--	[9]
Pp7030	*met2*	pPpKC1_Met2	0.28 ± 0.00 [b]	This study
Pp7031	*arg4 met2*	pPpKC1_Met2	0.30 ± 0.00 [b]	This study
Pp7032	*his4 met2*	pPpKC1_Met2	0.30 ± 0.00 [b]	This study
Pp7033	*lys2*	pPpKC1_Lys2	0.32 ± 0.01 [b]	This study
Pp7034	*arg4 lys2*	pPpKC1_Lys2	0.28 ± 0.00 [b]	This study
Pp7035	*his4 lys2*	pPpKC1_Lys2	0.27 ± 0.00 [b]	This study
Pp7036	*pro3*	pPpKC1_Pro3	0.28 ± 0.01 [b]	This study
Pp7037	*tyr1*	pPpKC1_Tyr1	0.27 ± 0.00 [b]	This study
Pp7029	*pha2*	pPpKC1_Pha2	0.16 ± 0.00 [b]	This study
Pp6668	*sub2*	pPpKC1_sub2	0.35 ± 0.01	This study
Pp6911	*his4 pep4*	pPpKC3_pep4	0.34 ± 0.00	This study
Pp6912	*prb1*	pPpKC1_prb1	0.32 ± 0.00	This study
Pp6676	*prc1*	pPpKC1_prc1	0.35 ± 0.00	This study
Pp6686	*yps1*	pPpKC1_yps1	0.34 ± 0.01	This study
Pp6671	*yps2*	pPpKC1_yps2	0.35 ± 0.00	This study
Pp6907	*yps7*	pPpKC1_yps7	0.34 ± 0.00	This study
Pp6909	*kex1*	pPpKC1_kex1	0.34 ± 0.00	This study
Pp6910	*his4 kex2*	pPpKC3_kex2	0.30 ± 0.00	This study
Pp6673	*prtP*	pPpKC1_prtP	0.34 ± 0.00	This study
Pp6687	*ctse*	pPpKC1_ctse	0.34 ± 0.01	This study
Pp6669	*kpx1*	pPpKC1_kpx1	0.35 ± 0.00	This study
Pp6906	*kpx4*	pPpKC1_kpx4	0.32 ± 0.02	This study
Pp6670	*kpx8*	pPpKC1_kpx8	0.32 ± 0.01	This study
Pp6908	*kpx10*	pPpKC1_kpx10	0.32 ± 0.00	This study
Pp6677	*kpx17*	pPpKC1_kpx17	0.32 ± 0.00	This study
Pp6680	*kpx20*	pPpKC1_kpx20	0.32 ± 0.00	This study
Pp6681	*kpx21*	pPpKC1_kpx21	0.37 ± 0.00	This study
Pp6684	*kpx24*	pPpKC1_kpx24	0.36 ± 0.00	This study
Pp6685	*kpx25*	pPpKC1_kpx25	0.36 ± 0.00	This study
Pp7013	*his4 pep4 prb1*	pPpKC1_prb1	0.25 ± 0.00	This study
Pp7076	*yps2 yps1*	pPpKC1_yps1	0.33 ± 0.00	This study
Pp7077	*yps7 yps2*	pPpKC1_yps2	0.32 ± 0.00	This study
Pp7078	*his4 kex2 yps1*	pPpKC1_yps1	0.24 ± 0.01	This study
Pp7079	*his4 pep4 kex2*	pPpKC3_kex2	0.25 ± 0.00	This study
Pp7080	*his4 pep4 prb1 kex2*	pPpKC3_kex2	0.17 ± 0.00	This study
Pp7015	*his4 pep4 aox1*	Aox1 flipper cassette	--	This study
Pp7016	*prb1 aox1*	Aox1 flipper cassette	--	This study
Pp7019	*prc1aox1*	Aox1 flipper cassette	--	This study
Pp7014	*his4 pep4 prb1 aox1*	Aox1 flipper cassette	--	This study
Pp7017	*yps1aox1*	Aox1 flipper cassette	--	This study
Pp7018	*yps7 aox1*	Aox1 flipper cassette	--	This study
Pp7020	*kex1 aox1*	Aox1 flipper cassette	--	This study

Supporting information

Additional File 1: Primers used in the current study.

#	Primer name	Sequence(5'-3')
1	PciIFRTSfiI1F	TCACATGTGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGCCGAT CAGGCCCAACTTACTCCAACCTTCTACGA
2	BglIIFRTSfiI2R	TTAGATCTGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCGGCCCTA GTGGCCGAGTTAGAGTTGACAACACGGCAG
3	P(AOX1)forw	AAGGTACCAGATCTAACATCCAAAGACGAAAG
4	PAox1R	CGTTTCGAATAATTAGTTGTTTTTTGATCTTC
5	FLPF	TTATTCGAAACGATGCCACAATTTGATATATTATG
6	FLPR	TTATATGCGTCTATTTATGTAGGATGAAAGG
7	Aox1TTF	GACGCATATAAGTTTTAGCCTTAGACATGACTG
8	Aox1TTR	CGTTCCGTTCCGCACAAACGAAGGTCTCAC
9	PAGR4F	CTTCGTTTGTGCGGAACGGAACGTATCTTAG
10	PARG4R	GTAACAACACTAGCTGGTAATAAGTTTAGAAC
11	EM72F	CTTATTACCAGCTAGTGTTGTTACTTTATACTTCCG
12	ZeoR	CAAACTCAGTATATTAGTCCTGCTCTTCTGCGAC
13	Arg4TTF	GAGCAGGACTAATATACTGAGTTTGTTAATGATA
14	Arg4TTKpnIR	GTGGTACCAATGCGAGGATGCTGCTGGAGAC
15	PucOriPciIR	ACATGTGAGCAAAAGGCCAGCAAAAGG
16	PucOriKpnIF	CCAGCAGCATCCTCGCATTGGTACCACTGAGCGTCAGAC
17	3UTRMET2F	TCGGCCGATCAGGCCGCTACAAAGTAAAAATTTGCCTCACGC
18	3UTRMET2R	GGACTTATGGTAGTTGGATTTAAATTCTAGTTGGGCTTGTGTACCTTTG
19	5UTRMET2F	GCCCAACTAGAATTTAAATCCAACTACCATAAGTCCTAGCTC
20	5UTRMET2R	TCGGCCCTAGTGGCCAATTCAATATGTTGAGGCG
21	Up5UTRMET2F	CCACCAGTAACTTTCCCTATATACTCAGC
22	Down3URMET2R	GATACGTAGTCTGCTCTTTGCTTTCTG
23	3'UTRLys2F	TCGGCCGATCAGGCCTTCTATATGTAAGTGATATTAAAC
24	3'UTRLys2R	CGCCCAAGTTGAAATATTCAGAAGTACGGGGTAGAAGGCC
25	5'UTRLys2F	TGAATATTTCAACTTGGGCGTCAGC
26	5'UTRLys2R	TCGGCCCTAGTGGCCGGTTCGTTCTTAAGAGTGCG
27	5'Lys2OutsideF	GAGAAGAAGAGGAAACTGCC
28	3'Lys2OutsideR	CACATGTGGACATACTCCCATG
29	3UTRPRO3F	TTTCTAGAGAATAGGAACTTCGGTAAATTCACTGACTGCCTCTTTCTTTC
30	3UTRPRO3R	CTTGGAATGGTTAATTTAAATATTTTCAGTATCAAACCCGTTGAACTTG
31	5UTRPRO3F	GATACTGAAAATATTTAAATTAACCATTCCAAGAAATGCATCTTTCCG
32	5UTRPRO3R	TCTCTAGAAAGTATAGGAACTTCCTTTATTAGTGATTAAGCAAACTAAA GTGGGAG
33	Up5UTRPRO3F	GGCTTGTCAGAATGTTCAGCTTCGGC
34	Down3UTRPRO3R	GTAAGGTCCGCTGGGTCCATAAAACTGTC
35	N3UTRTYR1F	TCGGCCGATCAGGCCCATACTTCAGCAATTTTCATTGAGCAAGG
36	N3UTRTYR1R	CTGCTGCATTTAAATAACAGCCGCTGTGTCCG
37	N5UTRTYR1F	GCGGCTGTTATTTAAATGCAGCAGATCAGTATAGTTTGAACTTG
38	N5UTRTYR1R	TCGGCCCTAGTGGCCTCTCTTCTCTATCACATCCGATCACC
39	Up5UTRTYR1F	GATCACGTTCAAGAGAGGTTTGGATTCC
40	ND3UTRTYR1R	CAGCATTGTATGTAGTTCATCCCTAGC
41	3UTRPHA2F	TCGGCCGATCAGGCCAAAAGGGTTAAGTGTAAGATGTAAATATATTAAT TTCG
42	3UTRPHA2R	CTGCAATGGCTGGATATTTAAATGATCGATATGACTCCCCTTCTGG
43	5UTRPHA2F	GTCATATCGATCATTTAAATATCCAGCCATTGCAGTTTGGATTC

44	5UTRPHA2R	TCGGCCCTAGTGGCCGGTTAGGTTATCCTATATGGGGGAACG
45	Up5UTRPHA2F	GTCGTTCATGAAAGACCTGCGC
46	Down3UTRPHA2R	GTTGAATTCCAGAAGCCTTGAGATCTATG
47	3UTRsub2F	TCGGCCGATCAGGCCTGACTCATTGACCCCAGCTCAAC
48	3UTRsub2R	GGGACTGACCCGGGTGAGGAAAACACTCATTGAAATTCCTG
49	5UTRsub2F	CCTCACCCGGGTCAGTCCCAACTTGTTGG
50	5UTRsub2R	TCGGCCCTAGTGGCCGATCCCTGTAATTTCAGCGATGGAG
51	Up5UTRsub2F	ACGATTAAGGCAAATCTTCCGGTTC
52	Down3UTRsub2R	GAAACAAATCAGTGACGGCGATGTC
53	3UTRapr1F	TCGGCCGATCAGGCCCTCAGTTTATGACCTAGGCAAAGATGC
54	3UTRapr1R	GATAAAGGTCCCCGGGACCTCGGTTGTAAGCGGTAATTC
55	5UTRapr1F	CCGAGGTCCCGGGGACCTTTATCACGTTGAATCTAGTTG
56	5UTRapr1R	TCGGCCCTAGTGGCCGCTTGTGTATCTTAGCAGAATGAACTTTGG
57	Up5UTRapr1F	GAAAATAGTGTATCACTGCCAGCATC
58	Down3UTRapr1R	CTCATCTATACCCCAGGACCAG
59	3UTRprb2F	TCGGCCGATCAGGCCCACTGTCACCATTAGCACCAAACTG
60	3UTRprb2R	GCCTCTAATCCCGGGAAAGTTTAACTTCATACAGAATAACTTCATG
61	5UTRprb2F	AAACTTTCCCGGGATTAGAGGCGGTTGAACTCTG
62	5UTRprb2R	TCGGCCCTAGTGGCCGTTGCTTCCTCCGACGATACTG
63	Up5UTRprb2F	GCAGTATCCTGCTCATCTTCCCGTAC
64	Down3UTRprb2R	CATGAACGTGTTGAACTTGGACGCC
65	3UTRkpx16F	TCGGCCGATCAGGCCGGCATCTGCAAGGACAGACC
66	3UTRkpx16R	CACCTATCCCGGGAAAAGGCACATAAAGCAATCAATC
67	5UTRkpx16F	GCCTTTTCCCGGGATAGGTGATCCCTCAAAGAAGG
68	5UTRkpx16R	TCGGCCCTAGTGGCCGGCCCCATATGATCAGCCAG
69	Up5UTRkpx16F	CAAGTTCAAATGGCTCCATGGAGC
70	Down3UTRkpx16R	GCATTGAGGAAGTACATGGTCACG
71	3UTRyps1F	TCG GCC GAT CAG GCC TTC AGC TAC CTT TCT CTC TGT TTG G
72	3UTRyps1R	C TGG CCC GGG CGT CTG GTT GTT TGT ATT AGC
73	5UTRyps1F	ACG CCC GGG CCA GGG ACC TAA TTA TGA CAT C
74	5UTRyps1R	TCG GCC CTA GTG GCC CGC TGA AGT CCA ACT GTT GAA CG
75	Up5UTRyps1F	CGA ACC TAA TCA ATG ACG GTT ACG AG
76	Down3UTRyps1R	TCG GCA TTA TCT GGT AGA TCC GG
77	3UTRyps2F	TCG GCC GAT CAG GCC CTG AGT GCA AGT AGA ATT AAG CTG CTA G
78	3UTRyps2R	CCA ATA ACC CGG GTC TGA ACA TCC TGA TTG AAA GC
79	5UTRyps2F	GTT CAG ACC CGG GTT ATT GGT GAT CAA GGT TCC TTC
80	5UTRyps2R	TCG GCC CTA GTG GCC GTC GGT GTC TCA CAT TAA CAC TAG TTC
81	Up5UTRyps2F	AAT TCA TGA TTC CGG AGT GCG TGT AAT C
82	Down3UTRyps2R	TTG ATT GGC GTA GCT GGT GAT GAC
83	3UTRctsdF	TCGGCCGATCAGGCCAATTAG CCG ACG AAA AGC ATA TCA GAG AC
84	3UTRctsdR	ATT CAT GCC CGG GTT GAG AGG CAT ATC GAG AAG
85	5UTRctsdF	CTC TCA ACC CGG GCA TGA ATT TAT TGG TGA TTG CTT AAA G
86	5UTRctsdR	TCG GCC CTA GTG GCC TCT TTA CAT CAT TGG CCA GTC TGT TGA C
87	Up5UTRctsdF	CAA TTG ACA CCA AAG GAC AGT TTA GAC TC
88	Down3UTRctsdR	GTG TGA AAT GCG CTG ATC GAA CTG
89	3UTRkex1F	TCG GCC GAT CAG GCC TGG ACG ACC TGG AAT CCA AC
90	3UTRkex1R	T GAC AAC CCG GGT CGG CAG ATG AGT CTT TG
91	5UTRkex1F	CC GAC CCG GGT TGT CAT TAT TGG TGG TAA GGC
92	5UTRkex1R	TCG GCC CTA GTG GCC TCG ATT GGA GAG GCT GAC ACC
93	Up5UTRkex1F	AGT GAA GAG AAT TCA CGA GTA CAA GAG AC
94	Down3UTRkex1R	GGT AAA ACC GTC GTC ATT GCT ATT GC
95	3UTRkex2F	TCG GCC GAT CAG GCC CTC TCG GAA GTC TCC AAA GCC
96	3UTRkex2R	TGA AGG CCC GGG ATA GTC GCC TTC CGG TTT C
97	5UTRkex2F	GAC TAT CCC GGG CCT TCA TCA TCG GAG TC
98	5UTRkex2R	TCG GCC CTA GTG GCC GAT CTC ATC CCA GCC GAT GAC

99	Up5UTRkex2F	CCT ACA TCA AAT AAA TCC GCC TGC G
100	Down3UTRkex2R	CTC CGC ATA TAG TAC CCA TCC AGG
101	3UTRprtPF	TCG GCC GAT CAG GCC AAA GCA GAT GCC AAT CCT ACC AG
102	3UTRprtPR	TTA AAG ACC CGG GCT GTT TTC ACA CTT GAG TCA G
103	5UTRprtPF	AAA ACA GCC CGG GTC TTT AAA CTG TCC AAT GGA AAC C
104	5UTRprtpR	TCG GCC CTA GTG GCC CGC TGC TGT GAT TTT CTC AAT TCT TCC
105	Up5UTRprtPF	TAT TCC TCG GAA CAA TCC TCT GTA AC
106	Down3UTRprtPR	TCC TGA CTA CTA CTC CTA ACA GTA GAA AGG
107	3UTRctse2F	TCG GCC GAT CAG GCC GTG TTA CGT TGG CAG TTT GAC TAA GG
108	3UTRctse2R	AGA AGT ACC CGG GCG AAC ATG AAC ATA TTG GCT G
109	5UTRctse2F	ATG TTC GCC CGG GTA CTT CTC TGT TCA CTT TGG GTC TTA TTC
110	5UTRctse2R	TCG GCC CTA GTG GCC TGA GTG TTG GTC CCT GCA TTA TTG
111	Up5UTRctse2F	TTG ATA AGC GGC TAC CAA GTC AGA C
112	Down3UTRctse2R	AAC ATT GAC CCT TGA GTT GTT ACT CGG
113	3UTRkpx1F	TCG GCC GAT CAG GCC TCT GTA GGA GAT GCC AAT GTC ATT G
114	3UTRkpx1R	GCTGATCTTATCCCGGGAAGTAATTTTTTAAACATATTGATAAACAGAC
115	5UTRkpx1F	AAA TTA CTT CCC GGG ATA AGA TCA GCA GGT ATG AAT G
116	5UTRkpx1R	TCG GCC CTA GTG GCC CTA CAC CAA AGC CAG GTT GCC AAA C
117	Up5UTRkpx1F	GTG ACC CTA TCT GGA AAG TCG AGA C
118	Down3UTRkpx1R	CTT CAC CAG GTC CAA AGT TGA TGG
119	3UTRkpx4F	TCG GCC GAT CAG GCC CAA CTA CGA CCC AAG CAT ATC AGA TG
120	3UTRkpx4R	GAA TTG CTA CCC GGG CAC CTG GAT TGA ATG CAA G
121	5UTRkpx4F	AGG TGC CCG GGT AGC AAT TCC GAG TTA ACA TAA CTG
122	5UTRkpx4R	TCG GCC CTA GTG GCC CAA TGG CGA CGA TTC CAG CAT AG
123	Up5UTRkpx4F	ATG GAG TGG CCC GTG ATT GAA ATA TTG
124	Down3UTRkpx4R	GGT TCT TCC AGT ATT AAA CCT AAC TTG ACA GG
125	3UTRkpx8F	TCGGCC GAT CAG GCC CTT CAG GTG CAT CTT CTG CTA CTC AAA ATG
126	3UTRkpx8R	CAA ATA TGT CCC GGG CAT AAG TAT CAA TGT ACT TCT CAA TG
127	5UTRkpx8F	GAT ACT TAT GCC CGG GAC ATA TTT GCC CTC TGA TTG CAC
128	5UTRkpx8R	TCG GCC CTA GTG GCC GCA ACT TGT TAG CCT TGA AAG GCG ATT G
129	Up5UTRkpx8F	GCC TCT GAC AGA GCG TTG ACC TG
130	Down3UTRkpx8R	CAA AGA TCT TGG TGG CTT CGT CC
131	3UTRkpx10F	TCG GCC GAT CAG GCC GTT ATC GTT GCT GAT TTG AAC AGA TGC TC
132	3UTRkpx10R	GA AAA GAC CCG GGC TTC GTA GGA GAG G
133	5UTRkpx10F	TA CGA AGC CCG GGT CTT TTC GAC GTG GTT GAT AAA G
134	5UTRkpx10R	TCG GCC CTA GTG GCC GAA TTC AAG CTG CTT CAA ACA GCA C
135	Up5UTRkpx10F	GTC TTG TTC CAG TGA ACA ACC AGC
136	Down3UTRkpx10R	TGG GTA GGG TTG TAT GGG AAC G
137	3UTRkpx17F	TCG GCC GAT CAG GCC GGT CGT TGT TTC GCG CTC
138	3UTRkpx17R	T ACT GCC CGG GTG GCT GGT GGT AGT G
139	5UTRkpx17F	GCC ACC CGG GCA GTA AGG AGT CAT CAA GAG
140	5UTRkpx17R	TCG GCC CTA GTG GCC GTG TGC TAT TGA CAA GGT GGT CTT ATA G
141	Up5UTRkpx17F	GCT TAT TCT CAG CAC TGG ATA CAC CTC
142	Down3UTRkpx17R	GAT TCG CCA AGA TCC TGG CC
143	3UTRkpx19F	TCG GCC GAT CAG GCC TGT TAC CAC TAC TCA AGC CAC ACA AG
144	3UTRkpx19R	GAC GTT CCC GGG TCT TTT CGA CGT GGT TGA TAA AG
145	5UTRkpx19F	GAA AAG ACC CGG GAA CGT CTA GTG ATG TGC TAC
146	5UTRkpx19R	TCG GCC CTA GTG GCC CAC GCA AAC TTT GCA AAT CCT GGA AAG
147	Up5UTRkpx19F	GAT TCC AGA AGG TCA ACG TTG TTT CTA C
148	Down3UTRkpx19R	AAC AAC CAG CTT GTC CCT GAT TGA G
149	3UTRkpx20F	TCG GCC GAT CAG GCC CGG CAG CCT GTA GGA TAA TAA GAA G
150	3UTRkpx20R	AAT GCT CCC GGG AGT TTT CAC GTG TTC TAC GG
151	5UTRkpx20F	TGA AAA CTC CCG GGA GCA TTC CAA AAC ATT GCG
152	5UTRkpx20R	TCG GCC CTA GTG GCC CTT ACA CTG TGA GTG AGG CTC AAG
153	Up5UTRkpx20F	CAG CAC TAT TTT GCT TGG CTT GC

154	Down3UTRkpx20R	GTG TAT TGA TGC GCA GAT ACT CAG C
155	3UTRkpx21F	TCG GCC GAT CAG GCC TAA GGC TTG TTC GTA AGA AGG CAA AG
156	3UTRkpx21R	TAA CTG CCC GGG CAA CAT GGA AAC ACG TG
157	5UTRkpx21F	ATG TTG CCC GGG CAG TTA GTC CTG TAC TTG G
158	5UTRkpx21R	TCG GCC CTA GTG GCC AGG ACA CTA CGA TAA CAA CTC ATG TTA C
159	Up5UTRkpx21F	TGC AGT ATG TTA CGC ACT GCA TTA TAT CG
160	Down3UTRkpx21R	GAT AAA ATT CCA CGG CTG ACC GG
161	3UTRkpx24F	TCG GCC GAT CAG GCC TTG TGG ATC TAC TGA AGG ACT TGT GAG
162	3UTRkpx24R	GAC TAG GAC CCG GGT CAT AAT CGC ATC TAC TGT TGT G
163	5UTRkpx24F	GAT TAT GAC CCG GGT CCT AGT CTA TGG ACT AAC AGC
164	5UTRkpx24R	TCG GCC CTA GTG GCC CTG GCA GAT GAT GTG GAT TCT GC
165	Up5UTRkpx24F	CTT TGA AAG TGC ACT ACA ACT GGT CC
166	Down3UTRkpx24R	TTC GTA TCG TCC ACG ATA GGT AAA TGT C
167	3UTRkpx25F	TCG GCC GAT CAG GCC ATT CCT GAA GGA CTG CAA GTC TG
168	3UTRkpx25R	CTT AGA TCC CGG GAA GTT TAG AGG AAA TCT GTC TCA AAT AAG
169	5UTRkpx25F	TCTAA ACT TCC CGG GAT CTA AGT TTG GAG ATT CTG ACC
170	5UTRkpx25R	TCG GCC CTA GTG GCC GTT GGA ATC GAA TGG CTC TAT GAT TGG
171	Up5UTRkpx25F	GAC AGC CTC GAT TAT TTC TCT TTC TCT TCG
172	Down3UTRkpx25R	AAG GCG TAC CAA GAG CCT TTA GC
173	nCDSsub2F	GCACACTCGCTTTTGATACCATCTC
174	nCDSsub2R	ATCCGAGTCATCAAGTACATCCTTGG
175	nCDSpep4F	CTCTCTACTCTAGGTATTGGTGCTGAAG
176	nCDSpep4R	ACCTACTGCATCTTTGCCTAGGTC
177	nCDSprb1F	AAACTCTTGGGCCAAGTTTTCAACAG
178	nCDSprb1R	GATTGGCTATCTTATCTGCCATAGCAG
179	nCDSprc1F	ATGAGAATTCTCTGGCTGATCATATGGG
180	nCDSprc1R	TCCTAAAGCTATTGGTCTGTCCTTGC
181	nCDSyps1F	AAA ACG TTG TTG GCG TTC AAC AGT TG
182	nCDSyps1R	ATA CTA TAC ACA CGC CGA GAA TAA CTA CC
183	nCDSyps2F	GAA CTA GTG TTA ATG TGA GAC ACC GAC
184	nCDSyps2R	CCT AGC AGC TTA ATT CTA CTT GCA CTC
185	nCDSctsdF	ACA GTC AAC AGA CTG GCC AAT GAT G
186	nCDSctsdR	GGT TTT GTC TCT GAT ATG CTT TTC GTC G
187	nCDSkex1F	TAC TGC TTC CTT TAG TAG CGG TGT C
188	nCDSkex1R	TGA GAA GTT TGC TCA TCA CAC TAT TGT CC
189	nCDSkex2F	CGG AAT CCA AGG TGA ATT GAC TAT TGG
190	nCDSkex2R	CGC TTA GCA TGC TGG ATC TTA ATT GG
191	nCDSprtPF	CCA GTG GAA GAA TTG AGA AAA TCA CAG C
192	nCDSprtPR	AAG GCC TGG TAG GAT TGG CAT C
193	nCDSctse2F	AGG CAA TAA TGC AGG GAC CAA CAC
194	nCDSctse2R	AAG GTA CAA GTG TTT CCA CCC TTA GTC
195	nCDSkpx1F	AGG GCT TTG GTG TTA GCT GAC TC
196	nCDSkpx1R	CCA ATA ACA CCA TGG CAA CTA CAG C
197	nCDSkpx4F	GTG TTG TTA CTG TCT ACA GGC TAT GC
198	nCDSkpx4R	AGT TTT GCG GCT ACC AGC ATT TG
199	nCDSkpx8F	CCA ATC GCC TTT CAA GGC TAA CAA G
200	nCDSkpx8R	ACA TGG CGT TGC ACT TAA AGA TGC
201	nCDSkpx10F	TAG ATC CAG AGT TCT CGT TCC CAG
202	nCDSkpx10R	CGG AGC ATC TGT TCA AAT CAG CAA C
203	nCDSkpx17F	CTC TAT AAG ACC ACC TTG TCA ATA GCA CAC
204	nCDSkpx17R	GTC TTA CTC CTT TCG GTA TTC TGC TCC
205	nCDSkpx19F	TCA GCA TTT GCT ACC CAA CTG AGA G
206	nCDSkpx19R	AGT TCT CGT TCC CAG CTG TAG TG
207	nCDSkpx20F	AGA TTT GGA ACC CAA GTT GAC TTG AG
208	nCDSkpx20R	CAA CAC TTC TTA TTA TCC TAC AGG CTG C

209	nCDSkpx21F	CGT GGT TTG AGT AAC ATG AGT TGT TAT CG
210	nCDSkpx21R	ATT GCT GTT CTA TTG TCA GCA TTT GCT G
211	nCDSkpx24F	GAT CGC TTG ATA ATA GCT CCA TGT AAG C
212	nCDSkpx24R	GCT CAC AAG TCC TTC AGT AGA TCC A
213	nCDSkpx25F	TCC GTC AAT TAC GCT AAC GGC ATC
214	nCDSkpx25R	AAG TCC CAA AGA GCA GAA CGC TG

Additional File 2: Elements of *E. coli*/*P. pastoris* gene knockout shuttle vectors and their function.

Elements	Origin	Function
P_Aox1	*Pichia pastoris* CBS7435	*P. pastoris* AOX1 promoter for expression of Flippase
Flippasea	*Saccharomyces cerevisiae* BY4741	Site-specific *FLP* recombinase, recycling of the marker genes (mutated to remove certain restriction sites)
Aox1_TT	*Pichia pastoris* CBS7435	Transcription terminator of *AOX1* gene in *P. pastoris* for Flippase transcription termination
P_Arg4	*Pichia pastoris* CBS7435	*ARG4* promoter for expression of selection marker genes in *P. pastoris*
Arg4_TT	*Pichia pastoris* CBS7435	Transcription terminator for expression of selection marker genes
EM 72 Syn B	Synthetic sequence, amplified from pPpT4 [10]	Constitutive prokaryotic promoter; drives expression of antibiotic resistance genes in *E. coli*
Sh_ble	Synthetic gene, amplified from pPpT4 [10]	Confers resistance to antibiotic Zeocin™
Arg4_CDS	*Pichia pastoris* CBS7435	*P. pastoris* wild type gene coding for argininosuccinate lyase; selection marker
KanMX6	KanMX6 amplified from pPpT4_Kan [10]	*KanMX6* gene; selection marker
His4_CDS	*Pichia pastoris* CBS7435	*P. pastoris HIS4* wild type gene; selection marker
pUC Ori	pUC8 (Genbank Acc. Nr. L08959)	pUC origin of replication for plasmid maintenance *in E. coli*
FRT	Synthetic FRT site	*FLP* recombinase recognition sequence for marker recycling
bla_cds	β-lactamase gene from pUC8 (Genbank Acc. Nr. L08959)	Ampicillin resistance in *E. coli*; selection marker
Stuffer	Synthetic gene HRP0 (Genbank Acc. Nr. HE963800.1)	For easier confirmation of restriction with *SfiI* restriction enzyme

Additional File 3:

Plasmid sequences of the constructs used in this study in Genbank format.

Chapter 5

Restriction site free cloning (RSFC) plasmid family for seamless, sequence independent cloning in *Pichia pastoris*

Thomas Vogl[1*#], **Mudassar Ahmad[1*],** Florian W. Krainer[1], Helmut Schwab[1], Anton Glieder[1]

[1] Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, Graz A-8010, Austria

*** These authors contributed equally.**

Contribution:

TV, MA, FWK designed the experiments, analysed and interpreted the data. MA, TV and FWK performed the experiments. TV and MA wrote the manuscript. TV, MA, FWK, HS and AG conceived of the study. AG and HS supervised the research. All authors read and approved the final manuscript.

Vogl *et al. Microb Cell Fact (2015) 14:103*
DOI 10.1186/s12934-015-0293-6

MICROBIAL CELL FACTORIES

TECHNICAL NOTES

Open Access

Restriction site free cloning (RSFC) plasmid family for seamless, sequence independent cloning in *Pichia pastoris*

Thomas Vogl[1,2*†], Mudassar Ahmad[1†], Florian W Krainer[1], Helmut Schwab[1] and Anton Glieder[1]

Abstract

Background: Tagging proteins is a standard method facilitating protein detection, purification or targeting. When tagging a certain protein of interest, it is challenging to predict which tag will give optimal results and will not interfere with protein folding, activity or yields. Ideally, multiple tags and positions are tested which however complicates molecular cloning and expression vector generation. In conventional cloning, tags are either added on PCR primers (requiring a distinct primer and PCR product per tag) or provided on the vector (typically leaving a restriction site scar).

Results: Here we report a vector family of 40 plasmids allowing simple, seamless fusions of a single PCR product with various N- and C-terminal tags, signal sequences and promoters. The restriction site free cloning (RSFC) strategy presented in this paper relies on seamless cloning using type IIS restriction endonucleases. After cutting out a stuffer (placeholder) fragment from the vectors, a single PCR product can be directly inserted in frame into all 40 plasmids using blunt end or TA ligations, requiring only verification of the orientation. We have established a RSFC vector family for the commonly used protein expression host *Pichia pastoris* and demonstrated the system with the secretory expression of horseradish peroxidase (HRP). HRP fusions to four tags (Myc, FLAG, His, Strep) and two fusion proteins (GFP and MBP) showed a 31-fold difference in volumetric activities. C-terminal tagging caused in some cases almost a complete loss of function, whereas N-terminal tags showed moderate differences.

Conclusions: The RSFC vectors provide an unprecedented toolbox for expression optimization in *P. pastoris*. The results obtained with HRP underline the importance of comparing different tags to maximize activities of fusion proteins. In a similar fashion the RSFC strategy can be applied in other expression hosts to screen for optimal promoters, signal sequences or to facilitate the evaluation of (iso-) enzyme families.

Keywords: Protein tagging, Protein tags, Seamless cloning, *Pichia pastoris*, Expression optimization, Cloning strategy, Type IIS restriction endonucleases

Background

Protein tags are commonly applied tools facilitating purification (affinity tags), enabling immuno-detection (epitope tags) or increasing solubility. Fusions to fluorescent proteins help elucidating the cellular localization and fusions to signal sequences provide specific intracellular targeting or secretion [1, 2]. However, as an extrinsic addition to a protein of interest (POI), such fusions may also show detrimental effects by affecting protein conformation, yields, activity or stability [1, 3]. The specific interactions of the POI with a certain tag are generally hard to foresee and may also depend on the position of the tag (N- or C-terminal). Unknown proteolytic processing or intracellular targeting of the POI may also influence the suitability of a specific fusion site. In addition, the same tagged protein may behave differently depending on the host system used (e.g. bacteria,

*Correspondence: thomas.vogl@aon.at
†Thomas Vogl and Mudassar Ahmad contributed equally
[2] Queensland University of Technology, 2 George St., Brisbane, QLD 4000, Australia
Full list of author information is available at the end of the article

yeast, higher eukaryotes) [4]. As there are large numbers of affinity, epitope tags and fusion proteins available it is challenging to predict the optimal choice for a certain POI. Therefore, commonly multiple tags are tested in N- or C- terminal positions and screened for optimal results [4–7].

However, preparing expression constructs containing multiple tags may require tedious cloning work. Tags are commonly provided on the plasmid adjacent to the multiple cloning site (MCS). This requires unique vectors for each tag and N-/C-terminal position. The gene of interest (GOI) needs to be cloned into the MCS via unique restriction endonuclease (RE) recognition sites. These restriction site scars remain in the protein coding sequence (CDS) and are later translated into additional amino acids, which may interfere with the POI's properties. Also cloning strategies based on recombination such as Gateway (e.g. [8]) leave the recombination sequence as a scar in the CDS.

Ideally, tags should be fused seamlessly to the GOI i.e. without any restriction site scars or additional sequences from the MCS. Seamless cloning can be achieved by various strategies [9]. Frequently, tags are directly added by PCR as a 5′ overhang of a primer and thereby seamlessly attached to the CDS. This approach requires however a unique primer for each tag, N-/C-terminal position and each GOI.

We aimed to design a simple, seamless system to facilitate testing of multiple tags in N-/C-terminal position at minimal cost and effort (e.g. without the need to order numerous primers).

Several novel cloning methods are completely independent of REs and allow simple assembly of multiple fragments solely by short overlaps (around 25 bp) relying on in vitro 'recombination' (e.g. annealing of single stranded overhangs and enzymatic linkage). These methods include SLIC (sequence and ligation–independent cloning) [10], SLiCE (Seamless Ligation Cloning Extract) [11], Gibson assembly [12], CPEC (Circular Polymerase Extension Cloning) [13] and are concisely compared on the website of the Joint BioEnergy Institute (JBEI), Emeryville, CA, USA [14, 15]. All these methods may be used to seamlessly add a tag to a protein by adding the tag sequence to a PCR primer. However, there is an additional overhang required for in vitro recombination with the vector, requiring relatively long primers. Most inconveniently a new primer is needed for each tag, each position and each POI to be tested.

Therefore we have based our strategy on type IIS REs. In contrast to type II REs, which recognize and cut within a palindromic sequence, type IIS REs cut outside of a non-palindromic recognition sequence [16, 17]. Thereby RE site scars can be circumvented making type IIS REs prominent tools for seamless cloning [9]. There are various type IIS enzymes available that create different types of overhangs including up to 4 bp overhangs suitable for sticky end cloning (e.g. *Eam*1104I [18], *Bsa*I [19, 20]), single base pair overhangs that can be applied for TA cloning (e.g. *Xcm*I [21, 22], *Eam*1105I [23, 24], *BciVI* [25]) or blunt end cloning (*Mly*I/*Sch*I [25, 26]), see Figure 1a.

In this study we have evaluated type IIS REs for blunt end and TA cloning and designed a restriction site free cloning (RSFC) strategy that enables simple, seamless cloning of a PCR product in frame with any desired upstream or downstream sequence in a vector. Based on this strategy, we have designed a RSFC vector family of 40 plasmids for the methylotrophic yeast *Pichia pastoris*, a commonly used protein production host for industrially relevant biocatalysts and biopharmaceuticals [27–29]. The vectors feature different epitope and affinity tags (Myc, FLAG, His, Strep) and fusion proteins (eGFP and MBP) in N- and C-terminal position that are provided for intracellular and secretory expression.

Results and discussion

Restriction site free cloning (RSFC)

Blunt end vs. TA cloning concept

We aimed to design a vector system in which a single PCR product of a GOI can be directly fused, sequence independently to various N- or C-terminal tags provided on different plasmids. Thereby only two primers are required to test seamless fusions of multiple tags with the GOI. This design is achieved by inserting a stuffer (placeholder) fragment flanked by two type IIS RE sites in opposite orientations in all vectors (Figure 1b–d). The CDSs of different N- and/or C-terminal tags or fusion proteins are provided upstream/downstream of the stuffer fragment. By digestion using the respective type IIS RE, the stuffer fragment including the RE sites is cut out, resulting in RE site free vector backbones that can be directly ligated with the same PCR product (Figure 1b).

Commonly used type IIS RE based cloning strategies such as Golden Gate cloning [19, 20] cannot be used for this purpose as they rely on type IIS enzymes creating short overhangs such as *Eam*1104I or *Bsa*I (Figure 1a). The use of these enzymes requires also RE digestion of the PCR product and the overhangs created on the vectors would differ between tags and impede seamless fusions.

Direct, sequence independent cloning of PCR products is in this context only possible by using TA cloning or blunt end ligations. These methods are in general not directional (with a few exceptions e.g. [25, 26]) and require verification of the orientation (e.g. by colony PCR, cPCR). TA cloning is based on the property of *Taq*-Polymerase to add a single deoxyadenine (dA) nucleotide

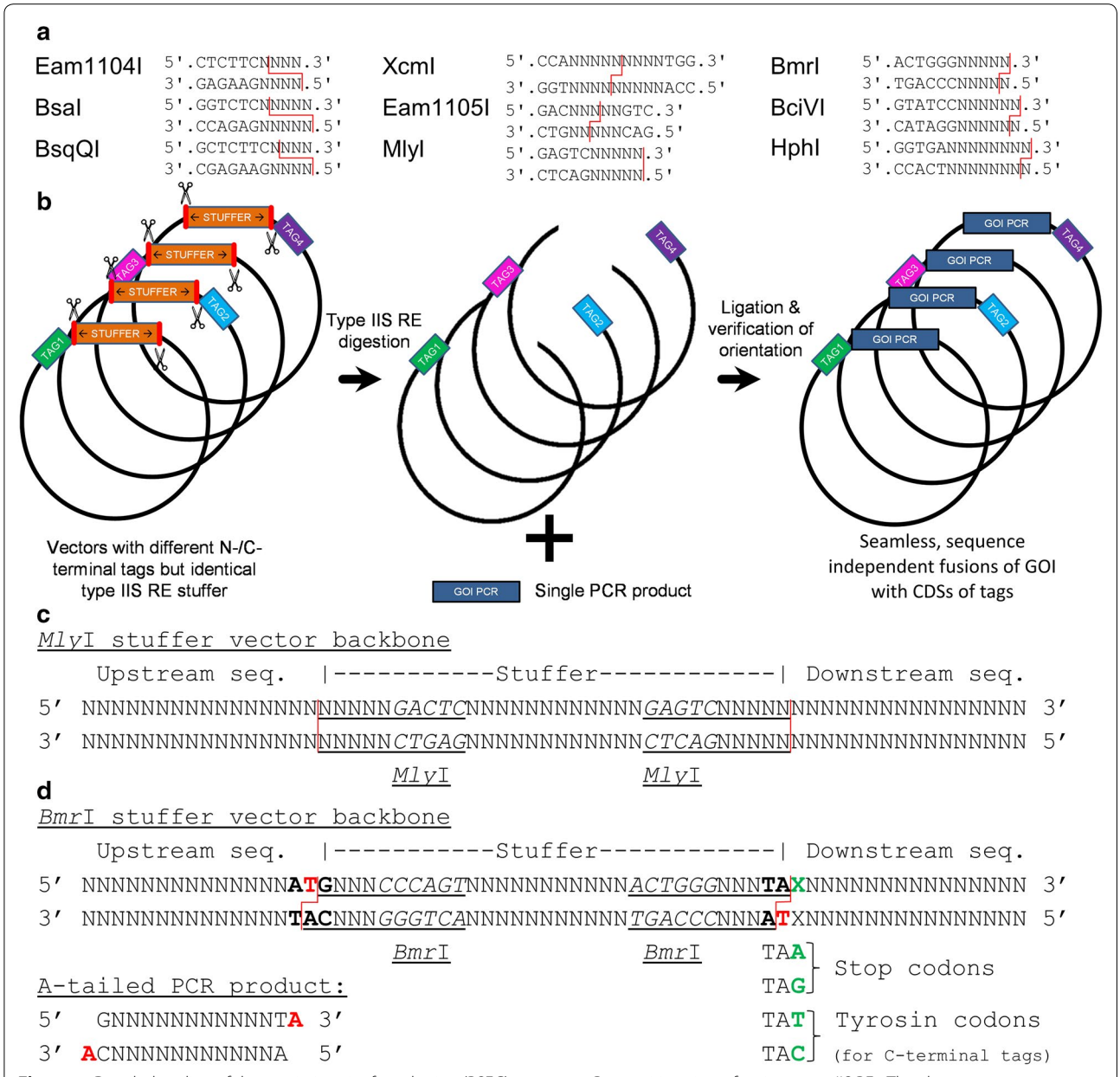


Figure 1 Detailed outline of the restriction site free cloning (RSFC) strategy. **a** Recognition sites of various type IIS REs. The cleavage patterns are indicated as *red lines*. **b** Schematic workflow of restriction site free cloning. After removal of a stuffer fragment using type IIS REs, a single PCR product can be ligated into all vectors in a seamless, sequence independent fashion. The strategy is shown for four vectors but can be extended to as many as desired. **c** Design of the *Mly*I stuffer fragment for blunt end ligations. The *Mly*I recognition sequence is written in italics, the entire cleavage pattern is underlined. Variable bp are denoted as 'N'. *Upstream* sequences may include promoters, N-terminal tags and signal sequences, downstream sequences may include C-terminal tags and stop codons. **d** Design of the *Bmr*I stuffer fragment for TA cloning. Same explanation as (**c**), in addition the incorporation of the dA and dT residues for TA cloning via Start- and Stop/Tyr-codons are shown (*red*). By varying the last nucleotide 'X' of the Stop/Tyr codon, either translation can be terminated or a C-terminal tag linked in frame. A dA-tailed PCR product suitable for ligation is also shown.

at the 3′ ends of amplified DNA [21]. These PCR products can be directly cloned using a vector with a single 3′ deoxythymidine (dT) overhang. TA cloning works more efficiently than blunt end cloning [21], however the required dA nucleotide complicates seamless fusions to tags as it must be universally incorporated in the transitions between tag and vector. In this respect, blunt end

ligations, that are completely sequence independent, are more favorable.

We designed test vectors based on type IIS REs for blunt end and TA cloning to compare their suitability. There is only one blunt end type IIS RE available that cuts outside of its recognition sequence (*Mly*I). *Mly*I has also been established for directional blunt end ligations of PCR products using a *lac*O, *lac*Z based blue-white screening in *Escherichia coli* [25, 26]. There are several type IIS REs available, that create a single base 3′ overhang (e.g. *Bmr*I, *Bci*VI, *Hph*I, see Figure 1a). We tested commercially available preparations of these three enzymes all of which showed sufficient cleavage efficiencies (data not shown). *Hph*I and *Bci*VI have been previously used for TA cloning [21, 25], yet these restriction sites were present more frequently in the vector backbones we wanted to use. Therefore we used *Bmr*I.

The basic sequence design of the transitions between the vector, the type IIS restriction sites and the stuffer fragment are shown in Figure 1c, d. For blunt end cloning using *Mly*I, the design is completely sequence independent (Figure 1c). For TA cloning, 3′ dT residues must be provided on the vector backbone and incorporated in the transition between vector and GOI. We solved this by using the dT nucleotide of the start codon (ATG) and the dA nucleotide of a partial stop codon (TAX), creating a 3′ dT base on the reverse strand (Figure 1d). Depending on the desired sequence context, 'X' may be provided on the vector side as A/G for a stop codon (translation termination) or T/C (coding for tyrosine, for linkage of C-terminal tags).

Cloning efficiencies

We compared the basic blunt end and TA cloning based system at first with expression vectors for *Schizosaccharomyces pombe* as these plasmids required fewer modifications in the vector backbones than the *P. pastoris* plasmids we intended to use. See Additional file 1: Figure S1 for plasmid maps and the "Materials and methods" section for details on the design. After cutting out the stuffer fragment using *Mly*I or *Bmr*I, the vector backbones were dephosphorylated to counter act self-ligation. Primers for insert amplification were phosphorylated prior to ligation (see "Materials and methods" section for experimental details and a simple, cost effective protocol). Both cloning strategies resulted in similar transformation efficiencies (via electroporation), approximately 10^2–10^3 colony forming units (cfu)/µg DNA (in the ligation reaction) with self-made competent cells (competence with circular, supercoiled plasmids: 10^6–10^7 cfu/µg DNA) and in both cases all 10 out of 10 clones tested contained an insert. We verified the orientation by cPCR; as statistically expected approximately half the clones contained an insert in the correct orientation (blunt end/*Mly*I: 5 of 10, TA cloning/*Bmr*I: 7 of 10). Additional file 2: Figure S2 outlines a simple cPCR strategy to test the correct orientation (using sequencing primers of the vector and the primers used for amplifying the insert). The vector/insert transitions were confirmed by sequencing and did not show any mutations. However, when cloning an insert into a larger set of vectors using blunt end ligations (see *P. pastoris* vectors below) we noticed occasionally single bp deletions of the insert adjacent to the vector transition (<5% of constructs). Sequencing of additional transformants resulted in all cases in correct sequences. Notably, the deletions were always in the 5′ ends of the insert and occurred more often after repeated freeze/thaw cycles of the PCR product. We therefore recommend aliquoting the PCR product and vector backbones and using them only once.

In general these RSFC ligations resulted in lower efficiencies (cfu/µg DNA) than comparable sticky end ligations, but still yielded sufficient numbers of transformants for our standard cloning applications. *Mly*I based blunt end ligations worked similarly efficient as *Bmr*I based TA cloning. Previously, TA cloning has been reported to be more efficient than blunt end cloning [21], however the difference may arise from the different enzymes used for vector preparation in our study.

We mutated *Mly*I sites present in the vector backbones to enable the stuffer removal (see "Materials and methods" section for details). All mutations but one resulted in no differences in DNA yields compared to the parental plasmids. Mutating a *Mly*I site in the *E. coli* pUC origin of replication to a sequence previously reported [25, 26] decreased plasmid yields to approximately 30% of the unmutated parental plasmid (wildtype pUC: ~400 ng/µl, *Mly*I mutated pUC: ~120 ng/µl). The *MlyI* site appears also in other high copy number origins of replication (ori) and switching to a lower copy number *ori* would also result in lower plasmid yields. We intended to use the RSFC plasmids only for sub cloning and aimed to perform expression in *P. pastoris*. To this end mini prep yields (typically >5 µg) were sufficient. However if similar plasmids should be used for expression in *E. coli*, we would recommend to screen other possible mutations of the *Mly*I site using degenerate primers to restore wild type plasmid yields.

However, the blunt end/*Mly*I based strategy allowed completely seamless cloning whereas seamless TA cloning was hindered by the requirement for dT/dA bases in the insert/vector transition. This problem is similar to the use of typeIIS enzymes creating longer sticky end overhangs that need to be complementary between all plasmids of a vector family (for example in plasmids by BioGrammatics, Inc., Carlsbad, CA, USA and 'Electra'

plasmids by DNA2.0, Inc., Menlo Park, CA, USA). As outlined in Figure 1d, the TA strategy can be still used for fusion of the same PCR product to different tags, however N-terminal tags are always linked via an ATG (coding for methionine/start), whereas C-terminal tags must be linked via tyrosine codons. Tyrosine is naturally a relatively rarely occurring and bulky amino acid, making it structurally rather unfavorable as a linker to an adjacent tag. In 'Electra' plasmids by DNA2.0 this issue is solved by adding an additional C-terminal 'linker' amino acid to all vectors, whereas in the RSFC strategy only vectors with C-terminal tags require a linker amino acid. Still we have solely focused on the blunt end/*Mly*I based strategy in the following plasmid design for *P. pastoris*. The blunt end/*Mly*I based ligations required no A-tailing step of PCR products but reached similar ligation efficiencies as TA cloning and allowed completely sequence independent fusions.

In summary, our cloning approach, relying on blunt end or TA ligations between a phosphorylated PCR product and a dephosphorylated vector backbone created by type IIS RE digestion, allowed seamless, sequence independent cloning at reasonable efficiencies. PCR products can be directly used for ligations and do not need RE digestion, therefore any insert sequence can be used (TA cloning with proof reading polymerases requires a separate dA-tailing step). There have previously been type IIS based cloning efforts using blunt end and TA ligations for the cloning of PCR fragments [21–26]. However, these strategies did not allow seamless fusions and are in part with *lac*O, *lac*Z based blue white screening [25, 26], despite the convenience of directional cloning, even incompatible with seamless fusions. To distinguish our approach from these efforts and other type IIS based strategies such as Golden Gate cloning [19, 20], we have termed our approach restriction site free cloning (RSFC).

RSFC plasmids for *P. pastoris* as toolbox for optimizing protein production

Tags and fusion proteins

We applied the RSFC cloning strategy to design a plasmid family for *P. pastoris* allowing seamless fusions of a GOI with various tags and fusion proteins in N- and C-terminal position. There are different expression plasmids available for *P. pastoris* based on various cloning strategies such as Gateway [8], TA cloning [22, 25], sticky end type IIS ligations (plasmids by BioGrammatics, 'Electra' plasmids by DNA2.0) and 'classical' typeII RE/ligation based systems ([30–32] and *P. pastoris* plasmids by Life Technologies, Carlsbad, CA, USA). The pCri vector family [32] is a multi-host platform, allowing to clone a single PCR product via restriction digestion and a MCS into different vectors. For *P. pastoris* only three pCri plasmids with a His tag are available. Therefore none of the vector systems currently available for *P. pastoris* offer different tags and only the BioGrammatics and Electra plasmids by DNA2.0 vectors allow seamless, yet sequence dependent cloning still requiring restriction digestion of the insert.

We designed a set of 40 RSFC plasmids for *P. pastoris* (termed pPpRSFC) offering different tags (Myc, FLAG, His, Strep) and fusion proteins (enhanced green fluorescent protein, eGFP and maltose binding protein, MBP) in N- and C-terminal position, see Table 1 for exact properties and Figure 2 for a schematic vector map. We have assigned numbers (#1 to #40) to the plasmids and are using them hereafter when referring to a specific construct.

After stuffer removal by *Mly*I digestion, a single PCR product can be cloned in a seamless and sequence independent fashion into all vectors, fused to tags and fusion proteins ranging from 18 to 1,101 bp in length. Epitope and affinity tags are included and constructs with affinity tags contain a TEV protease cleavage site to allow tag removal. The hexameric His tag is provided with and without TEV protease cleavage site. MBP is provided as a fusion protein with the potential to improve solubility and act as a purification tag, although in *P. pastoris* problems with proteolytic degradation have been reported [33]. eGFP is an enhanced version of GFP allowing simple fluorescence detection of tagged proteins.

When cloning a GOI into the pPpRSFC vectors, the forward primer must be designed starting at the DNA sequence coding for the 2nd amino acid of the POI (omitting the ATG start codon). The reverse primer must be designed starting (on the reverse strand) at the DNA sequence coding for the last amino acid/penultimate codon (omitting the stop codon). Especially a stop codon on the PCR product would interfere with tag fusions, therefore the start and stop codon are always provided on the vectors and must be omitted from PCR inserts.

P. pastoris is not only suitable for intracellular expression but can also produce secreted heterologous proteins at high titers while secreting only little endogenous protein [27–29]. Therefore we designed all plasmids also for secretory expression using the *S. cerevisiae* mating factor alpha pre-pro signal sequence (MF alpha), the most commonly applied signal sequence in *P. pastoris*. The MF alpha sequence is processed by two proteases (Ste13 and Kex2) that cleave the amino acid sequence KREAEA at the end of MF alpha [34]. Kex2 cleaves efficiently after KR whereas the Ste13 cleavage after the EA repeat may be incomplete, depending on the following amino acids of the POI. In several cases removal of the EAEA repeats has led to a more homogenous product [35, 36]. Therefore we designed the basic MF alpha pPpRSFC plasmids (#2, 3, 5, 6, 36, 37, 39, 40) with and without the EAEA

Table 1 RSFC vector family designed for *P. pastoris*

#	Name	Tag/Fusion protein, position[a] and length[b]			TEV protease cleavage site	Mode of expression	EAEA repeat	Selection marker[c]	Plasmid size (bp)
1	PPpRSFC	–	–	–	NA	Intracellular	NA	Zeocin	4,840
2	pPpRSFC-MFalpha	–	–	–	NA	Secretory	Yes	Zeocin	5,104
3	pPpRSFC-MFalpha-noEAEA	–	–	–	NA	Secretory	No	Zeocin	5,092
4	pPpRSFC-pGAP	–	–	–	NA	Intracellular	NA	Zeocin	3,771
5	pPpRSFC-pGAP-MFalpha	–	–	–	NA	Secretory	Yes	Zeocin	4,035
6	pPpRSFC-pGAP-MFalpha-noEAEA	–	–	–	NA	Secretory	No	Zeocin	4,023
7	pPpRSFC-N-eGFP	eGFP	N	240	No	Intracellular	NA	Zeocin	5,584
8	pPpRSFC-C-eGFP	eGFP	C	239	No	Intracellular	NA	Zeocin	5,584
9	pPpRSFC-MFalpha-N-eGFP	eGFP	N	239	No	Secretory	Yes	Zeocin	5,848
10	pPpRSFC-MFalpha-C-eGFP	eGFP	C	239	No	Secretory	Yes	Zeocin	5,848
11	pPpRSFC-N-Myc	MYC	N	11	No	Intracellular	NA	Zeocin	4,870
12	pPpRSFC-C-Myc	MYC	C	10	No	Intracellular	NA	Zeocin	4,870
13	pPpRSFC-MFalpha-N-Myc	MYC	N	10	No	Secretory	Yes	Zeocin	5,134
14	pPpRSFC-MFalpha-C-Myc	MYC	C	10	No	Secretory	Yes	Zeocin	5,134
15	pPpRSFC-N-FLAG	FLAG	N	9	No	Intracellular	NA	Zeocin	4,864
16	pPpRSFC-C-FLAG	FLAG	C	8	No	Intracellular	NA	Zeocin	4,864
17	pPpRSFC-MFalpha-N-FLAG	FLAG	N	8	No	Secretory	Yes	Zeocin	5,128
18	pPpRSFC-MFalpha-C-FLAG	FLAG	C	8	No	Secretory	Yes	Zeocin	5,128
19	pPpRSFC-N-His-ncs	His	N	7	No	Intracellular	NA	Zeocin	4,858
20	pPpRSFC-C-His-ncs	His	C	6	No	Intracellular	NA	Zeocin	4,858
21	pPpRSFC-MFalpha-N-His-ncs	His	N	6	No	Secretory	Yes	Zeocin	5,122
22	pPpRSFC-MFalpha-C-His-ncs	His	C	6	No	Secretory	Yes	Zeocin	5,122
23	pPpRSFC-N-His	His	N	7	Yes	Intracellular	NA	Zeocin	4,879
24	pPpRSFC-C-His	His	C	6	Yes	Intracellular	NA	Zeocin	4,879
25	pPpRSFC-MFalpha-N-His	His	N	6	Yes	Secretory	Yes	Zeocin	5,143
26	pPpRSFC-MFalpha-C-His	His	C	6	Yes	Secretory	Yes	Zeocin	5,143
27	pPpRSFC-N-MBP	MBP	N	367	Yes	Intracellular	NA	Zeocin	5,959
28	pPpRSFC-C-MBP	MBP	C	366	Yes	Intracellular	NA	Zeocin	5,959
29	pPpRSFC-MFalpha-N-MBP	MBP	N	366	Yes	Secretory	Yes	Zeocin	6,223
30	pPpRSFC-MFalpha-C-MBP	MBP	C	366	Yes	Secretory	Yes	Zeocin	6,223
31	pPpRSFC-N-Strep	Strep	N	9	Yes	Intracellular	NA	Zeocin	4,885
32	pPpRSFC-C-Strep	Strep	C	8	Yes	Intracellular	NA	Zeocin	4,885
33	pPpRSFC-MFalpha-N-Strep	Strep	N	8	Yes	Secretory	Yes	Zeocin	5,149
34	pPpRSFC-MFalpha-C-Strep	Strep	C	8	Yes	Secretory	Yes	Zeocin	5,149
35	pPpRSFC-HIS	–	–	–	NA	Intracellular	NA	HIS4	7,683
36	pPpRSFC-HIS-MFalpha	–	–	–	NA	Secretory	Yes	HIS4	7,947
37	pPpRSFC-HIS-MFalpha-noEAEA	–	–	–	NA	Secretory	No	HIS4	7,935
38	pPpRSFC-HIS-pGAP	–	–	–	NA	Intracellular	NA	HIS4	6,614
39	pPpRSFC-HIS-pGAP-MFalpha	–	–	–	NA	Secretory	Yes	HIS4	6,878
40	pPpRSFC-HIS-pGAP-MFalpha-noEAEA	–	–	–	NA	Secretory	No	HIS4	6,866

NA not applicable, *ncs* no TEV protease cleavage site.

[a] N- or C- terminal fusion to the POI.

[b] Length in amino acids (intracellular N- terminal tags are because of the start codon one aa longer, the TEV protease cleavage site (seven aa) is not included in this number).

[c] Zeocin selection is applicable for *E. coli* and *P. pastoris*, HIS4 plasmids are based on ampicillin selection in *E. coli* and used in combination with a histidine auxotrophic (*his4*) *P. pastoris* strain.

sequence. Plasmids bearing tags always contain the EAEA repeat (Table 1).

Promoters, integration events and resistance markers

The pPpRSFC plasmids are based on the pPpT4 vector family reported by Näätsaari et al. [30] and also used as a platform for the *P. pastoris* Electra vectors by DNA2.0. The majority of pPpRSFC plasmids contain the promoter of the *alcohol oxidase 1* gene (p*AOX1*). This strong, tightly regulated methanol inducible promoter is most commonly used in *P. pastoris* [37]. We have also designed basic plasmids bearing the *glyceraldehyde-3-phosphate dehydrogenase* promoter (p*GAP*) to enable methanol free, constitutive expression (see Table 1).

In contrast to the yeast *S. cerevisiae*, where stable, autonomously replicating plasmids are available, circular plasmids bearing a yeast ARS (autonomously replicating sequence) are not stable in *P. pastoris* and genomic integration of plasmid cassettes is the method of choice for heterologous gene expression [27, 38]. Most commonly *P. pastoris* integration cassettes are created by linearizing plasmids or generation of linear cassettes by PCR [39, 40] and targeted to the *AOX1* locus via homologous sequences. Depending on the linearization site in the plasmid, different homologous recombination events can be targeted [38]. The pPpRSFC plasmids allow linearization to target gene replacement at the *AOX1* locus. Thereby the endogenous *AOX1* gene is deleted and the minor *AOX2* gene must take over the function of oxidizing methanol to formaldehyde. Due to the lower expression levels of *AOX2*, these *aox1* knockout strains show a Mut^S (methanol utilization slow) phenotype, which may result in higher yields than a Mut^+ phenotype [41, 42]. This can be achieved by linearization using *Bgl*II. If the *Bgl*II site is present in the insert, the vectors can still be linearized using the rare 8 bp *Swa*I sites as a failsafe backup. If a Mut^+ phenotype is desired, the vectors can be linearized using unique REs cleaving in the 5′ or 3′ homologous sequence (e.g. *Pme*I, *Ase*I or *Eco*NI, *Bsr*BI). However, due to low homologous recombination frequencies in *P. pastoris* wildtype strains [30], even when targeting a gene replacement at the *AOX1* locus (expected Mut^S phenotype), still the majority of transformants are Mut^+. Therefore it is necessary to validate the Mut phenotype by growth on methanol plates.

The RSFC plasmids are based on a modular design, the promoter, N- or C-terminal tags, terminator, the resistance marker and the 3′ homologous sequence can be easily exchanged using unique restriction enzyme sites (Figure 2). Most plasmids are based on Zeocin selection, however basic expression plasmids (#35–40) were also constructed with a histidine marker to be used with auxotrophic strains. The tagged expression cassettes from the Zeocin plasmids can be easily shuffled to the histidine plasmid backbones using unique *Pci*I and *Bam*HI sites.

In the pPpRSFC plamids not only the transition between the insert and the vector is seamless, also the transition of the promoter to the start codon and the stop codon to the terminator are seamless. In standard RE based cloning, the MCS may interfere with translation initiation [43] and this appears relevant for the *AOX1* promoter as extensions of the 5′ UTR (also caused by a MCS) were shown to negatively affect expression [44].

Applications of RSFC vectors to optimize HRP expression in *P. pastoris*

Effects of tags and fusions proteins in N- and C-terminal position

With the set of pPpRSFC plasmids available, we aimed to validate the system with a typical application. We tested expression of horseradish peroxidase (HRP), a commonly used reporter enzyme for signal amplification in diagnostic kits and immunohistochemistry. Secretory expression of HRP has been previously demonstrated in *P. pastoris* [42, 45–47]. Cytoplasmic expression promised little chance of success as HRP is a secretory plant peroxidase that requires formation of disulfide bridges and is typically glycosylated in the secretory pathway [48, 49]. Still, we tested the basic pPpRSFC plasmid (#1, untagged, p*AOX1*) for cytoplasmic HRP expression. This construct showed neither activity in the supernatant (Figure 3) nor in the cytoplasm (data not shown). Therefore different tags were only evaluated for secretory expression. A single PCR fragment of HRP was cloned into the vectors as outlined above. All p*AOX1* plasmids were linearized via *Bgl*II sites to target a gene replacement event at the *AOX1* locus, and screened for a Mut^S phenotype, which has been reported to be more favorable for HRP expression than Mut^+ [42].

The different tags and positions had diverse effects on volumetric HRP activities (Figure 3) and led to valuable insights. For all tags, the N-terminal version was giving higher activities than the C-terminal version. For the larger fusion proteins (eGFP and MBP), C-terminal tagging even led to almost complete loss of activity (#10 and #30). Comparing the tagged construct with the highest activity (#21) with the construct of the lowest activity (#10) gives a 31 fold difference. N-terminal tagging with the relatively large eGFP (and MBP) did not strongly affect activity, whereas shorter tags (Myc, FLAG, His, Strep) showed varying effects. The N-terminally His tagged construct with TEV protease cleavage site (#25) showed the lowest activity of all N-terminally tagged proteins. However, the N-terminal His tagged construct without TEV protease cleavage site (#21) showed activity similar to other tags, hinting a negative effect of the

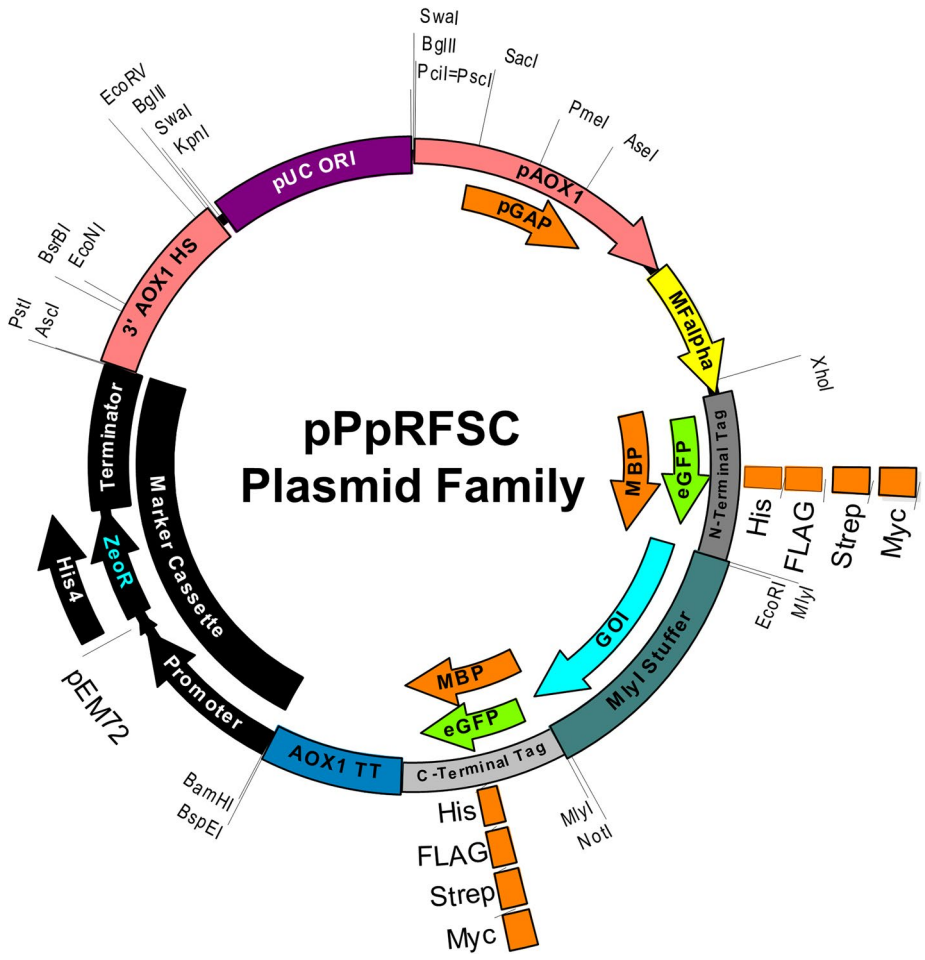


Figure 2 Representative map of *P. pastoris* RSFC plasmids. The features of all RSFC plasmids designed are summarized in this schematic map. Different promoters, N/C-terminal tags, resistance markers are shown. p*GAP* plasmids do not contain the 3′ *AOX1* homologous sequence for recombination (3′ AOX1 HS). HIS4 vectors contain in addition an ampicillin resistance cassette. The mating factor alpha signal sequence (MF alpha) is optional and only present in secretory plasmids. See Table 1 for the part combinations created in this study. Features are not drawn to scale. Exact plasmid maps are provided in the Additional file 3 in GenBank format.

TEV protease cleavage site in this context. Changes of the MFalpha sequence by removal of the EAEA sequence decreased activity 1.6 fold with the methanol inducible *AOX1* promoter (#2 vs. #3). With the constitutive *GAP* promoter (#5 vs. #6), removal of the EAEA sequence even led to a 17 fold decrease in activity. A possible mechanistic explanation would be that the EAEA repeats improved secretion due to increased Kex2 cleavage efficiencies [50, 51]. p*GAP* driven HRP expression was therefore, depending on the presence of EAEA repeats, competitive to the methanol inducible p*AOX1*. Due to shorter process times (not requiring methanol induction) p*GAP* driven expression may even be more favorable for HRP production regarding space time yields and process setup.

The effects seen on volumetric activities by using different tags may be caused by various reasons. On the one hand the tags may have interfered to a different extent with protein folding or access of the substrate to the active site thereby negatively affecting activity. On the other hand they also may have affected the protein yields by altering the protein stability, interfering with the secretion process or even on the mRNA level with transcript stability. Also the tags or linker sequence may have targeted proteolytic degradation, as previously reported for MBP in *P. pastoris* [33]. However, as we aimed only to evaluate the suitability of the RSFC strategy for screening different tags, we did not further investigate the underlying causes. The pPpRSFC plasmid family proved to be a simple tool to optimize volumetric activities of tagged HRP, showing that especially the tag positions and presence of EAEA repeats are crucial factors.

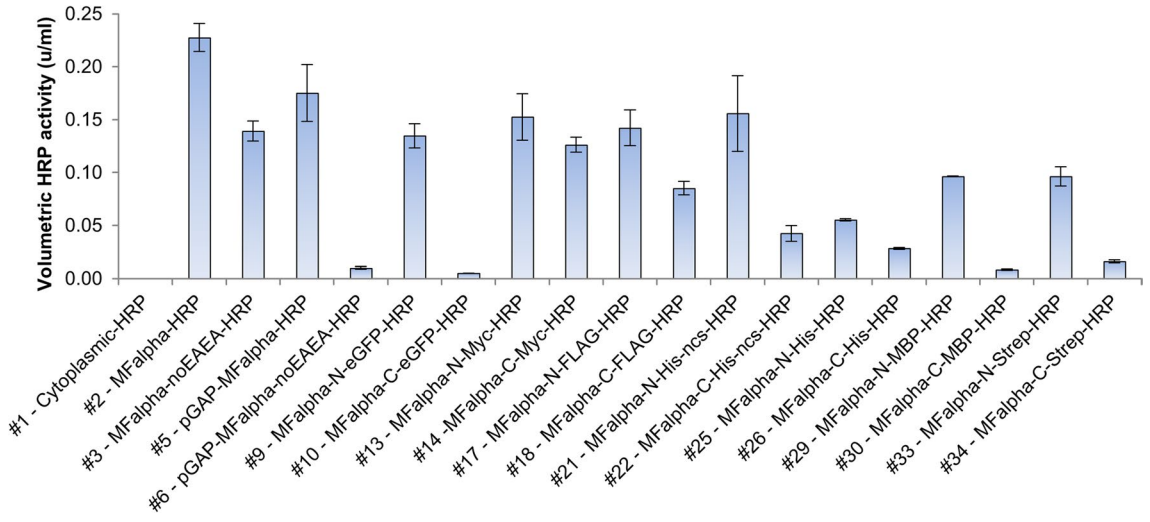


Figure 3 Type of tag and position (N/C-terminal) strongly affect volumetric HRP activities. Volumetric HRP activities in the supernatant with ABTS as substrate after cultivation on methanol for 72 h are shown. The pPpRSFC plasmids used are indicated via the *numbers* given in Table 1, the use of a signal sequence (MFalpha), different promoter than p*AOX1* and if applicable tag and position (N/C) are given. Mean values and standard deviations of biological triplicates are shown.

Fluorescence microscopy of strains expressing eGFP tagged HRP

GFP has routinely been used in *P. pastoris* as an intracellular reporter for comparing promoter activities [52–54] and to facilitate screening of protein production by testing fusions proteins [55], especially for membrane proteins [56–58]. Concerning GFP fusions of secretory proteins, conflicting results were obtained. In some cases GFP was successfully used as secretion reporter and for protein fusions [59–62]. In other cases problems with secretion (e.g. intracellular retention) were noticed [63–66]. As we had also designed N- and C-terminal fusions with eGFP (including the MFalpha signal sequence for secretion, #9 and #10), we performed fluorescence microscopy to investigate possible cellular retention and bottlenecks in the HRP secretion process.

The N-terminal eGFP-HRP fusion exhibited largely unchanged HRP activity, whereas the C-terminally tagged version had almost completely lost activity (Figure 3). We also included controls of intracellular eGFP expression (#1) and secretory eGFP alone (without an HRP fusion, created by self-ligating #9). Fluorescence microscopy images of methanol grown cells are shown in (Figure 4). While cytoplasmic expression showed bright fluorescence of the whole cell (Figure 4d), all secretory constructs (Figure 4a–c) showed punctate structures. These structures appeared somewhat similar to ER or Golgi mistargeting observed previously when expressing a GFP tagged membrane protein (human μ-opioid receptor, a G-protein coupled receptor) [65]. Most notably also the control of eGFP alone (Figure 4c), without an HRP fusion showed this retention. eGFP was apparently poorly secreted in *P. pastoris* and effects evoked by HRP may be masked and outweighed by the poor eGFP secretion.

We also measured eGFP fluorescence in the supernatant and the cell fraction (Additional file 4: Figure S3). Fluorescence in the supernatant could be detected for secretory constructs (Additional file 4: Figure S3a–c), while the cytoplasmic eGFP expression control (Additional file 4: Figure S3d) showed only marginal fluorescence in the supernatant. However, also for the secretory constructs (Additional file 4: Figure S3a–c) intracellular fluorescence surpassed fluorescence in the supernatant approximately 5- to 12-fold. These results suggested together with the microscopy images, that large amounts of eGFP were withheld in the secretion process. In this respect, eGFP fusion proteins may be used with caution when investigating secretory processes in *P. pastoris*. However, these effects may also be protein dependent, as there were cases reported where GFP was successfully used to evaluate signal sequences [59, 60] and some GFP fusion proteins were sufficiently secreted [61, 62].

Conclusions and outlook

The RSFC cloning strategy outlined here and the pPpRSFC plasmid family are simple tools to optimize expression of tagged proteins with little cloning efforts. RSFC requires at first the design and assembly of the vector family to be used. However, subsequent screening is

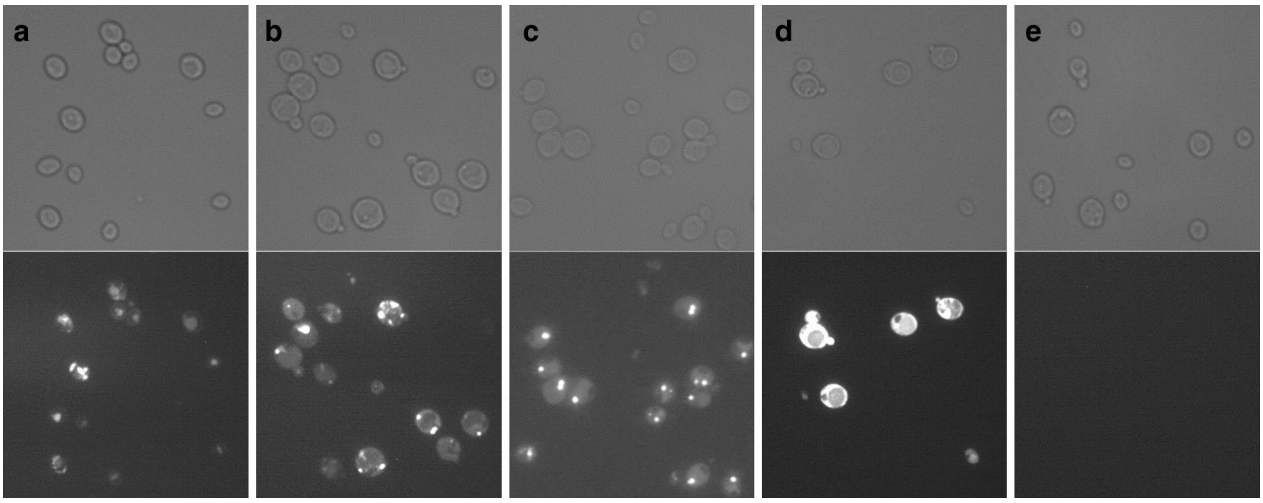


Figure 4 Fluorescence microscopy of fusions of HRP to GFP. Bright field images are shown on *top*, fluorescence images *below*. **a** HRP N-terminally tagged with eGFP (#9-MFalpha-N-eGFP-HRP), **b** HRP C-terminally tagged with eGFP (#10-MFalpha-C-eGFP-HRP), **c** control of eGFP with MFalpha (self-ligated #9), **d** control of cytoplasmic eGFP expression (#1-eGFP), **e** negative control of empty Mut^5 strain. For the bright field image of panel (**c**) brightness was decreased −11%, contrast was increased +44% for better comparability with the other panels.

drastically facilitated as large amounts of vector backbones can be prepared at once by *Mly*I digestion. Subsequently, the backbones ready for cloning can be frozen as aliquots and thawed when needed.

There have been systems reported previously that allow testing of the expression of a POI in different expression hosts by using only two PCR products [67]. This approach is based on ligation-independent cloning (LIC) similar to [10–13, 68]. While these methods allow highly efficient, seamless cloning, they rely on the annealing of single stranded overhangs, thereby requiring identical sequences between vector and insert. Therefore these methods are not suitable for seamless, sequence independent fusions possible with RSFC. However, as a downside of RSFC the blunt end ligations work less efficiently than annealing based in vitro recombination methods [10–13] and confirmation of the correct orientation is required. Otherwise only about 50% of the transformants show the desired orientation which is a disadvantage for library approaches. Nevertheless, after stuffer removal, inserts can also be cloned directionally into RSFC plasmids by in vitro recombination methods (such as Gibson assembly [12]). However this task requires the design of a separate primer for every tag and position to be tested as the overhang required for annealing changes between the vectors. We recommend using in vitro recombination methods with pPpRSFC plasmids when testing only a few constructs. When testing a larger number of constructs, the increased costs for primers and materials associated with in vitro recombination methods outweigh the costs for cPCRs to test the orientation of blunt end ligations. When performing a single experiment and cloning a low number of GOIs with only one tag, it will be more effort to set up a RSFC vector than to order a few long primers. However, if routinely a large number of GOIs should be screened with a set of tags in different positions, RSFC vectors are a powerful strategy.

A limitation of the RSFC system reported here is the use of *Mly*I, the only type IIS enzyme performing a blunt end cleavage. The recognition sequence of *Mly*I is five bp long (Figure 1a), thereby posing a problem as it appears statistically once per 512 bp ($4^5/2$) [69]. This may require frequent removal of *MlyI* sites in the vector backbones to be used. *Mly*I sites in CDSs of tags, fusion proteins and resistance markers can be easily removed by mutating the *Mly*I sequence to synonymous codons. However, mutating sites present in promoters, terminators or origins of replication have to be validated for unchanged functionality (or must be exchanged for parts free of *Mly*I sites). These issues could be solved by using artificial type IIS REs with longer recognition sequences. The catalytic domain of the archetypical type IIS enzyme *Fok*I has been fused to I-*Sce*I, a homing endonuclease with an 18 bp recognition sequence. This chimeric meganuclease showed sufficient cleavage resulting in 4 bp 'sticky' overhangs that could be ligated at 90% fidelity [69]. Following this strategy, the catalytic domain of *Mly*I (which is similar to *Fok*I [70]) could also be fused to I-*Sce*I. Statistically an 18 bp recognition sequence would appear approximately once in 10^{11} (4^{18}) bp, however I-*Sce*I recognizes also slightly degenerate sequences leading to an estimated appearance once in 10^8 bp [69, 71, 72], which

would still surpass the specificity of wildtype *Mly*I by several orders of magnitude.

Most vectors for *P. pastoris* have been conceptualized solely as straightforward expression vectors ([8, 22, 25, 30, 31] and *P. pastoris* plasmids by Life Technologies, BioGrammatics and DNA2.0) and few plasmid families allow to fine-tune expression [30, 31]. The 40 plasmids reported here extend the scope of applications and facilitate characterization and optimization of the production of heterologous proteins in *P. pastoris*. The RSFC strategy outlined here is not limited to tags and fusions proteins, but could also be applied to compare different promoters or signal sequences in other expression systems. Similarly, isoenzymes or families of homologous enzymes can be fused to tags to screen for better expression, solubility or other properties to identify enzymes combining desired biological, chemical and technological features.

Materials and methods

Chemicals and media

Phusion DNA Polymerase, restriction endonucleases and other DNA modifying enzymes were acquired from Thermo Fisher Scientific (Waltham, MA, USA) or New England Biolabs (Ipswich, MA, USA). Miscellaneous chemicals were purchased from Becton, Dickinson and Company (Franklin Lakes, NJ, USA), Carl Roth (Karlsruhe, Germany) and Fresenius Kabi Austria (Graz, Austria).

Plasmids were isolated using a GeneJET Plasmid Miniprep Kit by Thermo Fisher Scientific. The standard protocol was optimized for *Mly*I based constructs to compensate the decreased plasmid yields. A single colony of a strain bearing the respective plasmid was streaked on an agar plate containing the respective antibiotic. After incubation overnight, a cell pellet (approximately 0.1 g wet cells) was scratched of the plate and used for the isolations (final elution volume: 100 µl of ddH_2O).

Agarose embedded DNA, restriction digests and PCRs were purified using a Wizard SV Gel and PCR Clean-Up System by Promega.

P. pastoris strains were grown on full medium (yeast extract, peptone, 2% glucose, YPD), buffered minimal dextrose (BMD) and buffered minimal methanol medium with 0.5% methanol (BMM) as described by Weis et al. [16]. As only exception we used 2% glucose in the BMD medium and for HRP expression, media were supplemented with 1 mmol/l ferrous sulfate heptahydrate ($FeSO_4.7H_2O$). *Escherichia coli* strains were selected on LB-medium containing 25 µg/ml Zeocin™ (Life Technologies, Carlsbad, CA, USA). *P. pastoris* transformants were selected on YPD agar plates containing 100 µg/ml Zeocin. Primers were ordered from Integrated DNA Technologies (Leuven, Belgium), see Additional file 5: Table S1 for the sequences.

Plasmid construction

pombe RSFC test vectors pGAZ2-TA-BmrI-stuffer and pGAZ2-Blunt-MlyI-stuffer

For all cloning work an *E. coli* Top10 F' strain was used. The vectors for initially comparing blunt end and TA cloning were based on a replicative *S. pombe* vector pGAZ2 (Additional file 1: Figure S1, unpublished results). For the TA-cloning vector 'pGAZ2-TA-BmrI-stuffer', a stuffer fragment was amplified using primers TA_fwd_HindIII+BmrI+stuffer and TA_rev_BamHI+BmrI+stuffer (see Additional file 5: Table S1) and cloned into pGAZ2 via *Hind*III and *Bam*HI sites. The stuffer fragment was selected as a sequence that has no significant homology to *E. coli* and *S. pombe* genomes and lacks *Mly*I, *Bmr*I, *Hind*III and *Bam*HI RE sites; we used a part of a *P. pastoris* alpha, alpha trehalase gene. The 'pGAZ2-Blunt-MlyI-stuffer' vector required mutating two *Mly*I sites in the vector backbone. This was done by PCR amplifying the vector using primers pUC_mut_MlyI_fwd + pUC_mut_MlyI_rev and ZeoCDS_mut_MlyI_fwd + ZeoCDS_mut_MlyI_rev using Pfu Ultra polymerase (Agilent Technologies, Santa Clara, CA) followed by *Dpn*I digestion to remove template vector. The *Mly*I site in the pUC was mutated to the sequence reported by Rao et al. [25], the *Mly*I site in the zeocin resistance gene was mutated to a synonymous codon. After transformation, introduction of the correct mutations were confirmed by Sanger sequencing. Both plasmids do not provide seamless fusions, as the stuffer fragments were for convenience inserted via *Hind*III and *Bam*HI sites leaving RE site scars. For test purposes the gene coding for *Thermomyces lanuginosus* endo-beta-1,4-D-xylanase was amplified using primers Xyla_fwd and Xyla_rev and cloned into the two vectors (detailed protocol see below).

P. pastoris pPpRSFC plasmid family

The expression plasmids used in this study are based on the pPpT4 *P. pastoris*/*E. coli* shuttle vector family (e.g. GenBank accession number JQ519690.1) reported by Näätsaari et al. [30]. Two *Mly*I sites in the backbone (pUC and zeocin resticane gene) were mutated in the same way as in the *S. pombe* plasmids of this study (Additional file 1: Figure S1; same primers as in Additional file 5: Table S1) and confirmed by sequencing. The *AOX1* promoter, *Mly*I stuffer fragment and *AOX1* terminator were amplified in separate PCR reactions using primers PAOX1_PciIF/OePAox1StufferR, OestufferF/OeStufferR and OeAox1TTstufferF/Aox1TT_BamHIR respectively.

In the subsequent overlap extension PCR reactions the fragments were joined together using primer pair PAOX1_PciIF/AOX1TT_BamHIR followed by restriction with *Pci*I/*Bam*HI and were cloned in a vector backbone with mutated *Mly*I sites to create an intermediatory plasmid backbone termed 'pPp'. The stuffer fragment sequence was selected from as a sequence that has no homology to *E. coli* and *P. pastoris* and lacks unique RE used in the pPpRSFC plasmid family. We selected a part of a gene involved in the *S. cerevisiae* biotin metabolism. An *Eco*RI site in the stuffer was mutated using primers pairs OeEcoRIF and OeEcoRIR. There appeared a few additional mutations in the stuffer that had no functional consequences and where therefore left unchanged (see plasmid sequences in Additional file 3).

For constitutive plasmids, the *GAP* promoter was amplified via primers GAP_PciIF/OeGapStuffR and was cloned into the pPp backbone using PciI/EcoRI to create #4 (pPpRSFC-pGAP). The 3′ AOX1 homologous sequence was amplified via primers 3′AOX1_PstIASCIF/3′AOX1_KpnISwaIR and was cloned into pPp using *Kpn*I/*Pst*I restriction sites to create #1 (pPpRSFC). For secretory expression plasmids, the MFAlpha sequence was amplified using primer pair AlphaFSSF/AlphaEcoRIR (or aEAEAEcoRIR for insertion of Glu-Ala repeats). The *AOX1*/*GAP* promoters were amplified via primers PAOX1_PciIF + OeAlphaPAox1R/GAP_PciIF+ OeGapAlphaR. The MFAlpha sequence was fused with p*AOX1*/p*GAP* using primers PAOX1_PciIF+ AlphaEcoRIR (expression cassette for #3) or PAOX1_PciIF+ aEAEAEcoRIR (expression cassette for #2),/GAP_PciIF+ AlphaEcoRIR (expression cassette for #6) or GAP_PciIF+ aEAEAEcoRIR (expression cassette for #5). The pAOX1-MFAlpha PCR products were cloned into pPpRFSC via PciI/EcoRI sites to create #3 and #2. The pGap-MFAlpha fusion construct was cloned into pPpRSFC-pGAP via PciI/EcoRI restriction site to construct pPpRSFC-#6 and #5.

p*GAP* plasmids do not contain the 3′ *AOX1* sequence for homologous integration in the *AOX1* locus. Plasmids #1 to #6 were made initially and completely sequenced. In the creation of the following plasmids, only newly inserted parts (and the RE sites used for cloning) were confirmed by sequencing. A full description of how the 28 plasmids (#7–#40) with the N- and C- terminal tags were created would be beyond the scope of this section and is provided in the Additional file 5: Table S1 (spreadsheets on plasmid construction). For further details consult the annotated plasmid sequences provided in Additional file 3.

The HRP gene (isoenzyme A2A [46, 47]) was amplified using primers HRP-A2-RSFC-fwd and HRP-A2-RSFC-rev and cloned in the respective vectors (detailed protocol see below).

RSFC cloning of inserts and colony PCRs

For blunt end cloning, the vector backbone was dephosphorylated using either Thermo Scientific shrimp alkaline phosphatase or FastAP according to the manufacturer's recommendations. The backbone was gel purified and used for ligations with phosphorylated PCR products. Prior, PCR primers were phosphorylated using Thermo Scientific T4 Polynucleotide Kinase according to the manufacturer's recommendations. Subsequently the reaction mixtures containing the phosphorylated primers were desalted on nitrocellulose filters (Merck Millipore, Darmstadt, Germany) and added to the PCR (Phusion polymerase). Ligations were performed using the blunt end protocol provided for Thermo Scientific T4 DNA Ligase.

For TA cloning, phosphorylated Phusion PCR products were purified (Promega Wizard SV Gel and PCR Clean-Up System) and dA-tailed using Taq-Polymerase (GoTaq Flexi, Promega [Fitchburg, WI, USA], standard buffer, 0.2 mmol/l dATP, 30 min incubation at 72°C) and directly used for ligation (blunt end protocol provided for Thermo Scientific T4 DNA Ligase).

To verify the correct orientation by colony PCR, primers were selected as outlined in Additional file 2: Figure S2. A tiny amount of an *E. coli* colony from a transformation of the respective ligation was added to a GoTaq Flexi reaction. The manufacturer's protocol was followed except reducing the reaction volume to 20 µl and increasing the initial denaturing step to five min to break the cells.

P. pastoris *transformations and screening*

For testing the pPpRSFC plasmids, the *P. pastoris* CBS7435 wildtype strain was used. Plasmids bearing the *AOX1* promoter were linearized with *Bgl*II, plasmids with p*GAP* were linearized with *Swa*I. All linearized plasmids were transformed into competent *P. pastoris* cells prepared by the condensed protocol reported by Lin-Cereghino et al. [73]. Only low amounts of DNA (0.5 µg) were used for the transformations to avoid multicopy integration. A landscape of 80 clones was screened and checked for the desired Mut^S phenotype on minimal methanol plates. Ten Mut^S clones were rescreened for uniform expression; a single representative clone was used for the subsequent characterizations. Screenings, rescreening and characterizations were performed in deep well plates as described previously [74]. BMD 2% was used instead of BMD 1% (giving higher yields, data not shown) and the methanol induction was performed in 12 h intervals for 72 h.

HRP activity assay, eGFP fluorescence microscopy and measurements

HRP activity assays with 2,2′-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) diammonium salt (ABTS) were performed as described previously [42]. For intracellular HRP activity measurements, cells were broken using Yeast Protein Extraction Reagent (Y-PER from Thermo Scientific).

The cell suspensions of eGFP expressing strains were centrifuged and washed in an equal amount of water before fluorescence microscopy (Leica Microsystems, Germany, DM LB2, DFC350FX) at 1,000-fold magnification, fluorescence images were taken using filter set 'I3' [excitation filter BP 450–490]. eGFP fluorescence (ex/em 488/507 nm) and OD_{600} were measured in a Synergy MX plate reader (Biotek, Winooski, VT, USA) using micro titer plates (Nunc MicroWell 96-Well Optical-Bottom Plates with Polymer Base, Black; Thermo Fisher Scientific). Cell suspensions were diluted to be within the linear range. The background measurements of diluted medium were subtracted. Subsequently the relative fluorescence units were normalized per OD_{600}.

Additional files

Additional file 1: Figure S1. Vector maps of the ***S. pombe*** vectors used in this study

Additional file 2: Figure S2. Simple strategy for confirming the orientation of the insert. The forward or reverse primer used for amplifying the insert can be used together with the forward or reverse sequencing primer of the vector to confirm the correct orientation. Upon correct primer choice only the forward orientation gives a PCR fragment. The sequencing primers designed for Sanger sequencing allow sequencing of the insert from both sides. Depending on the vector, different primers should be used (e.g. when the MFalpha signal sequence or a fusion protein is present, see the primer list for all sequencing primers available).

Additional file 3: Plasmid sequences of the constructs used in this study in Genbank format.

Additional file 4: Figure S3. Fluorescence measurements of fusions of HRP to eGFP. Samples are labeled in the same way as in Figure 4. eGFP fluorescence of supernatants and cell pellets of methanol induced cells were normalized per cell density (OD_{600}).

Additional file 5: Table S1. Primers used in this study and detailed plasmid construction. Primers used for creating the ***S. pombe*** and ***P. pastoris*** plasmids are separated. Also primers for construction of the plasmids are separated from primers for sequencing and insertion of GOIs. In addition separate spreadsheets are providing information on the exact construction of the plasmids by listing the PCR products and restriction enzymes used for assembly.

Authors' contributions

TV, MA, FWK designed the experiments, analyzed and interpreted the data. MA, TV and FWK performed the experiments. TV and MA wrote the manuscript. TV, MA, FWK, HS and AG conceived of the study. AG and HS supervised the research. All authors read and approved the final manuscript.

Author details

[1] Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria. [2] Queensland University of Technology, 2 George St., Brisbane, QLD 4000, Australia.

Acknowledgements

We gratefully acknowledge the Austrian Science Fund (FWF) project number W901 (DK 'Molecular Enzymology' Graz), the Higher Education Commission of Pakistan and support from NAWI Graz. We would like to thank Harald Pichler and Christine M. Winkler for fruitful discussions regarding eGFP tagging experiments and Sarah Trunk for excellent technical assistance. We would also like to thank Rafael E. Luna for critical reading of the manuscript and helpful comments in frame of a scientific story telling workshop held by the DK 'Molecular Enzymology' Graz.

Compliance with ethical guidelines

Competing interests

The authors declare that they have no competing interests.

Received: 31 March 2015 Accepted: 30 June 2015
Published online: 14 July 2015

References

1. Malhotra A (2009) Tagging for protein expression. Methods Enzymol 463:239–258
2. Su WW (2005) Fluorescent proteins as tools to aid protein production. Microb Cell Fact 4:12
3. Arnau J, Lauritzen C, Petersen GE, Pedersen J (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr Purif 48:1–13
4. Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S (2005) Comparison of affinity tags for protein purification. Protein Expr Purif 41:98–105
5. Niiranen L, Espelid S, Karlsen CR, Mustonen M, Paulsen SM, Heikinheimo P et al (2007) Comparative expression study to increase the solubility of cold adapted Vibrio proteins in Escherichia coli. Protein Expr Purif 52:210–218
6. Gordon E, Horsefield R, Swarts HGP, de Pont JJHHM, Neutze R, Snijder A (2008) Effective high-throughput overproduction of membrane proteins in Escherichia coli. Protein Expr Purif 62:1–8
7. Mohanty AK, Wiener MC (2004) Membrane protein expression and production: effects of polyhistidine tag length and position. Protein Expr Purif 33:311–325
8. Sasagawa T, Matsui M, Kobayashi Y, Otagiri M, Moriya S, Sakamoto Y et al (2011) High-throughput recombinant gene expression systems in Pichia pastoris using newly developed plasmid vectors. Plasmid 65:65–69
9. Lu Q (2005) Seamless cloning and gene fusion. Trends Biotechnol 23:199–207
10. Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4:251–256
11. Zhang Y, Werling U, Edelmann W (2012) SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res 40:e55
12. Gibson DG, Young L, Chuang R, Venter JC, Hutchison CA, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345
13. Quan J, Tian J (2009) Circular polymerase extension cloning of complex gene libraries and pathways. PLoS ONE 4:e6441
14. Online manual to 'j5 automated DNA assembly'. http://j5.jbei.org/j5manual/pages/22.html. Accessed 26 Mar 2015
15. Hillson NJ, Rosengarten RD, Keasling JD (2012) j5 DNA assembly design automation software. ACS Synth Biol 1:14–21
16. Bath AJ, Milsom SE, Gormley NA, Halford SE (2002) Many type IIs restriction endonucleases interact with two recognition sites before cleaving DNA. J Biol Chem 277:4024–4033

17. Roberts RJ, Vincze T, Posfai J, Macelis D (2010) REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 38(Database issue):D234–D236
18. Padgett KA, Sorge JA (1996) Creating seamless junctions independent of restriction sites in PCR cloning. Gene 168:31–35
19. Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3:e3647
20. Engler C, Gruetzner R, Kandzia R, Marillonnet S (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS One 4:e5553
21. Mead DA, Pey NK, Herrnstadt C, Marcil RA, Smith LM (1991) A universal method for the direct cloning of PCR amplified nucleic acid. Biotechnology (N Y) 9:657–663
22. Lan D, Wang W, Sun L, Wang Y, Yang B (2011) A novel vector for expression cloning of large numbers of polymerase chain reaction products in *Pichia pastoris*. Afr J Microbiol Res 5:211–216
23. Ido E, Hayami M (1997) Construction of T-tailed vectors derived from a pUC plasmid: a rapid system for direct cloning of unmodified PCR products. Biosci Biotechnol Biochem 61:1766–1767
24. Reisinger C, Kern A, Fesko K, Schwab H (2007) An efficient plasmid vector for expression cloning of large numbers of PCR fragments in *Escherichia coli*. Appl Microbiol Biotechnol 77:241–244
25. Rao B, Zhong X, Wang Y, Wu Q, Jiang Z, Ma L (2010) Efficient vectors for expression cloning of large numbers of PCR fragments in *P. pastoris*. Yeast 27(5):285–292. doi:10.1002/yea.1753
26. Wu Q, Zhong X, Zhai C, Yang J, Chen X, Chen L et al (2010) A series of novel directional cloning and expression vectors for blunt-end ligation of PCR products. Biotechnol Lett 32:439–443
27. Cregg JM, Tolstorukov I, Kusari A, Sunga J, Madden K, Chappell T (2009) Expression in the yeast *Pichia pastoris*. Methods Enzymol 463:169–189
28. Gasser B, Prielhofer R, Marx H, Maurer M, Nocon J, Steiger M et al (2013) *Pichia pastoris*: protein production host and model organism for biomedical research. Futur Microbiol 8:191–208
29. Vogl T, Hartner FS, Glieder A (2013) New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*. Curr Opin Biotechnol 24:1094–1101
30. Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A (2012) Deletion of the *Pichia pastoris* KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One 7:e39720
31. Sears IB, O'Connor J, Rossanese OW, Glick BS (1998) A versatile set of vectors for constitutive and regulated gene expression in *Pichia pastoris*. Yeast 14:783–790
32. Goulas T, Cuppari A, Garcia-Castellanos R, Snipas S, Glockshuber R, Arolas JL et al (2014) The pCri system: a vector collection for recombinant protein expression and purification. PLoS One 9:e112643
33. Li Z, Leung W, Yon A, Nguyen J, Perez VC, Vu J et al (2010) Secretion and proteolysis of heterologous proteins fused to the *Escherichia coli* maltose binding protein in *Pichia pastoris*. Protein Expr Purif 72:113–124
34. Julius D, Brake A, Blair L, Kunisawa R, Thorner J (1984) Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-alpha-factor. Cell 37:1075–1089
35. Briand L, Perez V, Huet JC, Danty E, Masson C, Pernollet JC (1999) Optimization of the production of a honeybee odorant-binding protein by *Pichia pastoris*. Protein Expr Purif 15:362–369
36. Treerattrakool S, Eurwilaichitr L, Udomkit A, Panyim S (2002) Secretion of Pem-CMG, a peptide in the CHH/MIH/GIH family of *Penaeus monodon*, in *Pichia pastoris* is directed by secretion signal of the alpha-mating factor from *Saccharomyces cerevisiae*. J Biochem Mol Biol 35:476–481
37. Vogl T, Glieder A (2013) Regulation of *Pichia pastoris* promoters and its consequences for protein production. New Biotechnol 30:385–404
38. Daly R, Hearn MTW (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. J Mol Recognit 18:119–138
39. Liu Z, Pscheidt B, Avi M, Gaisberger R, Hartner FS, Schuster C et al (2008) Laboratory evolved biocatalysts for stereoselective syntheses of substituted benzaldehyde cyanohydrins. ChemBioChem 9:58–61
40. Pscheidt B, Liu Z, Gaisberger R, Avi M, Skranc W, Gruber K et al (2008) Efficient Biocatalytic Synthesis of (R)-Pantolactone. Adv Synth Catal 350:1943–1948
41. Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the *Pichia pastoris* expression system. Yeast 22:249–270
42. Krainer FW, Dietzsch C, Hajek T, Herwig C, Spadiut O, Glieder A (2012) Recombinant protein expression in *Pichia pastoris* strains with an engineered methanol utilization pathway. Microb Cell Fact 11:22
43. Crook NC, Freeman ES, Alper HS (2011) Re-engineering multicloning sites for function and convenience. Nucleic Acids Res 39:e92
44. Staley CA, Huang A, Nattestad M, Oshiro KT, Ray LE, Mulye T et al (2012) Analysis of the 5′ untranslated region (5′UTR) of the alcohol oxidase 1 (AOX1) gene in recombinant protein expression in Pichia pastoris. Gene 496:118–127
45. Morawski B, Lin Z, Cirino P, Joo H, Bandara G, Arnold FH (2000) Functional expression of horseradish peroxidase in *Saccharomyces cerevisiae* and *Pichia pastoris*. Protein Eng 13:377–384
46. Krainer FW, Pletzenauer R, Rossetti L, Herwig C, Glieder A, Spadiut O (2014) Purification and basic biochemical characterization of 19 recombinant plant peroxidase isoenzymes produced in *Pichia pastoris*. Protein Expr Purif 95:104–112
47. Näätsaari L, Krainer FW, Schubert M, Glieder A, Thallinger GG (2014) Peroxidase gene discovery from the horseradish transcriptome. BMC Genom 15:227
48. Veitch NC (2004) Horseradish peroxidase : a modern view of a classic enzyme. Phytochemistry 65:249–259
49. Lin-Cereghino GP, Stark CM, Kim D, Chang JWJ, Shaheen N, Poerwanto H et al (2013) The effect of α-mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*. Gene 519:311–317
50. Kjeldsen T (2000) Yeast secretory expression of insulin precursors. Appl Microbiol Biotechnol 54:277–286
51. Thim L, Hansen MT, Norris K, Hoegh I, Boel E, Forstrom J et al (1986) Secretion and processing of insulin precursors in yeast. Proc Natl Acad Sci USA 83:6766–6770
52. Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP et al (2008) Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. Nucleic Acids Res 36:e76
53. Xuan Y, Zhou X, Zhang W, Zhang X, Song Z, Zhang Y (2009) An upstream activation sequence controls the expression of AOX1 gene in *Pichia pastoris*. FEMS Yeast Res 9:1271–1282
54. Vogl T, Ruth C, Pitzer J, Kickenweiz T, Glieder A (2014) Synthetic core promoters for *Pichia pastoris*. ACS Synth Biol 3:188–191
55. Al-Samarrai TH, Kirk CA, Jones WT, Harvey D, Templeton MD (2010) A simple and rapid method for selecting high producers of recombinant proteins in individual clones of *P. pastoris*. Comb Chem High Throughput Screen 13:377–382
56. Brooks CL, Morrison M, Lemieux JM (2013) Rapid expression screening of eukaryotic membrane proteins in Pichia pastoris. Protein Sci 22:425–433
57. Mizutani K, Yoshioka S, Mizutani Y, Iwata S, Mikami B (2011) High-throughput construction of expression system using yeast Pichia pastoris, and its application to membrane proteins. Protein Expr Purif 77:1–8
58. Vogl T, Thallinger GG, Zellnig G, Drew D, Cregg JM, Glieder A et al (2014) Towards improved membrane protein production in Pichia pastoris: General and specific transcriptional response to membrane protein overexpression. N Biotechnol 31:538–552
59. Eiden-Plach A, Zagorc T, Heintel T, Carius Y, Breinig F, Schmitt MJ (2004) Viral preprotoxin signal sequence allows efficient secretion of green fluorescent protein by *Candida glabrata*, *Pichia pastoris*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*. Appl Environ Microbiol 70:961–966
60. Kottmeier K, Ostermann K, Bley T, Rödel G (2011) Hydrophobin signal sequence mediates efficient secretion of recombinant proteins in *Pichia pastoris*. Appl Microbiol Biotechnol 91:133–141
61. Tang J-B, Zhu P, Yang H, Sun L-M, Song S, Ji A (2008) Expression and secretion of recombinant ZZ-EGFP fusion protein by the methylotrophic yeast *Pichia pastoris*. Biotechnol Lett 30:1409–1414
62. Panjideh H, Coelho V, Dernedde J, Fuchs H, Keilholz U, Thiel E et al (2008) Production of bifunctional single-chain antibody-based fusion proteins in *Pichia pastoris* supernatants. Bioprocess Biosyst Eng 31:559–568
63. Cha HJ, Dalal NN, Bentley WE (2005) Secretion of human interleukin-2 fused with green fluorescent protein in recombinant *Pichia pastoris*. Appl Biochem Biotechnol 126:1–11

64. Passolunghi S, Brocca S, Cannizzaro L, Porro D, Lotti M (2003) Monitoring the transport of recombinant Candida rugosa lipase by a green fluorescent protein-lipase fusion. Biotechnol Lett 25:1945–1948
65. Sarramegna V, Talmont F, Seree de Roch M, Milon A, Demange P (2002) Green fluorescent protein as a reporter of human mu-opioid receptor overexpression and localization in the methylotrophic yeast Pichia pastoris. J Biotechnol 99:23–39
66. Surribas A, Resina D, Ferrer P, Valero F (2007) Rivoflavin may interfere with on-line monitoring of secreted green fluorescence protein fusion proteins in Pichia pastoris. Microb Cell Fact 6:15
67. Dortay H, Akula UM, Westphal C, Sittig M, Mueller-Roeber B (2011) High-throughput protein expression using a combination of ligation-independent cloning (LIC) and infrared fluorescent protein (IFP) detection. PLoS One 6:e18900
68. De Jong RN, Daniëls MA, Kaptein R, Folkers GE (2006) Enzyme free cloning for high throughput gene cloning and expression. J Struct Funct Genomics 7:109–118
69. Lippow SM, Aha PM, Parker MH, Blake WJ, Baynes BM, Lipovsek D (2009) Creation of a type IIS restriction endonuclease with a long recognition sequence. Nucleic Acids Res 37:3061–3073
70. Besnier CE, Kong H (2001) Converting MlyI endonuclease into a nicking enzyme by changing its oligomerization state. EMBO Rep 2:782–786
71. Scalley-Kim M, McConnell-Smith A, Stoddard BL (2007) Coevolution of a homing endonuclease and its host target sequence. J Mol Biol 372:1305–1319
72. Doyon JB, Pattanayak V, Meyer CB, Liu DR (2006) Directed evolution and substrate specificity profile of homing endonuclease I-SceI. J Am Chem Soc 128:2477–2484
73. Lin-Cereghino J, Wong WW, Xiong S, Giang W, Luong LT, Vu J et al (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. Biotechniques 38:44, 46, 48
74. Weis R, Luiten R, Skranc W, Schwab H, Wubbolts M, Glieder A (2004) Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena. FEMS Yeast Res 5:179–189

Chapter 6

RANDOM MUTAGENESIS OF *PICHIA PASTORIS* IDENTIFIES EFFECTORS OF RECOMBINANT PROTEIN SECRETION

Christine M. Winkler[1], **Mudassar Ahmad[1]**, Peter Neugebauer[1], Helmut Schwab[1,2] and Harald Pichler[1,2*]

[1]*Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria*

[2]*Austrian Center of Industrial Biotechnology (ACIB), Petersgasse 14, 8010 Graz, Austria*

Contribution:

CW (80%), and MA (Approximately 15%) designed the experiments, analyzed and interpreted the data. CW (72%), MA (10%), PN performed the experiments. CW, HP wrote the manuscript. HS and HP supervised the study. All Authors read and approved the final manuscript.

I was mainly involved in devising the strategy to reproduce results in CBS 7435 strain. For this, I designed and constructed different expression vectors. I also generated and characterized 50% of the clean gene knockouts in CBS 7435 Δ*his* strain background for this study.

RANDOM MUTAGENESIS OF *PICHIA PASTORIS* IDENTIFIES EFFECTORS OF RECOMBINANT PROTEIN SECRETION

Christine M. Winkler[1], **Mudassar Ahmad[1]**, Peter Neugebauer[1], Helmut Schwab[1,2] and Harald Pichler[1,2*]

[1]*Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria*

[2]*Austrian Center of Industrial Biotechnology (ACIB), Petersgasse 14, 8010 Graz, Austria*

Corresponding author:

[*]Harald Pichler

Institute of Molecular Biotechnology

Graz University of Technology

Petersgasse 14/2

8010 Graz

Austria

Tel.: +43 (316) 873 4089

Fax: +43 (316) 873 4071

E-mail: harald.pichler@tugraz.at

Keywords

Yeast, *Pichia pastoris*, protein secretion, random mutagenesis, horseradish peroxidase

Abstract

Previous attempts to improve recombinant protein secretion in the popular expression host *Pichia pastoris* were traditionally guided by knowledge available from the model yeast *Saccharomyces cerevisiae*. Obviously, this approach does not allow for the identification of secretion enhancing factors that are unique to *P. pastoris*. We have developed a novel insertion mutagenesis method for the unbiased targeting of genes in this yeast. Applying this method and performing medium throughput screening for enhanced horseradish peroxidase (HRP) secretion identified a set of genes that had not been associated with recombinant protein secretion before. Here, we show that the targeted deletion of *P. pastoris* genes *RIM101*, *SGT2* and *KEP1* increases the amount of HRP secreted to the culture medium. Furthermore, disruption of the previously undescribed gene *KEP1* (knockout enhances protein secretion) causes elevated levels of a yet unidentified endogenous protein in the culture supernatant. This phenotype was reversed by restoring *KEP1* expression in the respective knockout strain. Most interestingly, deletion of *KEP1* also benefits secretion of recombinant proteins structurally unrelated to HRP, including alternative pig liver esterase (APLE) and human growth hormone (hGH). Based on these results, we promote the knockout strain *kep1*Δ as a valuable tool in the industrial production of recombinant proteins.

Introduction

The methylotrophic yeast *P. pastoris* is widely used as a production host for recombinant protein. *P. pastoris* offers easy genetic manipulation, growth to high cell densities, availability of strong and regulable promoters and permits post-translational modifications of expressed proteins. Targeting the recombinant product for secretion to the culture medium is a popular strategy. Secretion circumvents the accumulation of potentially toxic proteins intracellularly. Moreover, proteins entering the secretory pathway are folded and may receive post-translational modifications in ER and Golgi apparatus. These modifications are prerequisites for many proteins of plant and mammalian origin to obtain their biologically active form. A further advantage of protein secretion to the culture supernatant is that the product is thereby purified from most intracellular proteins, simplifying downstream purification.

However, secretion from the cell still represents one of the limiting factors in high-yield production for many recombinant proteins. Reported yield-limiting steps in the secretory pathway include translocation of the nascent protein to the ER (Koganesawa et al. 2001; Pfeffer et al. 2012), folding and processing in ER and Golgi compartments (Kowalski et al. 1998; Whyteside et al. 2011), and, finally, passage through the cell wall (Marx et al. 2006; Larsen et al. 2013). Engineering of these bottlenecks has proven to be extremely challenging, as the factors involved underlie complex interactions. A common strategy to improve recombinant protein secretion has been to manipulate the organism based on knowledge obtained from the model yeast *S. cerevisiae*. Different examples for this strategy were reviewed by Damasceno et al. (2011) and Idiris et al. (2010). They discuss the co-overexpression of ER-resident chaperones, e.g. BiP/Kar2 and PDI, and other proteins like the ER-associated degradation (ERAD) inducing transcription factor Hac1 in *P. pastoris* and related yeasts. In most of the reviewed cases, the co-overexpression of secretion helpers with a

recombinant protein enhanced secretion titers of only a subset of the tested proteins. This finding might be explained by the additional burden that is imposed on the expression host by overexpressing two proteins at the same time. Another drawback of the knowledge-driven strain improvement is that it does not take physiological and regulatory differences between the two yeast species *P. pastoris* and *S. cerevisiae* into account. Different "–omics" based approaches have pointed to a set of previously known and several novel genes involved in recombinant protein secretion (Gasser et al. 2007; Stadlmayr et al. 2010; Baumann et al. 2011; Pfeffer et al. 2012). The large amount of data generated by these methods requires subsequent evaluation of the observed effects by genetic engineering techniques. The decision on whether to further investigate a distinct gene or not, is, again, often guided by expertise obtained from *S. cerevisiae*.

Random mutagenesis represents a powerful tool to find yet unidentified effectors of secretion. It involves the random inactivation of genes followed by screening for altered secretion levels of the recombinant product. The impaired gene can easily be identified by genome walking when integrative DNA cassettes are used for mutagenesis. We found that a linear DNA marker cassette having no sequences homologous to the *P. pastoris* genome sequence efficiently and randomly integrated into the genome upon transformation. A comparable approach had been used by van Dijk et al. (2001) for the insertional mutagenesis of *Ogataea angusta* (aka *Pichia angusta* or *Hansenula polymorpha*). Schroder et al. (2007) and Larsen et al. (2013) achieved the random integration of DNA fragments into the genome of *P. pastoris* by performing REMI (restriction enzyme mediated insertion) in which both the genomic DNA and the mutagenesis cassette are cut with restriction enzymes to increase insertion frequencies. In this study, we performed random mutagenesis of *P. pastoris* cells secreting HRP as a reporter protein. We subsequently screened for gene deletions either increasing or decreasing the amount of HRP secreted to the culture medium.

Our strategy allowed us to pinpoint a number of highly interesting genome loci of which many had not been associated with recombinant protein secretion before. We could show that deleting components of the Rim-pathway (Rim101, Rim20, Rim13), responsible for the reaction to alkaline pH in *S. cerevisiae*, and Sgt2, a member of the GET-complex, benefits the secretion of HRP. Most interestingly, we identified an up to now uncharacterized protein, which we termed Kep1 (knockout enhances protein secretion 1), to be an effector of recombinant protein secretion. Aside from enhancing HRP secretion, deletion of *KEP1* is equally beneficial for the secretion of two other structurally and functionally unrelated recombinant proteins, alternative pig liver esterase (APLE) and human growth hormone (hGH). Thus, *kep1*Δ host strains may promote enhanced secretion levels of a plenitude of recombinant proteins. Moreover, we found that the *kep1*Δ knockout strain releases an endogenous protein to the culture supernatant to a higher extent than the wild type. This phenotype could be reversed by restoring expression of *KEP1* in the respective knockout strains. Similarly, enhanced HRP secretion in the three knockout strains *kep1*Δ, *rim101*Δ and *sgt2*Δ receded to wild type levels upon expression of the previously knocked out genes.

Materials and methods

All *P. pastoris* strains constructed during this study are described in more detail in Suppl. table 1. An overview of all employed primers is given in Suppl. table 2. Unless otherwise mentioned, we used standard cloning techniques to construct the plasmids listed in Suppl. table 3.

Strains and culture conditions

Escherichia coli TOP10F' cells (Life Technologies, Carlsbad, CA) were used for cloning experiments and propagation of expression vectors. We used the *P. pastoris* strain GS115 (Life Technologies) for mutagenesis and initial screening, and the strain CBS7435 *his4*Δ (Näätsaari et al. 2012) was employed for all further experiments. Transformants were selected on YPD-Zeocin™ (1 % yeast extract, 2 % peptone, 2 % glucose, 2 % agar, 100 mg/l Zeocin™ (Invivogen-Eubio, Vienna, Austria)) or on minimal dextrose plates (1.34 % yeast nitrogen base (YNB), 4×10^{-5} % biotin, 2 % glucose, and 2 % agar). Unless otherwise stated, cells were grown in 96-well deep-well plates (Bel-Art Scienceware, Pequannock, NJ) at 28°C, 320 rpm and 80 % humidity. For the initial screening experiments, cells were cultured on buffered YPD (1 % yeast extract, 2 % peptone, 2 % glucose, 0.2 M potassium phosphate buffer, pH 6.0), for 72 h before harvest. Later, the time for expression from P_{GAP} was shortened to 35 h. In case of expression from P_{AOX1}, cells were pre-grown in $BMG_{lucose}Y$ (1 % yeast extract, 2 % peptone, 2 % glucose, 0.2 M potassium phosphate buffer, pH 6.0, 1.34 % YNB, 4×10^{-5} % biotin) for 32 h, followed by 48 h of induction with BMMY medium containing 1 % methanol instead of glucose. Deep well plates were spun at 1,500 x g for 10 min at 22°C to harvest cells.

The growth curves of wild type control and knockout strains were recorded in 300 mL baffled shake flasks, incubated at 28°C, 140 rpm, 80 % humidity. Each strain was cultivated in triplicate.

We determined OD_{600} of the culture and HRP activity in the culture supernatants following centrifugation at 3,000 x g for 6 min at multiple time points during 118 h of cultivation time.

Random mutagenesis

The starting strain for mutagenesis was constructed by transforming *P. pastoris* GS115 cells with the plasmid pPIC9toGAPαHRP (Suppl. Fig. 1). This modification of pPIC9 (Life Technologies) had a fragment of P_{GAP} and α-mating factor signal sequence, both originating from vector pGAPZαA (Life Technologies), and the HRP-C1A coding sequence (Genbank accession number HE963800.1) ligated between restriction sites *Bgl*II and *Not*I, thereby replacing P_{AOX1}. As a mutagenic cassette, the Zeocin™ resistance cassette was amplified from the vector pGAPZαA by PCR (Fig. 1). The mutagenic cassette was transformed into competent cells of the mutagenesis starting strain by electroporation. Transformation following the condensed protocol was essentially done as described (Lin-Cereghino et al. 2005). In each transformation reaction, 1-2 µg of purified PCR product (Promega Wizard SV Gel and PCR Clean-Up System, Promega, Madison,WI) was transformed into 80 µL of competent cells. We observed a transformation rate of ~200 CFU per microgram of transformed PCR product. Transformants were selected on YPD plates supplemented with Zeocin™ (100 µg/mL).

Identification of integration loci

The protocol for the isolation of genomic DNA was adapted from Hoffman and Winston (1987). The insertion locus of the mutagenesis cassette was determined by template-blocking PCR, as described in Bae and Sohn (2010). Flanking genomic regions amplified by PCR were sequenced by LGC Genomics GmbH (Berlin, Germany). The obtained sequences were identified by performing a BLAST Nucleotide Sequence Similarity Search (Altschul et al. 1997) against the

genome sequences of *P. pastoris* CBS7435 and GS115. The amino acid sequence corresponding to the affected ORF was compared to the proteome of *S. cerevisiae* and other related yeasts using NCBI Protein BLAST.

Targeted gene knockouts

Gene loci were targeted following the protocol of Ahmad et al. (manuscript in preparation) for gene knockout and subsequent marker recycling. The respective 3'- and 5'- homology regions of 0.7 – 1.1 kb in length were cloned into pPpKC1. Knockout cassettes contained the site-specific Flp recombinase and a Zeocin^{TM} selection marker placed between two recombination target sequences (FRT). Gene disruptions were verified by PCR with primers binding in the disruption cassette in combination with primers binding either up- or downstream of the targeted locus. Selection marker and other vector elements were excised by Flp-mediated recombination at the FRT sites flanking these sequences after shifting the cells to methanol as described in Ahmad et al. (manuscript in preparation). The successful excision of the marker was verified by counter-selection on medium containing antibiotic and by PCR using primers binding up- and downstream of the targeted locus.

Genes coding for the reporter proteins (HRP, APLE and hGH) were cloned into the plasmids pAaHSwa (Suppl. Fig. 2) and pGaHSwa (Suppl. Fig. 3), respectively, using the restriction sites *Xho*I and *Not*I. The vector pAaHSwa was assembled from the sequences of Ampicillin resistance gene and *E. coli* origin of replication, both originating from plasmid pUC8 (Genbank accession number L08959). The other parts of the vector, namely *AOX1* promoter, *AOX1* terminator, 3'-homology region of *AOX1*, *ARG4* promoter, *ARG4* terminator and *HIS4* ORF were amplified from *P. pastoris* CBS7435. The sequence of the *S. cerevisiae* α-mating factor was obtained as a synthetic DNA fragment. The vector pGaHSwa was constructed by inserting P_{GAP}, amplified from vector

pPpB1GAP (Näätsaari et al. 2012) into the vector backbone of pAaHSwa by overlap-extension PCR. Prior to transformation, the plasmids were linearized with *Swa*I, to target them to the *AOX1* locus. Replacement events at the *AOX1* locus were confirmed by Mut^S screening on minimal methanol plates (1.34 % YNB, 4×10^{-5} % biotin, 0.5 % methanol, 2 % agar) and PCR (Ahmad et al., manuscript in preparation).

Enzyme- and immunoassays

Enzymatic activity in culture supernatants was quantified with colorimetric assays, as described for HRP in Morawski et al. (2000) and APLE in Hermann et al. (2008). Absorption at 405 nm was quantified with a Biotek Gen5 spectrophotometer (Biotek, Winooski, VT) at room temperature, followed by normalization for optical density of the cultures. In the initial screenings, color development was compared between the mutant strains by eye.

To quantify the amount of secreted product by immunoblotting, aliquots of culture supernatants were mixed with SDS-sample buffer (Life Technologies), and heated to 40°C (APLE) or 75°C (HRP and hGH) for 15 min. For deglycosylation of HRP, samples were treated with EndoH (New England Biolabs, Beverly, MA) following the manufacturer's instructions, before adding sample buffer. Proteins were resolved on NuPAGE 4-12 % Bis-Tris gels (Life Technologies) and blotted onto Amersham Hybond ECL nitrocellulose blotting membranes (GE Healthcare, Buckinghamshire, UK) using Life Technologies's XCell II blot module according to the manufacturer's instructions.

Immunoblotting analysis was performed following published procedures (Haid and Suissa 1983). HRP and APLE were detected using rabbit polyclonal antibodies raised against HRP and porcine liver esterase, respectively (both Abcam, Cambridge, UK). Goat anti-rabbit polyclonal conjugated

with HRP was used as secondary antibody. Western blot detection was done with the SuperSignal West Pico Kit (Thermo Scientific, Waltham, MA). Primary anti-hGH antibody produced in goat and secondary donkey-anti-goat antibody were purchased from Santa Cruz Biotechnology (Santa Cruz, CA). The alkaline phosphatase-conjugated secondary antibody was detected with BCIP/NBT (Thermo Scientific).

Additional protocols for characterization of knockout strains

For the Calcofluor white plate assay, dilutions of cells between OD_{600}=0.05 and 5 x 10^{-7} were spotted onto YPD agar plates containing 10 µg/mL of Calcofluor white (Sigma Aldrich, St. Louis, WO). Plates were incubated for 3-4 days at 30°C. The alkaline phosphatase assay was performed as described in Larsen et al. (2013).

Complementation with FLAG-tagged proteins

For expression of the genes from their native promoter, the open reading frames and upstream regions of *P. pastoris RIM101* (NCBI GeneID: PP7435_Chr3-0578), *SGT2* (PP7435_Chr1-0883) and *KEP1* (PP7435_Chr4-0066) were amplified from *P. pastoris* CBS7435. The amplified fragments were fused to a C-terminal FLAG-tag, Zeocin™ resistance cassette and *E. coli* origin of replication by Gibson assembly (Gibson et al. 2009). As a template for the above mentioned plasmid backbone fragments we used plasmid pPpT4 (Näätsaari et al. 2012), with a FLAG-tag-fused *AOX1* terminator previously inserted by *Not*I and *Bam*HI restriction enzyme cloning. The plasmids obtained after Gibson assembly were linearized with restriction enzymes cutting upstream of the genes (189 bp upstream the *RIM101* start codon, 252 bp upstream of *SGT2*, and 448 bp upstream of the *KEP1* start codon) prior to transformation into the respective knockout strains, already expressing secreted HRP from P_{GAP}. The strains were cultured for 35 h in deep-well plates

and were subjected to HRP secretion analysis employing the ABTS assay as described above. For SDS-PAGE analysis of proteins in the culture supernatant, cultivation time was prolonged to 72 h. Proteins in 244 µL culture supernatant were precipitated in 33 % trichloroacetic acid (TCA). Pellets were washed twice with ice-cold acetone and resuspended in sample buffer and 200 mM Tris-HCl buffer, pH 7.4, followed by heating to 75°C for 15 min. Proteins were resolved on NuPAGE 4-12 % Bis-Tris gels and visualized with Coomassie blue. For the detection of intracellularly expressed FLAG-tagged proteins, cells were disrupted and proteins precipitated with TCA as described in Horvath and Riezman (1994). SDS-PAGE, Western blotting and signal detection were performed as described above. For immunodetection we used monoclonal anti-FLAG M2 antibody, produced in mouse, and HRP-conjugated anti-mouse antibody produced in goat (both Sigma-Aldrich).

Results

Mutant library construction

Many researchers working with *P. pastoris* have struggled with this yeast's preference for non-homologous recombination when they tried to selectively target DNA fragments to a specific locus in the genome (Näätsaari et al. 2012). Our random mutagenesis protocol described here, in fact, made use of the efficient and random integration of non-homologous DNA fragments into the genome sequence of *P. pastoris*. We employed the Zeocin^{TM} resistance cassette of Life Technologies' pGAPZ plasmid for random integration (Fig. 1). The 1172 bp-comprising DNA cassette consists of the *TEF1* promoter from *S. cerevisiae* (GenBank accession numbers D12478, D01130), the synthetic prokaryotic promoter *EM7*, the *She ble* ORF from *Streptoalloteichus hindustanus* that confers resistance to the antibiotic family of phleomycins (Gatignol et al. 1987; Drocourt et al. 1990; Calmels et al. 1991) and the *S. cerevisiae CYC1* transcription termination region (GenBank accession number M34014). We could not detect any significant homology of the cassette to the *P. pastoris* genome sequence using NCBI Nucleotide BLAST. Therefore, biased integration events should occur with negligible frequency. To generate the starting strain for mutagenesis, we transformed *P. pastoris* GS115 cells with the plasmid pPIC9toGAPαHRP (Suppl. Fig. 1 and described in the Materials and methods section) for secretory expression of HRP from P_{GAP}. The mutagenesis cassette was amplified from the plasmid pGAPZ by PCR and the purified product was used to transform the HRP-secreting strain. After several rounds of transformation, we obtained approximately 3000 transformants that were subsequently screened for altered levels of HRP activity in the culture supernatants.

Library screening for altered secretion levels

Supposed secretion mutants and the mutagenesis starting strain were cultivated in 96-well deep-well plates. We measured HRP enzyme activity in the culture supernatants by a kinetic assay with chromogenic 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) diammonium salt (ABTS) as substrate. In the next step, we subjected hits of the primary screen with higher or lower HRP activity than the reference strain to two further rounds of screening following the same procedures as described above. Finally, we selected 12 mutants with increased and 11 mutants with decreased or abolished HRP activity for genome walking in order to identify the integration locus of the mutagenesis cassette.

Genome walking and bioinformatics search

Genomic regions flanking the insertion loci of the mutagenesis cassette were amplified and sequenced following the template-blocking PCR protocol of Bae and Sohn (2010). This protocol for PCR-based genome walking aims to prevent unspecific amplification that could lead to false positive results. We blasted the resulting nucleotide sequences against the genome sequence of *P. pastoris* GS115 and CBS7435. In most cases, the sequences obtained for one particular mutant strain could be assigned to a single integration event. Only two out of 23 analyzed mutants had a second copy of the mutagenesis cassette integrated. Altogether, we identified nine genes that were disrupted in mutants with positive effect on HRP activity in the supernatant (Table 1), and eight genes in mutants with negative effect (Table 2). In four mutants with zero activity of HRP the mutagenesis cassette had integrated into the HRP expression cassette itself, thereby abolishing expression.

To identify the proteins that are encoded by the genes disrupted in our screening hits, we blasted the corresponding amino acid sequences against the database of *S. cerevisiae* proteins on NCBI. The results are listed in Table 1 and 2. For two protein sequences, i.e. CCA41154.1 and CCA40244.1, no significant homology to any described *S. cerevisiae* protein was detected. The pBLAST results with the highest score for these two proteins were Mrp51 and Cwc2, respectively. It should be mentioned that the reverse approach, where we searched for the homologs of *S. cerevisiae* Mrp51 and Cwc2 in *P. pastoris,* identified other, more reliable hits as judged from sequence homology. Together with the detected low sequence coverage and identity, the pBLAST outcome indicated that no obvious homologs for these proteins exist in baker's yeast.

Three genes appeared as multiple hits in our screening. We found them disrupted in more than one mutant subjected to genome walking. Mapping of the insertion locus showed that the mutagenesis cassette had integrated at unique basepair positions within these mutants, identifying them as independent insertion events. The significance of these repeated hits influenced our decision on which genes to investigate further. Other factors taken into account were the signal intensity in the HRP screening and the proposed function of their *S. cerevisiae* homologs, as reviewed in the Discussion section. Based on the listed criteria we selected the homologs of Rim101, Sgt2, Kcs1 and the *P. pastoris* CBS7435 protein with the accession number CCA40244.1, which we termed Kep1 (knockout enhances protein secretion 1), for a more detailed analysis. The latter showed no significant homology to any described protein of *S. cerevisiae*. We did not further investigate gene deletions that reduced HRP activity for reasons addressed in the Discussion.

Targeted gene deletions and effect on secretion of model proteins

Random mutagenesis approaches harbor the danger that the observed phenotype is falsely attributed to a specific mutation, while it is in fact caused by another genetic alteration. To address this problem, we constructed clean knockouts of the most interesting genes found in the screening. The four selected ORFs (Table 1, in bold) were disrupted in the laboratory wild type strain CBS7435 *his4* strain with the knockout strategy described by Ahmad et al. (manuscript in preparation). In short, 5' and 3' homology regions of the targeted locus were cloned into a vector system that allows selection of transformants with Zeocin^{TM}. Following successful gene disruption, the marker and other elements of the vector were excised by recombination. This recombination event was facilitated by a recombinase enzyme, which was also encoded on the vector.

To assess the effects of gene disruptions on secretion, we transformed wild type and knockout strains with expression cassettes of HRP, APLE and hGH. The proteins were expressed from the constitutive P_{GAP} or the inducible P_{AOX1} promoter, and were fused to the signal sequence of *S. cerevisiae* α-mating factor to drive secretion to the culture supernatant. In all cases, the expression cassettes were targeted to the *AOX1* locus to increase comparability between different expression strains. Plasmids constructed for the secreted expression of HRP are depicted in Suppl. Fig. 2 and 3. Correct integration into the *AOX1* locus by a double cross-over event generated the methanol-utilization slow (Mut^S) phenotype, which we selected for on minimal methanol plates.

For characterization of the knockout strains expressing secreted HRP from P_{GAP}, we performed growth tests on buffered YPD (BYPD) in baffled shake flasks. As shown in Fig. 2a, all knockout strains had a growth rate similar to the wild type control. By contrast, HRP activity in the culture supernatants differed significantly (Fig. 2b). While the *kcs1*Δ knockout had no considerable effect

on HRP secretion in shake flasks, we could confirm the positive effect of *rim101*Δ, *sgt2*Δ and *kep1*Δ strains that had been suggested by the initial screening results. The most striking result was that HRP activity in the supernatant of *kep1*Δ did not show the same growth-dependent increase as observed in *rim101*Δ and *sgt2*Δ (Fig. 2b). Instead, HRP secretion in the *kep1*Δ strain seemed to increase dramatically as soon as the cells reached the stationary phase after ~40 h of cultivation time. This sudden accumulation of HRP in the supernatant is unlikely to result from cell lysis of *kep1*Δ, as no defect in cell wall integrity was observed in alkaline phosphatase (Larsen et al. 2013) and Calcofluor white plate assays (Roncero and Duran 1985) (data not shown).

Similar results were obtained when the same strains were cultivated in 96-well deep well plates. After 35 h of cultivation, the activity of HRP in the supernatants of *rim101*Δ, *sgt2*Δ and *kep1*Δ strains was increased relative to the wild type control, whereas it was unaffected in *kcs1*Δ (Fig. 3a). The observed increases could on the one hand be explained with a higher specific activity of recombinant HRP in these knockout strain supernatants, for example through improved folding and more efficient integration of the heme cofactor, which is required for enzymatic activity (Veitch 2004). On the other hand, the activity could be increased through more secreted HRP protein. To resolve this issue, we quantified the amount of enzyme present in the culture supernatant after 72 h of cultivation by Western blot analysis with a commercial primary antibody detecting HRP. As shown in Fig. 3b, HRP in its *N*-glycosylated form runs as a smear at high molecular weight (Veitch 2004; Wuhrer et al. 2005; Morawski et al. 2000). Already at the first glance the signal for *kep1*Δ appeared stronger than for the other strains, supporting the hypothesis of increased protein secretion. Quantification of the signal intensities was facilitated by deglycosylation with EndoH (Fig. 3c). The results followed the trend observed in the activity screen. As seen previously in the shake flask experiments, the beneficial effect of *kep1*Δ on HRP

secretion became more pronounced with prolonged incubation, resulting in 360 % intensity of the wild type band after 72 h of cultivation, as compared to 180 % HRP activity measured after 35 h (Fig. 3c).

To exclude that the beneficial effects of single gene knockouts on HRP secretion were P_{GAP}-HRP specific, we tested for HRP secretion governed by the strong inducible P_{AOX1}. The promoter change resulted in ~180 % HRP activity in the culture supernatant of *rim101*Δ relative to the wild type, in agreement with P_{GAP} results (Fig. 3d). Strikingly, the effect of *sgt2*Δ and *kep1*Δ knockouts yielded ~330 % and ~350 % relative HRP activity, respectively. The knockout strain *kcs1*Δ slightly benefited to ~130 % relative activity using this promoter instead of P_{GAP}. These results offer vital evidence that beneficial effects detected in a screening with P_{GAP} can be transferred to expression with other promoters, and might even multiply proportionally to the strength of the promoter.

We expected the same favorable effect of the knockout strains on the secretion of other recombinant proteins, speculating that the higher level of HRP in the culture supernatant stemmed from a general effect on the protein secretion machinery. To confirm this hypothesis, we tested for secretion of two other industrially important proteins expressed from P_{GAP}, i.e. APLE and hGH (Fig. 4). APLE had been expressed in *P. pastoris* by our group before, but secretion to the culture supernatant appeared to be hampered by its bulky trimeric structure (Hermann et al. 2008). The small hormone hGH, on the contrary, is secreted at levels easily detected by SDS-PAGE, also in wild type background (Ecamilla-Treviño et al. 2000; Calik et al. 2008). Performing deep-well cultivations as described for HRP above, we found that only the *kep1*Δ background had a positive effect on APLE secretion. Activity assays with the esterase substrate para-nitrophenyl acetate (pNPA) and Western blot analysis with primary antibody recognizing APLE both pointed to approximately 50 % increased APLE secretion relative to the wild type background. In case of hGH, expression in

*kep1*Δ doubled the amount of protein detected by Western blot using an anti-hGH primary antibody. In addition, also *rim101*Δ and *sgt2*Δ seemed to be slightly beneficial for hGH secretion. We did not test the influence of the *kcs1*Δ background on hGH secretion, as its effect on recombinant protein secretion appeared to be minor. Taken together, the results suggest that *kep1*Δ has a universal effect on protein secretion, whereas the beneficial effect of *rim101*Δ and *sgt2*Δ depends on the secreted protein.

Complementation restores wild type secretion behavior in knockout strains

In our initial screening we found the genes of *RIM101, SGT2* and *KEP1* disrupted by the mutagenesis cassette. To verify that depletion or loss of function of the encoded protein was responsible for the observed secretion enhancement, we decided to express the previously deleted genes in the respective HRP-secreting knockout strains. Therefore, we PCR-amplified the 5'-untranslated regions and ORFs of *RIM101, SGT2* and *KEP1*, and fused them to a C-terminal FLAG-tag and a selection marker. The resulting constructs were linearized in the genes' 5'-untranslated region, several hundred basepairs upstream of the ATG start codon. We chose this strategy with the intention to integrate the genes in-frame with their native promoters, thereby enabling wild type expression levels. Analysis of HRP activity in the culture supernatants of these strains showed that expression of the FLAG-tagged genes restored wild type secretion behavior (Fig. 5a). The results confirm that the secretion phenotypes were indeed caused by depletion of the gene products of the targeted loci. Trying to detect the tagged proteins in cell lysates with an antibody specific for the C-terminal FLAG-tag, we obtained a specific signal of the expected 37 kDa for Sgt2-FLAG (results not shown). The calculated molecular weight of Kep1-FLAG is ~67 kDa. In this size range we only observed an unspecific signal, which was visible also in the negative control strains. Similarly, we were not able to detect Rim101-FLAG, most likely because of

proteolytic processing of the C-terminus (Lamb et al. 2001), or because of its low expression level as a transcription factor.

We noticed that apart from enhanced recombinant protein secretion, the knockout of *KEP1* led to the release of another, yet unidentified endogenous protein into the culture supernatant. This release resulted in a prominent band of >100 kDa on an SDS-PAGE loaded with TCA-precipitated culture supernatants of the *kep1*Δ strain (Fig. 5c). In contrast, the same band appeared weakly in the wild type control. Expression of FLAG-tagged *KEP1* from its native promoter reversed this phenotype, verifying dependency of the phenotype on the presence of Kep1.

Discussion

P. pastoris has been used for the expression of recombinant proteins in research and industry for decades (Lin-Cereghino and Cregg 2000; Cregg et al. 2000). In the recent years, the genome sequence of *P. pastoris* has been published, thereby making this yeast available to genetic engineering approaches and strengthening its position as a popular production platform. Nevertheless, a large part of the annotated protein coding genes still lack basic characterization. Similarly, many questions about the molecular mechanism and effectors of secretion are still unanswered, even though secretion of recombinant protein is a popular strategy in this host. Random mutagenesis, as we used it in this study, represents a versatile tool to identify so far uncharacterized ORFs and link them to a certain phenotype (Novick and Schekman 1979; Ross-Macdonald et al. 1999).

Sequencing projects annotated 5313 protein coding genes in the histidine auxotrophic GS115 (De Schutter et al. 2009) and 5007 in the wild type strain CBS7435 (Küberl et al. 2011). Although a certain percentage of these genes is supposedly essential for the yeast to survive and cannot be disrupted, our screening of approximately 3000 mutants does not cover all possible gene disruptions. This aspect might explain why our screening did not identify well-known effectors of protein secretion as reviewed in Damasceno et al. (2011) and Idiris et al. (2010), or hits found by Larsen et al. (2013), who used a comparable mutagenesis approach. Strikingly, we discovered several hits more than once in the screening. Among the 23 sequenced insertion loci, the genes encoding Dus1 and Sgt2 were found twice, the gene encoding Rim20 even three times. Moreover, the screening uncovered two other members of the Rim-signaling cascade, Rim13 and Rim101

(Table 1). These repetitive findings suggested that we had covered a significant part of the yeast's genome with the relatively low number of mutants we had screened.

The genes that - upon disruption - reduced the level of HRP activity in the culture supernatant included two genes that potentially influence synthesis of the iron-containing heme cofactor (Table 2). Hem3 catalyzes the third step of heme biosynthesis (Keng et al. 1992), whereas Fra1 is involved in the regulation of iron uptake in *S. cerevisiae* (Kumánovics et al. 2008). Imbalances in heme biosynthesis are highly likely to have a negative influence on HRP activity as the heme cofactor is required for functionality. We trust that these mutations most probably do not negatively influence protein secretion *per se*. Regarding other genes listed in Table 2, we cannot exclude that diminished HRP activity stems from an overall deleterious effect of the gene disruption on cell growth or protein synthesis rates.

Thus, we focused on hits that increased HRP activity in the supernatant and selected four genes for targeted knockout. Shake flask and deep-well plate cultivation confirmed the screening results for *rim101Δ*, *sgt2Δ* and *kep1Δ* (Fig. 2b, 3a, 3d). Surprisingly, *kcs1Δ* triggered increased HRP secretion only when the enzyme was expressed from P_{AOX1}, but not when expressed from P_{GAP}, the promoter employed in the initial screening. Similarly unexpected, neither the knockout of *kcs1Δ* nor of *rim101Δ* or *sgt2Δ* affected secretion of APLE and hGH (Fig. 4a, 4b, 4c). In the following paragraphs, we will discuss possible functions of the most interesting hits *RIM101*, *SGT2* and *KEP1* in *P. pastoris*.

The transcription factor Rim101 is the ultimate target of a signal transduction pathway sensing extracellular pH. The so-called Rim pathway, extensively studied in *S. cerevisiae*, has been associated with diverse functions as growth at alkaline pH, sporulation, invasive growth, cell wall

construction and ion homeostasis (Lamb and Mitchell 2003). The pathway comprises seven dedicated components, namely Rim8, Rim9, Rim13, Rim20, Rim21, Dfg16 and Ygr122w (Maeda 2012). In response to alkaline pH, Rim101 is proteolytically activated by the calpain-like protease Rim13, causing it to relocate to the nucleus (Futai et al. 2001; Lamb et al. 2001; Lamb and Mitchell 2003). The interaction between Rim101 and Rim13 is mediated by the protease scaffold Rim20 (Xu and Mitchell 2001). Rim101, Rim20 and Rim13 were independently found in our screening in *P. pastoris*. The discovery of these three proteins supposedly also acting in the same cascade in *P. pastoris* strongly indicated that the Rim-regulon is actually connected to HRP secretion. We identified a possible link between the Rim pathway and the heme-containing enzyme HRP in the influence of Rim101 on iron homeostasis and regulation of iron-consuming pathways (reviewed for different fungi in Canessa and Larrondo 2013). The repression of iron-consuming pathways, such as heme biosynthesis, is mediated through the interaction of Rim101 with the transcription factor HapX (Canessa and Larrondo 2013; Franken et al. 2011).

In *P. pastoris*, we did not observe any of the defects described for the *rim101Δ* knockout in *S. cerevisiae*, i.e. reduced ion-tolerance, reduced growth at alkaline pH, or increased resistance to Calcofluor white (Lamb et al. 2001). Sauer et al. (2004) reported that, in strong contrast to *S. cerevisiae*, the intracellular pH of *P. pastoris* cells is not influenced by external pH conditions. This obvious discrepancy between the two yeasts suggests that Rim101 might have alternate biological functions in *P. pastoris* then in *S. cerevisiae*.

Regarding Sgt2, this cytosolic protein was described as a member of the GET-complex, which mediates the insertion of tail-anchored (TA) proteins into the ER. TA proteins comprise a diverse group of proteins characterized by the presence of a single C-terminal transmembrane domain that requires post-translational insertion to prevent aggregation (Borgese and Fasana 2011; Johnson et

al. 2013). TA proteins play critical roles throughout the secretory pathway. Among them are many SNARE proteins, which mediate secretory vesicle fusion (Beilharz et al. 2003). Inefficient vesicle fusion was shown to cause secretion of the ER-resident chaperone Kar2 in *get* mutants, due to reduced retrograde transport (Schuldiner et al. 2005; Schuldiner et al. 2008). Deletion of *SGT2* could lead to a cascade of secretory pathway alterations, finally affecting HRP secretion. Apart from its role in TA protein sorting, Sgt2 was indicated to mediate the interaction between heat-shock proteins (Hsps) and protein aggregates (Wang et al. 2010; Kohl et al. 2011; Kiktev et al. 2012). As a conclusion of their study on yeast prions, Kiktev and colleagues (Kiktev et al. 2012) speculated that Sgt2 could trigger the elimination of protein aggregates. If HRP had a stronger tendency to aggregate in the cytosol prior to translocation into the ER lumen than the other recombinant proteins tested, APLE and hGH, this could explain why the beneficial effect of *sgt2Δ* was only observed for HRP secretion. The advantage of *sgt2Δ* is also more pronounced when the recombinant enzyme is expressed from the strong, inducible P_{AOX1} than when expressed from P_{GAP}, suggesting that the deletion comes into effect when the secretory machinery is overloaded with newly synthesized cargo (Fig. 3a, 3d). We did not further test the hypothesis of potential HRP aggregation in the cytosol.

In the NCBI database of CBS7435 proteins, Kep1 is annotated as hypothetical protein CCA40244.1. We were not able to identify a close homolog of this protein in *S. cerevisiae* or any other related yeast. However, the protein shares a short stretch of homologous amino acids with other, so far uncharacterized, proteins of the yeasts *Ogataea parapolymorpha* and *Dekkera bruxellensis*. The corresponding gene was chosen for targeted disruption because the respective mutant showed an outstanding activity of HRP in repeated rounds of screening. In the same mutant strain, a second ORF was found to be disrupted, but targeted gene knockout confirmed that the

disruption of CCA40244.1 was responsible for the observed phenotype. We termed the encoded protein Kep1 (knockout enhances protein secretion), because knockout of this gene benefited the secretion of all three tested model proteins. Western blot analysis confirmed that not the specific activity of the enzymes HRP and APLE, but the amount of protein secreted to the culture supernatant was increased in this mutant (Fig. 3b, 3c, 4b). The growth-uncoupled increase of HRP activity in the culture supernatant might hint at an explanation for this effect (Fig. 2b). The sudden accumulation of activity was noticed approximately at the same time as cells entered the stationary phase. This observation implies a physiological change of the cell in this phase, resulting in increased secretion. Interestingly enough, we did not note any indication for reduced cell wall stability in this mutant.

Another indication for the physiological change specific for the *kep1Δ* strain was the release of an unidentified endogenous protein to the culture supernatant, a phenotype that could be reversed by restoring *KEP1* expression (Fig. 5b). Ongoing work focusses on the identification of this protein and investigates its relationship to Kep1.

In conclusion, our work has demonstrated that random mutagenesis of *P. pastoris* provides a very powerful tool to characterize molecular processes like secretion. The results of this study allowed us to draw connections between the secretion of recombinant HRP and the genes *SGT2* and *RIM101*, which had not been associated with secretion before. As the most striking result, we identified the orphan gene *KEP1*, the deletion of which has the potential to globally enhance recombinant protein secretion. Our data suggests that the *kep1Δ* knockout strain could be exploited for the secretory expression of various proteins in research and industry.

Acknowledgements

We would like to thank Laura Näätsaari for providing us with a plasmid encoding HRP and protocols for the ABTS assay. Furthermore, we thank all members of the Harald Pichler team and especially Tamara Wriessnegger for technical advice and gift of reagents. This work was supported by the DK Molecular Enzymology (CW) and the Higher Education Commission of Pakistan (MA).

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–402. doi: 10.1093/nar/25.17.3389

Bae J-H, Sohn J-H (2010) Template-blocking PCR: an advanced PCR technique for genome walking. Anal Biochem 398:112–6. doi: 10.1016/j.ab.2009.11.003

Baumann K, Adelantado N, Lang C, Mattanovich D, Ferrer P (2011) Protein trafficking, ergosterol biosynthesis and membrane physics impact recombinant protein secretion in *Pichia pastoris*. Microb Cell Fact 10:93. doi: 10.1186/1475-2859-10-93

Beilharz T, Egan B, Silver PA, Hofmann K, Lithgow T (2003) Bipartite signals mediate subcellular targeting of tail-anchored membrane proteins in *Saccharomyces cerevisiae*. J Biol Chem 278:8219–23. doi: 10.1074/jbc.M212725200

Borgese N, Fasana E (2011) Targeting pathways of C-tail-anchored proteins. Biochim Biophys Acta 1808:937–46. doi: 10.1016/j.bbamem.2010.07.010

Calik P, Orman MA, Celik E, Halloran SM, Calik G, Ozdamar TH (2008) Expression system for synthesis and purification of recombinant human growth hormone in *Pichia pastoris* and structural analysis by MALDI-TOF Mass Spectrometry. Biotechnol Prog 24:221–6. doi: 10.1021/bp070294t

Calmels T, Parriche M, Burand H, Tiraby G (1991) High Efficiency Transformation of *Tolypocladium geodes* Conidiospores to Phleomycin Resistance. Curr Genet 20:309-314. doi: 10.1007/BF00318520

Canessa P, Larrondo LF (2013) Environmental responses and the control of iron homeostasis in fungal systems. Appl Microbiol Biotechnol 97:939–55. doi: 10.1007/s00253-012-4615-x

Cregg JM, Cereghino JL, Shi J, Higgins DR (2000) Recombinant protein expression in *Pichia pastoris*. Mol Biotechnol 16:23–52. doi: 10.1385/MB:16:1:23

Damasceno LM, Huang C-J, Batt CA (2011) Protein secretion in *Pichia pastoris* and advances in protein production. Appl Microbiol Biotechnol. doi: 10.1007/s00253-011-3654-z

Drocourt D, Calmels T, Reynes J, Baron M, Tiraby G (1990) Cassettes of the *Streptoalloteichus hindustanus ble* gene for transformation of lower and higher eukaryotes to phleomycin resistance. Nucleic Acids Res 18:4009. doi: 10.1093/nar/18.13.4009

Ecamilla-Treviño L, Viader-Salvado J, Barrera-Saldana H, Guerrero-Olazaran M (2000) Biosynthesis and secretion of recombinant human growth hormone in *Pichia pastoris*. Biotechnol Lett 22:109–114. doi: 10.1023/A:1005675920451

Franken ACW, Lokman BC, Ram AFJ, Punt PJ, van den Hondel CAMJJ, de Weert S (2011) Heme biosynthesis and its regulation: towards understanding and improvement of heme biosynthesis in filamentous fungi. Appl Microbiol Biotechnol 91:447–60. doi: 10.1007/s00253-011-3391-3

Futai E, Kubo T, Sorimachi H, Suzuki K, Maeda T (2001) Molecular cloning of PalBH, a mammalian homologue of the *Aspergillus* atypical calpain PalB. Biochim Biophys Acta 1517:316–9. doi: 10.1016/S0167-4781(00)00256-6

Gasser B, Sauer M, Maurer M, Stadlmayr G, Mattanovich D (2007) Transcriptomics-Based Identification of Novel Factors Enhancing Heterologous Protein Secretion in Yeasts. Appl Environ Microbiol 73:6499–6507. doi: 10.1128/AEM.01196-07

Gatignol A, Baron M, Tiraby G (1987) Phleomycin Resistance Encoded by the *ble* Gene from Transposon Tn5 as a Dominant Selectable Marker in *Saccharomyces cerevisiae*. Mol Gen Genet 207:342-348. doi: 10.1007/BF00331599

Gibson DG, Young L, Chuang R, Venter JC, Hutchison CA, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–5. doi: 10.1038/nmeth.1318

Haid A, Suissa M (1983) Immunochemical identification of membrane proteins after sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Methods Enzymol 96:192–205. doi:10.1016/S0076- 6879(83)96017-2

Hermann M, Kietzmann MU, Ivancić M, Zenzmaier C, Luiten RGM, Skranc W, Wubbolts M, Winkler M, Birner-Gruenberger R, Pichler H, Schwab H (2008) Alternative pig liver esterase (APLE) - cloning, identification and functional expression in *Pichia pastoris* of a versatile new biocatalyst. J Biotechnol 133:301–10. doi: 10.1016/j.jbiotec.2007.10.010

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

Horvath A, Riezman H (1994) Rapid protein extraction from *Saccharomyces cerevisiae*. Yeast 10:1305-10

Idiris A, Tohda H, Kumagai H, Takegawa K (2010) Engineering of protein secretion in yeast: strategies and impact on protein production. Appl Microbiol Biotechnol 86:403–417. doi: 10.1007/s00253-010-2447-0

Johnson N, Powis K, High S (2013) Post-translational translocation into the endoplasmic reticulum. Biochim Biophys Acta 1833:2403–9. doi: 10.1016/j.bbamcr.2012.12.008

Keng T, Richard C, Larocque R (1992) Structure and regulation of yeast *HEM3*, the gene for porphobilinogen deaminase, Mol. Gen. Genet. 234 233–243. doi: 10.1007/BF00283844

Kiktev DA, Patterson JC, Müller S, Bariar B, Pan T, Chernoff YO (2012) Regulation of chaperone effects on a yeast prion by cochaperone Sgt2. Mol Cell Biol 32:4960–70. doi: 10.1128/MCB.00875-12

Koganesawa N, Aizawa T, Masaki K, Matsuura A, Nimori T, Bando H, Kawano K, Nitta K (2001) Construction of an expression system of insect lysozyme lacking thermal stability: the effect of selection of signal sequence on level of expression in the *Pichia pastoris* expression system. Protein Eng 14:705–10. doi: 10.1093/protein/14.9.705

Kohl C, Tessarz P, von der Malsburg K, Zahn R, Bukau B, Mogk A (2011) Cooperative and independent activities of Sgt2 and Get5 in the targeting of tail-anchored proteins. Biol Chem 392:601–8. doi: 10.1515/BC.2011.066

Kowalski JM, Parekh RN, Mao J, Wittrup KD (1998) Protein Folding Stability Can Determine the Efficiency of Escape from Endoplasmic Reticulum Quality Control. J Biol Chem 273:19453–19458. doi: 10.1074/jbc.273.31.19453

Kumánovics A, Chen OS, Li L, Bagley D, Adkins EM, Lin H, Dingra NN, Outten CE, Keller G, Winge D, Ward DM, Kaplan J (2008) Identification of *FRA1* and *FRA2* as genes involved in regulating the yeast iron regulon in response to decreased mitochondrial iron-sulfur cluster synthesis. J Biol Chem 283:10276–86. doi: 10.1074/jbc.M801160200

Lamb TM, Xu W, Diamond A, Mitchell AP (2001) Alkaline response genes of *Saccharomyces cerevisiae* and their relationship to the *RIM101* pathway. J Biol Chem 276:1850–6. doi: 10.1074/jbc.M008381200

Lamb TM, Mitchell AP (2003) The Transcription Factor Rim101p Governs Ion Tolerance and Cell Differentiation by Direct Repression of the Regulatory Genes *NRG1* and *SMP1* in *Saccharomyces cerevisiae*. Mol Cell Biol 23:677–686. doi: 10.1128/MCB.23.2.677-686.2003

Larsen S, Weaver J, de Sa Campos K, Bulahan R, Nguyen J, Grove H, Huang A, Low L, Tran N, Gomez S, Yau J, Ilustrisimo T, Kawilarang J, Lau J, Tranphung M, Chen I, Tran C, Fox M, Lin-Cereghino J, Lin-Cereghino GP (2013) Mutant strains of *Pichia pastoris* with enhanced secretion of recombinant proteins. Biotechnol Lett. doi: 10.1007/s10529-013-1290-7

Lin-Cereghino J, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol Rev 24:45–66. doi: 10.1111/j.1574-6976.2000.tb00532.x

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

Maeda T (2012) The signaling mechanism of ambient pH sensing and adaptation in yeast and fungi. FEBS J 279:1407–13. doi: 10.1111/j.1742-4658.2012.08548.x

Marx H, Sauer M, Resina D, Vai M, Porro D, Valero F, Ferrer P, Mattanovich D (2006) Cloning, disruption and protein secretory phenotype of the *GAS1* homologue of *Pichia pastoris*. FEMS Microbiol Lett 264:40–7. doi: 10.1111/j.1574-6968.2006.00427.x

Morawski B, Lin Z, Cirino P, Joo H, Bandara G, Arnold FH (2000) Functional expression of horseradish peroxidase in *Saccharomyces cerevisiae* and *Pichia pastoris*. Protein Eng 13:377–84. doi: 10.1093/protein/13.5.377

Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A (2012) Deletion of the *Pichia pastoris KU70* homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One 7:e39720. doi: 10.1371/journal.pone.0039720

Novick P, Schekman R (1979) Secretion and cell-surface growth are blocked in a temperature-sensitive mutant of *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 76:1858–62.

Pfeffer M, Maurer M, Stadlmann J, Grass J, Delic M, Altmann F, Mattanovich D (2012) Intracellular interactome of secreted antibody Fab fragment in *Pichia pastoris* reveals its routes of secretion and degradation. Appl Microbiol Biotechnol 93:2503–12. doi: 10.1007/s00253-012-3933-3

Roncero C, Duran A (1985) Effect of Calcofluor white and Congo red on fungal cell wall morphogenesis: in vivo activation of chitin polymerization. J Bacteriol 163:1180–1185.

Sauer M, Branduardi P, Gasser B, Valli M, Maurer M, Porro D, Mattanovich D (2004) Differential gene expression in recombinant *Pichia pastoris* analysed by heterologous DNA microarray hybridisation. Microb Cell Fact 3:17. doi: 10.1186/1475-2859-3-17

Schroder LA, Glick BS, Dunn WA (2007) Identification of pexophagy genes by restriction enzyme-mediated integration. Methods Mol Biol 389:203–218

Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt JF, Weissman JS, Krogan NJ (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. Cell 123:507–19. doi: 10.1016/j.cell.2005.08.031

Schuldiner M, Metz J, Schmid V, Denic V, Rakwalska M, Schmitt HD, Schwappach B, Weissman JS (2008) The GET complex mediates insertion of tail-anchored proteins into the ER membrane. Cell 134:634–45. doi: 10.1016/j.cell.2008.06.025

Stadlmayr G, Benakovitsch K, Gasser B, Mattanovich D, Sauer M (2010) Genome-scale analysis of library sorting (GALibSo): Isolation of secretion enhancing factors for recombinant protein production in *Pichia pastoris*. Biotechnol Bioeng 105:543–55. doi: 10.1002/bit.22573

Van Dijk R, Faber KN, Hammond AT, Glick BS, Veenhuis M, Kiel JAKW (2001) Tagging *Hansenula polymorpha* genes by random integration of linear DNA fragments (RALF). Mol Genet Genomics 266:646–56. doi: 10.1007/s004380100584

Veitch NC (2004) Horseradish peroxidase: a modern view of a classic enzyme. Phytochemistry 65:249–259. doi: 10.1016/j.phytochem.2003.10.022

Wang F, Brown EC, Mak G, Zhuang J, Denic V (2010) A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. Mol Cell 40:159–71. doi: 10.1016/j.molcel.2010.08.038

Whyteside G, Alcocer MJC, Kumita JR, Dobson CM, Lazarou M, Pleass RJ, Archer DB (2011) Native-state stability determines the extent of degradation relative to secretion of protein variants from *Pichia pastoris*. PLoS One 6:e22692. doi: 10.1371/journal.pone.0022692

Wuhrer M, Balog CIA, Koeleman CAM, Deelder AM, Hokke CH (2005) New features of site-specific horseradish peroxidase (HRP) glycosylation uncovered by nano-LC-MS with repeated ion-isolation/fragmentation cycles. Biochim Biophys Acta 1723:229–39. doi: 10.1016/j.bbagen.2005.02.013

Xu W, Mitchell AP (2001) Yeast PalA/AIP1/Alix homolog Rim20p associates with a PEST-like region and is required for its proteolytic cleavage. J Bacteriol 183:6917–23. doi: 10.1128/JB.183.23.6917-6923.2001

Tables

Table 1 Summary of identified mutants with positive effect on HRP activity in culture supernatant

Accession N° of affected ORF in *P. pastoris* GS115	Accession N° of protein in *P. pastoris* CBS7435	N° of independent hits[a]	Homologs[b]	Max. score	Query coverage	Max. identity	Comments
XM_002490784.1	CCA37611.1	2	Dus1p	556	99%	65%	Dihydrouridine synthase; modifies pre-tRNA(Phe) at U17
XM_002490898.1	CCA38922.1	1	Rph1p	419	66%	54%	JmjC domain-containing histone demethylase
XM_002489483.1	CCA36351.1	3	Rim20p	290	93%	29%	Protein involved in proteolytic activation of Rim101p in response to alkaline pH
XM_002492818.1	CCA39524.1	1	Rim13p	95.1	88%	22%	Protein involved in proteolytic activation of Rim101p in response to alkaline pH
XM_002492805.1	**CCA39536.1**	1	**Rim101p**	144	15%	62%	Transcriptional repressor in response to alkaline pH
XM_002490178.1	**CCA37018.1**	2	**Sgt2p**	256	98%	40%	Glutamine-rich cytoplasmic cochaperone; acts in GET-pathway
XM_002493375.1[c]	CCA41154.1	1	(Mrp51p)	32.0	13%	27%	No significant homology with *S. cerevisiae* proteome
XM_002494291.1[c]	**CCA40244.1**	1	**(Cwc2p)**	31.6	6%	47%	No significant homology with *S. cerevisiae* proteome; termed *KEP1* (knockout enhances protein secretion)
no ORF annotated[d]	**CCA41142.1**	1	**Kcs1p**	244	67%	80%	Inositol hexakisphosphate (IP6) and inositol heptakisphosphate (IP7) kinase

Entries in bold underline open reading frames that were selectively knocked out for further analysis

[a] Repetitive hits found in screening with different insertion positions within the ORF

[b] Protein homolog with highest max. score in *S. cerevisiae* identified by Protein BLAST search

[c] Hits were identified by genome walking in the same strain

[d] ORF annotated for CBS7435, not for GS115

Table 2 Summary of identified mutants with negative effect on HRP activity in culture supernatant

Accession N° of affected ORF in *P. pastoris* GS115	Accession N° of protein in *P. pastoris* CBS7435	Number of independent hits[a]	Homologs[b]	Max. score	Query coverage	Max. identity	Comments
XM_002489715.1	CCA36575.1	1	Om45p	42.0	43%	29%	Mitochondrial outer membrane protein of unknown function
XM_002490919.1	CCA38902.1	1	Hua2p	33.1	13%	36%	Cytoplasmic protein of unknown function
XM_002491055.1	CCA38772.1	1	Pmt1p	855	98%	54%	Protein O-mannosyltransferase of the ER membrane
XM_002490379.1	CCA37218.1	1	Hem3p	350	85%	54%	Porphobilinogen deaminase, catalyzes third step in heme biosynthesis
XM_002493262.1	CCA39075.1	1	YML020W	286	50%	38%	Putative protein of unknown function
XM_002489709.1	CCA36569.1	1	Fra1p	614	97%	44%	Protein involved in negative regulation of transcription of iron regulon
XM_002491785.1[c]	CCA38070.1	1	Crc1p	311	96%	54%	Mitochondrial inner membrane carnitine transporter
XM_002493383.1[c]	CCA41145.1	1	Ubp15p	882	96%	40%	Ubiquitin-specific protease involved in protein deubiquitination

[a] Repetitive hits found in screening with different insertion positions within the ORF

[b] Protein homolog with highest max. score in *S. cerevisiae* identified by Protein BLAST search

[c] Hits were identified by genome walking in the same strain

Figures

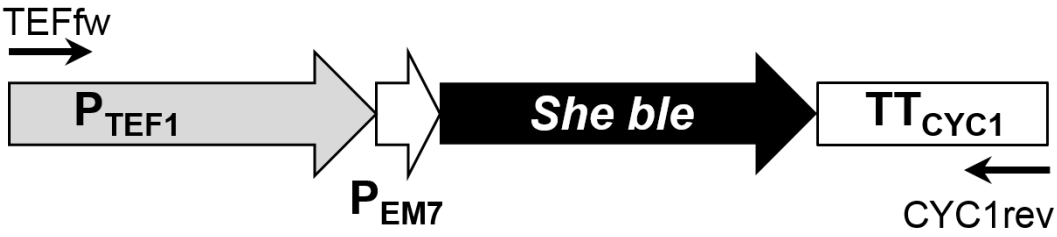


Fig. 1 Schematic drawing of Zeocin™ resistance cassette used for random mutagenesis of *P. pastoris* cells. The 1172 bp fragment composed of eukaryotic promoter P_{TEF1}, prokaryotic promoter P_{EM7}, *She ble* open reading frame and *CYC1* transcription terminator was amplified by PCR with the indicated primers.

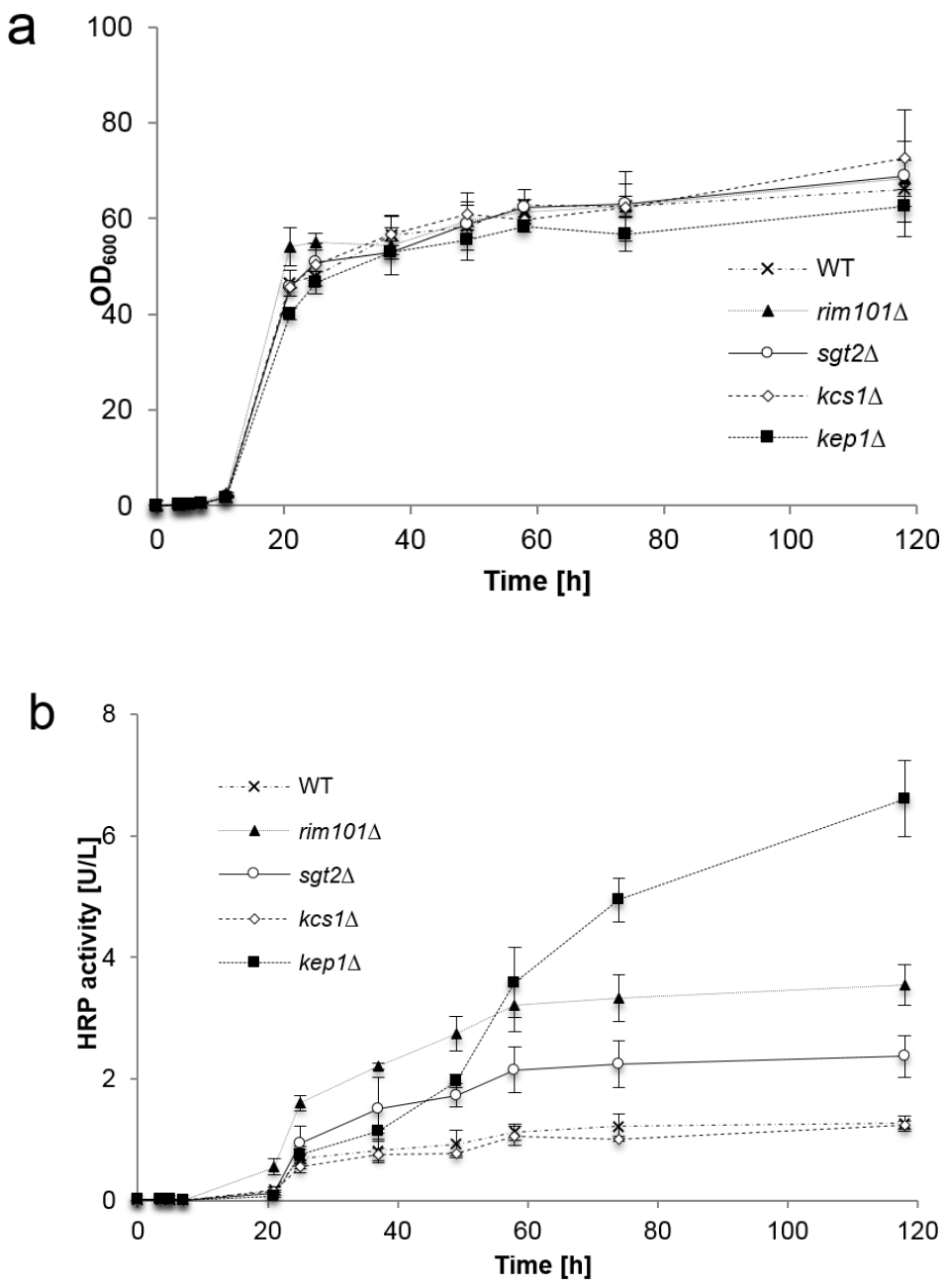
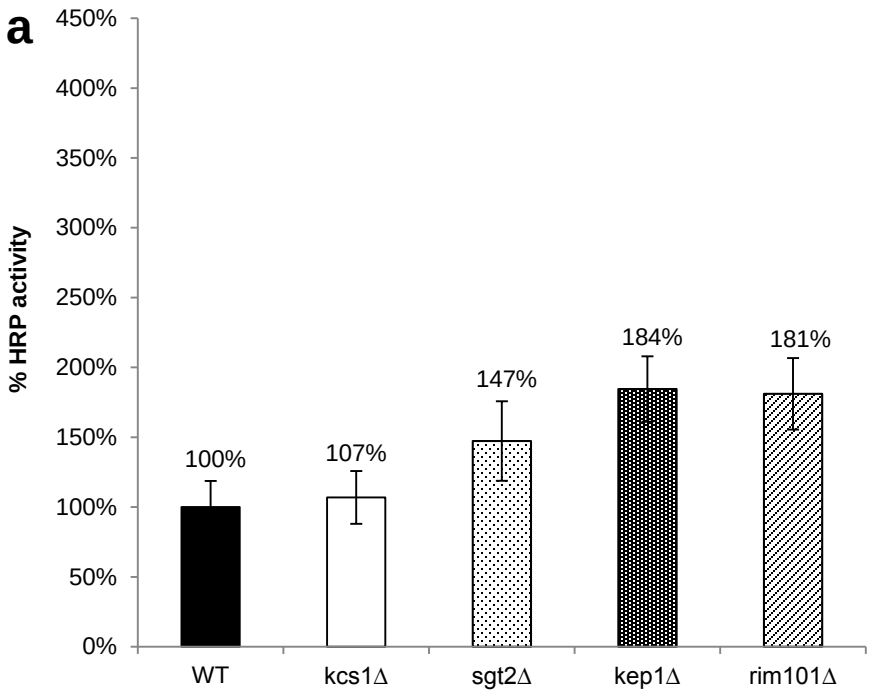
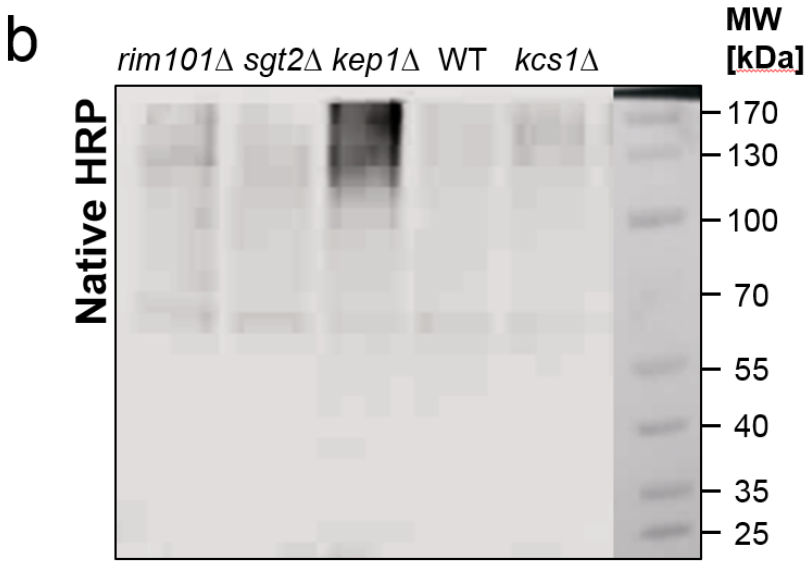


Fig. 2 HRP activity in supernatants of *kep1Δ* does not correlate with growth. Growth curve analysis of control and knockout strains secreting HRP from *GAP* promoter (a). Cells were grown in BYPD (2 % glucose) in baffled shake flasks at 28°C. Experiments were performed in biological triplicate. HRP activity was detected in culture supernatants with ABTS assay at the same time points (b)


a
% HRP activity
450%
400%
350%
300%
250%
200%
150%
100%
50%
0%
100%
107%
147%
184%
181%
WT
kcs1Δ
sgt2Δ
kep1Δ
rim101Δ


b
rim101Δ sgt2Δ kep1Δ WT kcs1Δ
MW
[kDa]
Native HRP
170
130
100
70
55
40
35
25

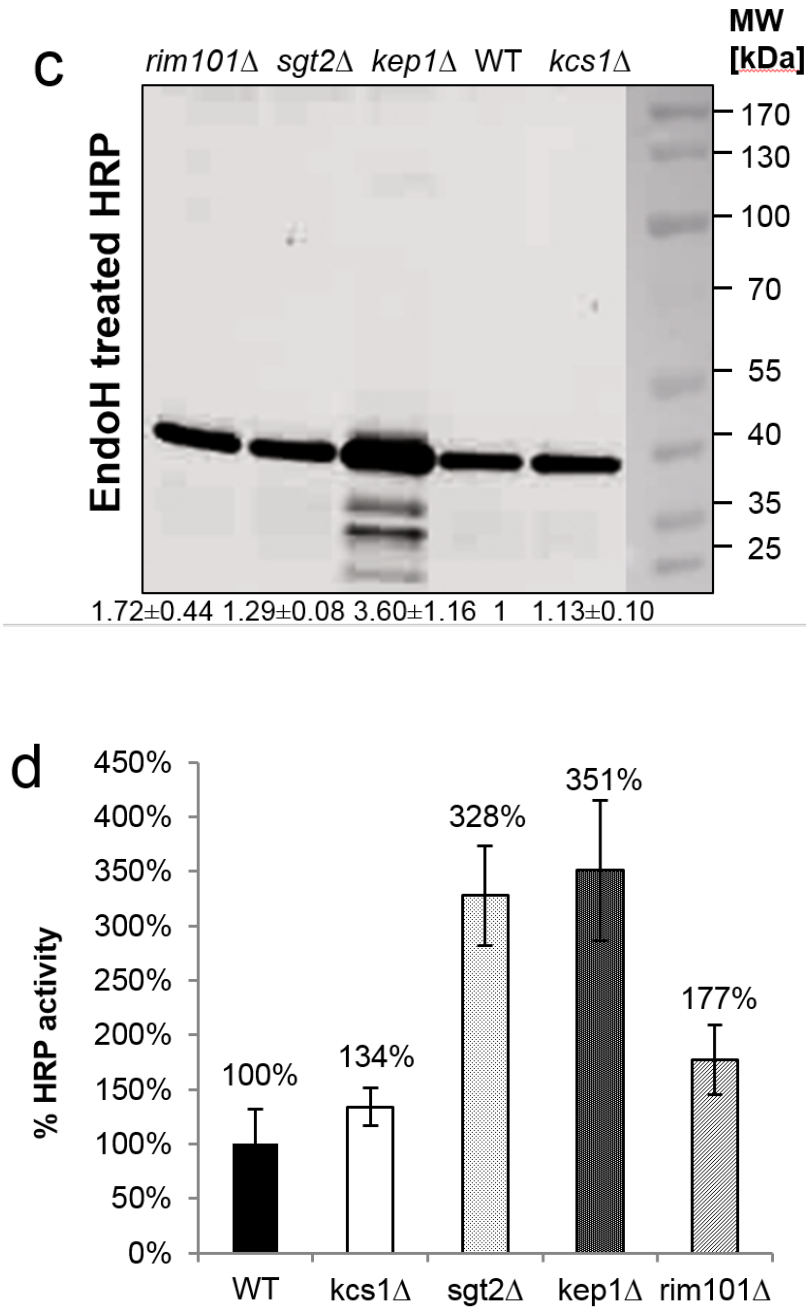


Fig. 3 Quantification of secreted HRP in deep-well plate culture supernatants. Relative HRP activity upon expression from *GAP* promoter (a). WT control and knockout strains expressing HRP from P_{GAP} were grown on BYPD (2 % glucose) in deep-well plates for 35 h before analysis. Peroxidase activity in supernatants was quantified with ABTS assay. Results represent the mean of four biologically independent experiments with 12 technical replicates per experiment.

Western blot analysis of native HRP running as smear (b). Western blot after de-glycosylation of proteins in culture supernatants with EndoH and densitometric scanning (c). Numbers indicate the relative levels of HRP. Expression from P_{GAP} for 72 h. The data presented are averaged from two independent experiments. Relative HRP activity upon expression from P_{AOX1} (d). Strains expressing HRP from P_{AOX1} were grown in BMGY for 32 h and induced with methanol for 48 h for activity assays as above. Results represent mean of three biologically independent experiments with 12 technical replicate samples per experiment.

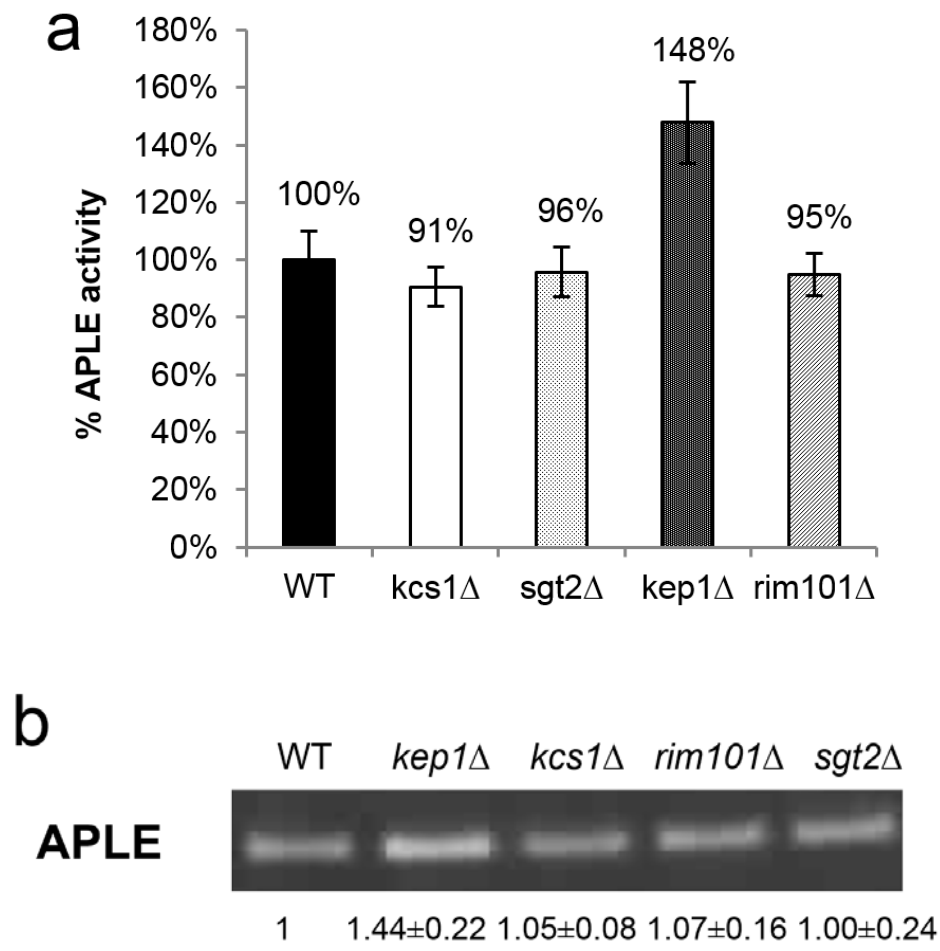
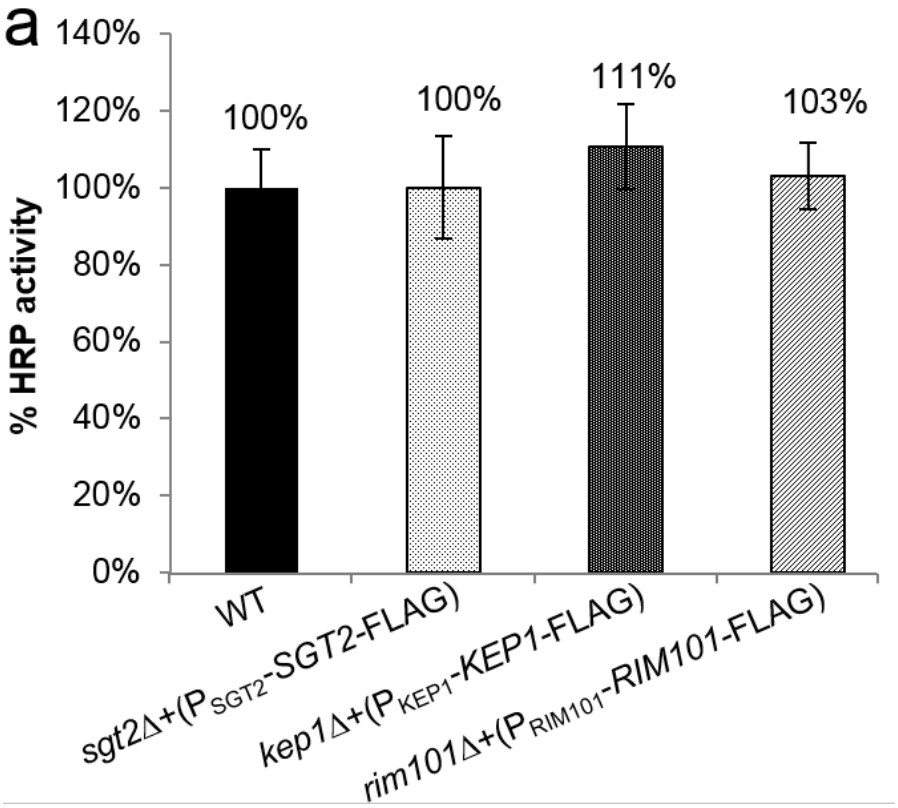


Fig. 4 Quantification of APLE and hGH secretion in deep-well plate cultivations. Strains expressing APLE from P_{GAP} were grown on BYPD (2 % glucose) in deep-well plates for 35 h before analysis. Relative activity of APLE in culture supernatants was measured with pNPA assay (a). Results represent the mean of six biologically independent experiments, with 3-12 technical replicates per experiment. Relative APLE protein levels in culture supernatants determined by Western blotting and densitrometric scanning (b). Numbers indicate the relative levels of APLE and are presented as the averages from two technical replicates. hGH protein levels, as determined by Western blotting and densitrometric scanning (c). Numbers indicate the relative levels of hGH in the supernatant. The data represents one experiment analyzed in duplicate.


a
140%
120%
100%
80%
60%
40%
20%
0%
% HRP activity
100%
100%
111%
103%
WT
sgt2Δ+(P SGT2-SGT2-FLAG)
kep1Δ+(P KEP1-KEP1-FLAG)
rim101Δ+(P RIM101-RIM101-FLAG)

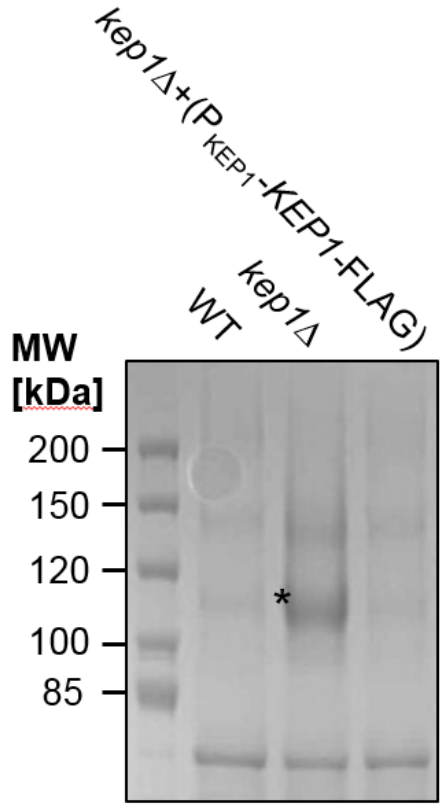

kep1Δ+(P KEP1-KEP1-FLAG)
kep1Δ
WT
MW
[kDa]
200
150
120
100
85
*

Fig. 5 Expression of targeted proteins in respective knockout strains restores wild-type HRP secretion levels. Relative HRP activity upon expression from *GAP* promoter (a). The knockout strains were transformed with expression constructs coding for FLAG-tagged versions of the previously knocked out target gene under control of its native promoter to reconstitute wild type-like expression. All strains expressed HRP from P_{GAP} and were grown on BYPD (2 % glucose) in deep-well plates for 35 h before analysis. Peroxidase activity in supernatants was measured with ABTS assay. Results represent the mean of six biologically independent experiments with 12 technical replicates per experiment. The knockout strain *kep1Δ* secretes an endogenous protein of >100 kDa (*) at higher levels than the WT strain, an effect that can be reversed by expression of FLAG-tagged *KEP1* (b). Strains were grown on BYPD (2 % glucose) in deep-well plates for 72 h. Proteins in culture supernatants were precipitated in 33 % TCA, resolved by SDS-PAGE and stained with Coomassie blue.

Supplementary

Suppl. table 1 All strains constructed during this study

Name	Description	Source
GS115	GS115 Δ*his4*	Life Technologies, Carlsbad, CA
GS115+GAPαHRP	GS115 Δ*his4 pPIC9toGAPαHRP*	This study
WT	CBS7435 *Δhis4*	Näätsaari et al. (2012)
kep1Δ	CBS7435 *Δhis4 Δkep1*	This study
kcs1Δ	CBS7435 *Δhis4 Δkcs1*	This study
rim101Δ	CBS7435 *Δhis4 Δrim101*	This study
sgt2Δ	CBS7435 *Δhis4 Δsgt2*	This study
WT+GAPαHRP	CBS7435 *Δhis4 Δaox1::pGaHSwaHRP*	This study
WT+AOX1αHRP	CBS7435 *Δhis4 Δaox1::pAaHSwaHRP*	This study
WT +GAPαAPLE	CBS7435 *Δhis4 Δaox1::pGaHSwaAPLE*	This study
WT+GAPαhGH	CBS7435 *Δhis4 Δaox1::pGaHSwahGH*	This study
kep1Δ+GAPαHRP	CBS7435 *Δhis4 Δkep1 Δaox1::pGaHSwaHRP*	This study
kep1Δ+AOX1αHRP	CBS7435 *Δhis4 Δkep1 Δaox1::pAaHSwaHRP*	This study
kep1Δ+GAPαAPLE	CBS7435 *Δhis4 Δkep1 Δaox1::pGaHSwaAPLE*	This study
kep1Δ+GAPαhGH	CBS7435 *Δhis4 Δkep1 Δaox1::pGaHSwahGH*	This study
KEP1flag	kep1Δ+GAPαHRP *pPpT4_KEP1_C.FLAG*	This study
kcs1Δ+GAPαHRP	CBS7435 *Δhis4 Δkcs1 Δaox1::pGaHSwaHRP*	This study
kcs1Δ+AOX1αHRP	CBS7435 *Δhis4 Δkcs1 Δaox1::pAaHSwaHRP*	This study
kcs1Δ+GAPαAPLE	CBS7435 *Δhis4 Δkcs1 Δaox1::pGaHSwaAPLE*	This study
rim101Δ+GAPαHRP	CBS7435 *Δhis4 Δrim101 Δaox1::pGaHSwaHRP*	This study
rim101Δ+AOX1αHRP	CBS7435 *Δhis4 Δrim101 Δaox1::pAaHSwaHRP*	This study
rim101Δ+GAPαAPLE	CBS7435 *Δhis4 Δrim101 Δaox1::pGaHSwaAPLE*	This study
rim101Δ+GAPαhGH	CBS7435 *Δhis4 Δrim101 Δaox1::pGaHSwahGH*	This study
RIM101flag	rim101Δ+GAPαHRP *pPpT4_RIM101_C.FLAG*	This study
sgt2Δ+GAPαHRP	CBS7435 *Δhis4 Δsgt2 Δaox1::pGaHSwaHRP*	This study
sgt2Δ+AOX1αHRP	CBS7435 *Δhis4 Δsgt2 Δaox1::pAaHSwaHRP*	This study
sgt2Δ+GAPαAPLE	CBS7435 *Δhis4 Δsgt2 Δaox1::pGaHSwaAPLE*	This study
sgt2Δ+GAPαhGH	CBS7435 *Δhis4 Δsgt2 Δaox1::pGaHSwahGH*	This study
SGT2flag	sgt2Δ+GAPαHRP *pPpT4_SGT2_C.FLAG*	This study

Suppl. table 2 All vectors constructed during this study

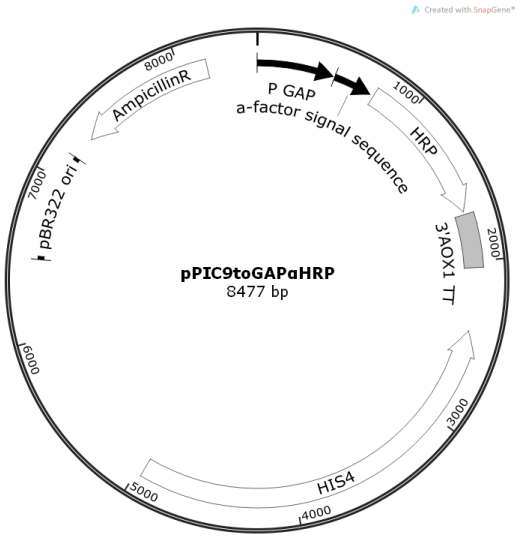
Name	Description	Source
pGAPZαA	pPpGAP-alpha ss-Zeocin	Life Technologies, Carlsbad, CA
pPIC9	pPpAOX1-alpha ss-HIS4	Life Technologies, Carlsbad, CA
pPIC9toGAPαHRP	pPpGAP-alpha ss-HIS4	This study
pAaHSwa	5'AOX1-PAOX1-alpha ss -TT-HIS4-3'AOX1	This study
pGaHSwa	5'AOX1-PGAP-alpha ss-TT-HIS4-3'AOX1	This study
pPpKC1	FRT-Zeocin-Flippase-FRT	Ahmad et al. (manuscript in prep.)
pPpKC1_KEP1	5'KEP1-FRT-Zeocin-Flippase-FRT-3'KEP1	This study
pPpKC1_RIM101	5'RIM101-FRT-Zeocin-Flippase-FRT-3'RIM101	This study
pPpKC1_KCS1	5'KCS1-FRT-Zeocin-Flippase-FRT-3'KCS1	This study
pPpKC1_SGT2	5'SGT2-FRT-Zeocin-Flippase-FRT-3'SGT2	This study
pPpT4_C.FLAG	pPpAOX1-FLAGtag-TT-Zeocin	This study
pPpT4_RIM101_C.FLAG	5'RIM101-RIM101-FLAGtag-TT-Zeocin	This study
pPpT4_SGT2_C.FLAG	5'SGT2-SGT2-FLAGtag-TT-Zeocin	This study
pPpT4_KEP1_C.FLAG	5'KEP1-KEP1-FLAGtag-TT-Zeocin	This study

Suppl. table 3 All primers used during this study (During the course of the project, the names we used to refer to certain gene loci, changed: KEP1=H8=FLO11=MOEP; H7=KCS1)

Name	**Sequence**
Amplification of mutagenesis cassette	
TEFfw	CCCACACACCATAGCTTCAAAATG
CYC1rev	AGCTTGCAAATTAAAGCCTTCGAG
Genome walking (Template-Blocking PCR)	
CSF27	GACGCGTAATACGACTCACTATAGGGA
CSR30	ATCTCCCTATAGTGAGTCGTATTACGCGTC
CP	ACGCGTAATACGACTCACTATAGGGAGATC
GSPCYC1a	GAGTTAGACAACCTGAAGTCTAGGTCCCTA
GSPCYC1b	GTACAGACGCGTGTACGCATGTAACATTATAC
GSPTEFa	TTCCAAACCTTTAGTACGGGTAATTAACGACAC
GSPTEFb	GCTGTGCTTGGGTGTTTTGAAGTGGT
Construction and verification of knockouts	
3UTRSgt2F	TCGGCCGATCAGGCCGGCGGTGATAAGAAGCCTTAAATTTATAATCTTTCT
3UTRSgt2R	CCTGGAAGAGCATGAATATTATGTTCGTTAAGGTTAATTCGGTTTGTAGCT
5UTRSgt2F	CCTTAACGAACATAATATTCATGCTCTTCCAGGAAACGTTACAAATAA
5UTRSgt2R	TCGGCCCTAGTGGCCGTTTCCCTCCAGCTTGAAAGCTTC
Up5UTRSgt2F	CTTGGAGACCAACTGCATAATATGGG
Down3UTRSgt2R	CCGAACTCGTTTCTCAACTACAAGATC

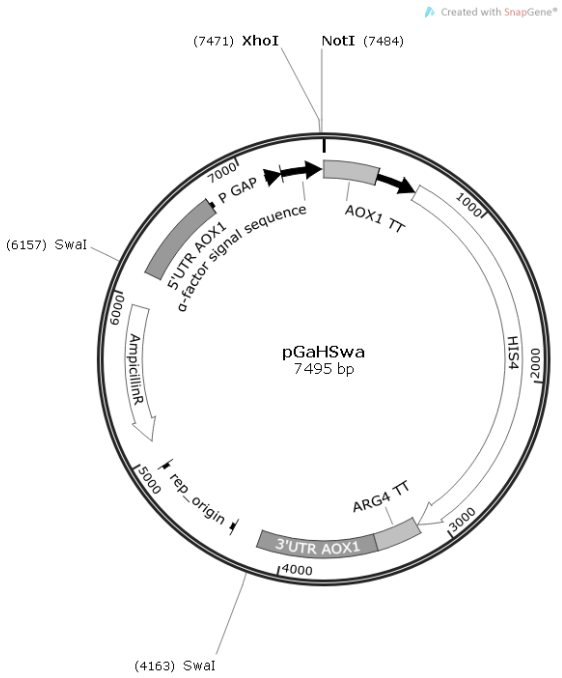
Up2SGT2fw	GCATCTTCAACTAGGACAGATAGCAC
Down2SGT2rev	GAGGCAATTCAGTTACTCAATGATCGAG
InSGT2fw	GATCCCTCGTATGTTAAGGCCTATTC
UpSgt2	GTGTACCAGTTGTTTGACGAAACTTTC
DownSgt2	GAGCTATGCGACATTACTGAGTAGATAG
3UTRhypprotH8F	TCGGCCGATCAGGCCCGGAAGTGATACTAAATTTGAATATGGAAGGGC
3UTRhypprotH8R	AATATCACTATGATCTTAAGTGAATTTAAATTATCTAGCTTTCCCAGGTACGCTC
5UTRhypprotH8F	GGAAAGCTAGATAATTTAAATTCACTTAAGATCATAGTGATATTATAACTCAATCCTCC
5UTRhypprotH8R	TCGGCCCTAGTGGCCCAACTCCTTCCAAGAATGATGTAAATACCTACC
Up5UTRhypprotH8F	CCTTCGCAGTATATCTACCCAGGC
Down3UTRhypprotH8R	CCCAGCTTGTCATCCTTGTCG
UpFlo11	GAAATCGAATTGCGAAGGGTACCTG
DownFlo11	GTTGACCGGCAACAAATACGATATC
Up2hypH8fw	CGAAACCATTATCGCGCTGAAATG
In1hypH8fw	GGACCTTTCCCATTGTTAAGTCTAG
In2hypH8fw	GACTATCGTTCCAAGGTGAACCAG
Down2hypH8rev	AGGTAACTCAGCAGGAGACTTATG
InhypH8rev	GCACTTCAACGTTCCATGATGCTC
3UTRhypprotH7F	TCGGCCGATCAGGCCCATAATACGCTGTATAATACATAATAATACAAAGAACTAGCCAT
3UTRhypprotH7R	AGAAATCGTCCCGGGAAAGTCTCCCAGTTGACTAACTTTAC
5UTRhypprotH7F	AGACTTTCCCGGGACGATTTCTCCGAGAACATCAA
5UTRhypprotH7R	TCGGCCCTAGTGGCCTAGTAGCTCTGGATGACGCATCTCTAT
Up5UTRhypprotH7F	GGGATCAGTTTACTGTATACCCAATCTTTGG
Down3UTRhypprotH7R	CCTTATGCGCACTTGCTATCTCAAC
UpKcs1	CACCTTCTGCGCAATCTCTC
DownKcs1	CGCATAGCTTACTTCTTTGGCAAC
5UTRRim101F	GCGGGATCCATCTTTGGAAACACAGGCTTGCCATC
5UTRRim101R	CTCGGCCCTAGTGGCCTCAACAGTTCAAAGACAAGTTCTTGTTG
3UTRRim101F	TACGGCCAATCGGGCCTAATTATCACGGCCATTCTGCTTACAAC
3UTRRim101R	GCCGGATCCCCTTCTTTCATAGTCGTAGTAGCAGATTGTTCC
UpRim101	CAAATAGCCCACTTTAATGACCGTTAAC
DownRim101	CATTACCCTCACTAGAACCAGAAAGAG
InRIM101fw	CTCAAATGGGCCATTCCAGTG
InRIM101rev	CTTGGTGATGAGCTGTATGATCCATG
PAox1SeqR	GGTTTCATTCAACCTTTCGTCTTTGGATG
PucSeqF	CTTTTTACGGTTCCTGGCCTTTTGC
UpAOX1	GAAATAGACGCAGATCGGGAAC
DownAOX1	CCAAATAGATTAGCTGTTTTGCCCTAATGTAC
Expression of FLAG-tagged genes (Gibson cloning)	
AOX1TT_BamHIR	AAGGATCCTCCGGAGCACAAACGAACGTCTCAC

16PNotIFwd	AAGCGGCCGCGAGTCGTGAGGACTATAAGGATGACGACGATAAG
AODTTpUC_fw	AAACTTGGATCTGATTACCTTAGGGCGCGCCCCCGTAGAAAAGATCAAAGGATCTTCTTG
AODTTpUC_rev	AAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGGCGCGCCCTAAGGTAATC
pUCH8fw	ACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACAATTCAAAGGTAACTCAGCAGGAGAC
pUCH8rev	TCCATTACATAAGTCTCCTGCTGAGTTACCTTTGAATTGTGAGCAAAAGGCCAGCAAAAG
H8flagfw	GTTACAATAATGACAAAGCAAAGCATGACTATAAGGATGACGACGATAAGTAATCAAGAG
H8flagrev	GATTACTTATCGTCGTCATCCTTATAGTCATGCTTTGCTTTGTCATTATTGTAACTCTTG
pUCRIM101fw	ACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACAGTACCTCAAGAACGGTACACTAGAG
pUCRIM101rev	ATTGTTGTTTTTCTCTAGTGTACCGTTCTTGAGGTACTGTGAGCAAAAGGCCAGCAAAAG
RIM101flagfw	GTTCTTTATATCCTACTATTGTTGTTGACTATAAGGATGACGACGATAAGTAATCAAGAG
RIM101flagrev	TTACTTATCGTCGTCATCCTTATAGTCAACAACAATAGTAGGATATAAAGAACTGCCTTC
pUCSGT2fw	CGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACACCTTTCAACTGCCGTTAATGGCTTAC
pUCSGT2rev	ATGATCCCTGAGTAAGCCATTAACGGCAGTTGAAAGGTGTGAGCAAAAGGCCAGCAAAAG
SGT2flagfw	AGTTCATGGGCGGTGATAAGAAGCCTGACTATAAGGATGACGACGATAAGTAATCAAGAG
SGT2flagrev	CATCCTCTTGATTACTTATCGTCGTCATCCTTATAGTCAGGCTTCTTATCACCGCCCATG
Construction of plasmid pGaHSwa	
OePGapAlphaR	GAAGGAAATCTCATTGTGTTTT
OePAox1GapR	CTACAAAAATTATTAGAGATTA
OePGapAlphaF	CAAAACACAATGAGATTTCCTT
OePAox1GapF	AAAATAATCTCTAATAATTTTT

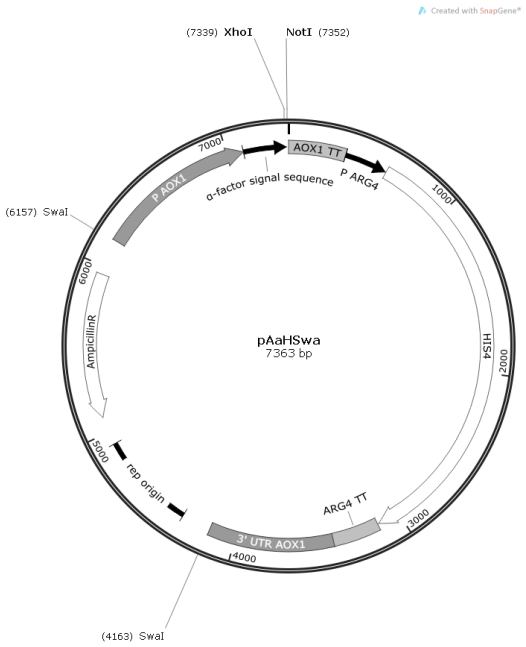


Suppl. Fig. 1 Plasmid pPIC9toGAP with HRP-C1A CDS integrated between *Xho*I and *Not*I sites. The plasmid was constructed by restricting pGAPZαA-HRP and pPIC9 with *Bgl*II and *Not*I.

The resulting fragments containing P_{GAP}+α-signal sequence+HRP, *HIS4* marker and pBR322 ori+beta-lactamase CDS were ligated to yield this plasmid.



Suppl. Fig. 2 Expression plasmid pGaHSwa. The plasmids can be targeted to the *AOX1* locus by restriction with *Swa*I. The target gene is inserted downstream of P_{GAP} and the α-mating factor signal sequence by cloning with *Xho*I and *Not*I. The *HIS4* gene acts as selectable marker.



Suppl. Fig. 3 Expression plasmid pAaHSwa. The plasmids can be targeted to the *AOX1* locus by restriction with *Swa*I. The target gene is inserted downstream of P_{AOX1} and the α-mating factor signal sequence by cloning with *Xho*I and *Not*I. The *HIS4* gene acts as selectable marker.

Conclusion and future outlook

During the course of this study, several valuable tools and methods for genetic engineering of *Pichia pastoris* were generated. These methods and tools presented here are strongly expected to significantly contribute to the optimization of recombinant protein expression – intracellular and secreted – and investigation of gene function in *P. pastoris*. A set of 27 expression vectors (pXYZ vector family) was newly designed and constructed based on CBS 7435 (NRRL Y-11430) strain background for which patent has expired and there are no material rights pending ***(1)***. Therefore, these vectors can be used for research and commercial purposes without paying licensing fees ***(1, 2)***. One of the main advantages of these expression vectors is that the expression of selection marker is uniformly controlled by a short and weaker *ARG4* promoter. The basal level of expression from this promoter is sufficient to select single copy transformants. Therefore, transformants are not burdened with expressing selection marker protein at a higher rate than needed. On the other hand, this basal level of expression of the *Sh ble* gene, which confers resistance to Zeocin^{TM} antibiotic, can be used to select multicopy transformants with higher efficiency at lower concentrations of the drug. The targeting efficiency of these vectors for integration at the *AOX1* locus was measured to be in the range of 60-70%, which is considerably higher than the reported efficiency of 5-25% with the previous developed *TEF1*-driven expression ***(3, 4)***. There is yet no explanation available on why targeting efficiencies of these vectors are so high? Additionally, these vectors were used to investigate the secretory potential of *Pp*_αMF secretion leader ***(5)***. In comparison with *Sc*_αMF, the secretion efficiency was determined for *Bacillus subtilis* levanase and horseradish peroxidase ***(6)*** as reporter proteins. The obtained data shows that *Pp*_αMF is inferior to *Sc*_ αMF leader sequence in directing the reporter proteins to the exterior of cell, at least for the tested proteins. Furthermore, it appears that cleavage of *Pp*-αMF pro-region is more dependent on *KEX2* protease activity than in the case of *Sc*_αMF ***(7–10)***. Further experiments are needed to gain more insight into the role of *Glu-Ala* spacer sequences and the dependence of *Pp*_αMF on *KEX2* protease activity.

A major part of this thesis was the development of a simple and potent system for targeted gene deletion in *P. pastoris*. The knockout vectors can easily be adopted to the gene of interest and *P. pastoris* strain by exchanging homology regions and selection markers with a single and efficient cloning step. In addition to that, the selection markers can be easily recycled by induced expression of *FLP/FRT* recombinase, thus, allowing the repeated use of same selection marker for multiple gene deletions. In order to verify our knockout strategy, we created deletions for five genes involved in the amino acid biosynthetic pathways i.e., *LYS2, MET2, TYR1, PRO3* and *PHA2*. Deletion of these gene resulted in auxotrophy for lysine, methionine, tyrosine and proline except for phenylalanine, which showed a bradytroph phenotype hinting that there is an alternative but less efficient pathway in *P. pastoris* for synthesis of this amino acid ***(11–14)***. In addition to these biosynthetic genes, we also created 26 gene deletions for known and putative proteases. The targeting efficiencies for different loci varied greatly from 4 – 88%. The knockout efficiency also appeared to be affected by the selection marker used, especially in case of genes that effect viability of deletion strain. For example, it was not possible to generate *PEP4* and *KEX2* knockout strains using Zeocin™ as a selection marker. However, by substituting this marker with *HIS4* we were able to generate Δ*pep4* and Δ*kex2* deletion strains right away. As expression of Zeocin™ is controlled by a short and weaker *ARG4* promoter, we reasoned that it could be due to the suboptimal expression level of this marker gene. However, we were not able to transform these two deletion strains with vectors where Zeocin™ expression is controlled by a stronger promoter e.g., P_{ILV5} or P_{TEF1} ***(15, 16)***. Further, analysis of these deletion strains revealed that they are less viable, therefore, increased selection pressure from Zeocin™ antibiotic resulted in cell death. Moreover, with slight modification in the architecture of these knockout vectors, they can be used to integrate cascade of expression cassettes at defined loci in *P. pastoris*. Therefore, this ability could make these knockout vectors an essential tool for metabolic engineering of *P. pastoris*.

The third part of this thesis, describes a novel cloning strategy, restriction site free cloning (RSFC), based on type II S restriction enzymes. The RSFC cloning strategy can be used to optimize expression of tagged proteins with minimal cloning effort. A single PCR product of the gene of interest (GOI) can be inserted into all the 40 plasmids irrespective of the upstream or downstream DNA sequences. Testing of expression of GOI in multiple hosts, based on ligation independent cloning, using two PCR products has been reported previously ***(17–21)***. These methods are dependent on the annealing of single stranded overhangs, therefore, require identical sequences present on both the vector and GOI. One critical disadvantage of the RSFC cloning strategy is blunt end ligation of the GOI insert, which is notoriously less efficient than recombination based methods and requires confirmation of orientation ***(19–21)***. Therefore, this strategy would not be suitable for large scale library approaches. The second limitation of this strategy is the use of *MlyI* restriction enzyme, which recognizes a five base pair sequence to perform blunt end cleavage. Therefore, frequent removal of *MlyI* sites from vector backbones, coding sequences, promoters, terminators, secretion signal and origin of replication may be required followed by analysis of unchanged functionality. Majority of the vectors currently available to achieve recombinant gene expression in *Pichia pastoris* are based on straight forward classical concepts e.g., plasmids available from Life Technologies™, Biogrammatics, DNA2.0 and Technical University of Graz ***(15, 22–25)***. Thus, the RSFC vectors constructed during this study would prove to be a valuable set of tools for production and characterization of recombinant proteins. Furthermore, we have used these expression vectors to investigate effects of promoter (P_{AOX1}, P_{GAP}), *Sc*_αMF secretion signal variants (with and without Glu-Ala repeats), fusion proteins (eGFP, MBP), N and C-terminal tags (Myc, FLAG, HIS, Strep) on horseradish peroxidase secretion in *P. pastoris*.

The final part of this thesis, deals with the development of a novel insertion mutagenesis method to randomly target and inactivate genomic loci in *P. pastoris*. This mutagenesis method was used in combination with targeted gene deletion, described in chapter 2, to further investigate the secretory pathway of this yeast using horseradish peroxidase as a reporter. It was possible to identify some connection between *HRP* secretion and the *SGT2* and *RIM101* genes, which has not been described before. Additionally, an orphan gene which we termed as *KEP1* (Knockout enhances protein secretion 1) was identified. Deletion of this gene enhanced protein secretion of alternative pig liver esterase and human growth hormone in addition to *HRP*. It remains to be seen whether this deletion would also increase secretion of other proteins than tested here. One promising strategy could be to combine these deletions in one strain to test the combinatorial effect on protein secretion

.

References:

1. M. Ahmad, M. Hirz, H. Pichler, et al. (2014), Protein expression in Pichia pastoris: Recent achievements and perspectives for heterologous protein production, http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4047484&tool=pmcentrez&rendertype=abstract.

2. H. Schwab, M. Ahmad, M. Hirz, et al. (2014) Developing a versatile tool set for heterologous gene expression in Pichia pastoris, New Biotechnology. 31, S190.

3. M. Romanos (1995) Advances in the use of Pichia pastoris for high-level gene expression, Current Opinion in biotechnology. 1–7.

4. R. Daly and M.T.W. Hearn (2005) Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production., Journal of molecular recognition : JMR. 18, 119–138.

5. A. Küberl, J. Schneider, G.G. Thallinger, et al. (2011) High-quality genome sequence of Pichia pastoris CBS7435., Journal of biotechnology. 154, 312–320.

6. J. Kurjan and I. Herskowitz (1982) Structure of a yeast pheromone gene (MFα): A putative α-factor precursor contains four tandem copies of mature α-factor, Cell. 30, 933–943.

7. M.W.T. Werten and F. a. De Wolf (2005) Reduced proteolysis of secreted gelatin and Yps1-mediated alpha-factor leader processing in a Pichia pastoris kex2 disruptant., Applied and environmental microbiology. 71, 2310–2317.

8. M. Egel-Mitani, H.P. Flygenring, and M.T. Hansen (1989) A novel aspartyl protease allowing KEX2-independent MF alpha propheromone processing in yeast., Yeast (Chichester, England). 6, 127–137.

9. H. Komano and R.S. Fuller (1995) Shared functions in vivo of a glycosyl-phosphatidylinositol-linked aspartyl protease, Mkc7, and the proprotein processing protease Kex2 in yeast., Proceedings of the National Academy of Sciences of the United States of America. 92, 10752–10756.

10. V. Olsen, N.X. Cawley, J. Brandt, et al. (1999) Identification and characterization of Saccharomyces cerevisiae yapsin 3, a new member of the yapsin family of aspartic proteases encoded by the YPS3 gene., Journal Biochemical Society. 339, 407–411.

11. R.J. Austin, R.E. Kuestner, D.K. Chang, et al. (2011) SILAC compatible strain of pichia pastoris for expression of isotopically labeled protein standards and quantitative proteomics, Journal of Proteome Research. 10, 5251–5259.

12. D. Thor, S. Xiong, C.C. Orazem, et al. (2005) Cloning and characterization of the Pichia pastoris MET2 gene as a selectable marker., FEMS yeast research. 5, 935–942.

13. M.M. Whittaker and J.W. Whittaker (2005) Construction and characterization of Pichia pastoris strains for labeling aromatic amino acids in recombinant proteins., Protein expression and purification. 41, 266–74.

14. G.H. Braus (1991) Aromatic amino acid biosynthesis in the yeast Saccharomyces cerevisiae: a model system for the regulation of a eukaryotic biosynthetic pathway., Microbiological reviews. 55, 349–70.

15. L. Näätsaari, B. Mistlberger, C. Ruth, et al. (2012) Deletion of the pichia pastoris ku70 homologue facilitates platform strain generation for gene expression and synthetic biology, PLoS ONE. 7,.

16. J. Ahn, J. Hong, H. Lee, et al. (2007) Translation elongation factor 1-alpha gene from Pichia pastoris: molecular cloning, sequence, and use of its promoter., Applied microbiology and biotechnology. 74, 601–608.

17. H. Dortay, U.M. Akula, C. Westphal, et al. (2011) High-throughput protein expression using a combination of ligation-independent cloning (LIC) and infrared fluorescent protein (IFP) detection., PloS one. 6, e18900.

18. R.N. de Jong, M.A. Daniëls, R. Kaptein, et al. (2006) Enzyme free cloning for high throughput gene cloning and expression., Journal of structural and functional genomics. 7, 109–18.

19. M.Z. Li and S.J. Elledge (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC., Nature methods. 4, 251–6.

20. J. Quan and J. Tian (2009) Circular polymerase extension cloning of complex gene libraries and pathways., PloS one. 4, e6441.

21. D.G. Gibson, L. Young, R. Chuang, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases., Nat Methods. 6, 343–5.

22. I.B. Sears, J. O'Connor, O.W. Rossanese, et al. (1998) A versatile set of vectors for constitutive and regulated gene expression in Pichia pastoris., Yeast (Chichester, England). 14, 783–790.

23. B. Rao, X. Zhong, Y. Wang, et al. (2010) Efficient vectors for expression cloning of large numbers of PCR fragments in P . pastoris, Yeast. 285–292.

24. D. Lan, W. Wang, L. Sun, et al. (2011) A novel vector for expression cloning of large numbers of polymerase chain reaction products in Pichia pastoris, 5, 211–216.

25. T. Sasagawa, M. Matsui, Y. Kobayashi, et al. (2011) High-throughput recombinant gene expression systems in Pichia pastoris using newly developed plasmid vectors., Plasmid. 65, 65–69.

EUROPEAN
CURRICULUM VITAE
FORMAT

CURRICULUM VITAE

PERSONAL INFORMATION

Name	**Mudassar Ahmad**
Address	101C-1511 McKercher Dr. Saskatoon, SK, Canada
Telephone	+1 306 477 9494
Mobile phone	+1 306 222 5711
E-mail	**Mudassar.ahmad@hotmail.com**
Nationality	Pakistan
Date of birth	18-01-1982

PROFESSIONAL EXPERIENCE

Dates (from – to)	since 2010
Name and address of employer	Graz University of Technology, Graz, Austria
Type of business or sector	University research
Occupation or position held	PhD student
Main activities and responsibilities	PhD Student
Dates (from – to)	01-2007 till 03-2009
Name and address of employer	National Institute for Molecular Biotechnology and Genetic Engineering, Faisalabad, Pakistan
Type of business or sector	Research Institute
Occupation or position held	Research Officer
Main activities and responsibilities	R & D in Biotechnology (Plant Biotechnology)

EDUCATION AND TRAINING

Dates (from – to)	03-2010, expected completion by 10-2015
Name and type of organization providing education and training	Graz University of Technology, Graz, Austria,
Principal subjects/occupational skills covered	PhD study Molecular cloning, strain and vector construction, engineering of Pichia genome for recombinant protein production, protein expression and purification, scientific writing, supervision of undergraduate students

Title of qualification awarded	Dr. rer. tech. (expected)
Dates (from – to)	2005-2007
Name and type of organization providing education and training	National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan and Quid-e-Azam University Islamabad, Pakistan
Principal subjects/occupational skills covered	Biotechnology/Student Biotechnology Related Skills
Title of qualification awarded	M.Phil. Biotechnology
Dates (from – to)	2000- 2004
Name and type of organization providing education and training	Agriculture University of Faisalabad, Punjab, Pakistan
Principal subjects/occupational skills covered	Bachelor curriculum in plant breeding and genetics Basic knowledge of biology and supporting subjects
Title of qualification awarded	B. Sc. (Hons) Agriculture with Major in Plant Breeding and Genetics

PERSONAL SKILLS AND COMPETENCES

MOTHER TONGUE	**Urdu, Punjabi**
OTHER LANGUAGES	
	English
• Reading skills	excellent
• Writing skills	I wish they were excellent.
• Verbal skills	Good
SOCIAL SKILLS	Almost None
ORGANIZATIONAL SKILLS AND COMPETENCES	**Teaching and organizing courses for students**
TECHNICAL SKILLS AND COMPETENCES	Laboratory skills: recombinant protein expression and secretion in *Pichia pastoris* Conventional microbiology methods; protein analysis: SDS PAGE, Western blot, protein activity assays, protein expression in bacteria; protein purification; ELISA; cell fractionation techniques; molecular biology and cloning methods: PCR methods; Northern and Southern blots, gene expression, gene tagging, gene deletion, mutagenesis; *in vivo* cell imaging and microscopy techniques, Teaching and tutor experience; scientific writing.

TEACHING ACTIVITIES

Dates (from – to)	01-2011 - 03-2014
Name of organization	Graz University of Technology, Institute for Molecular Biotechnology
Position	Laboratory project supervisor
Type and title of course	Laboratory project for master students: recombinant protein secretion in *P.pastoris*

PUBLICATIONS

Ahmad M*, Hirz M*, Pichler H, Schwab H (2014) Protein expression in Pichia pastoris: Recent achievements and perspectives for heterologous protein production. Appl. Microbiol. Biotechnol. 98:5301–5317.(*equal contribution)

Schwab H, **Ahmad M**, Hirz M, Kolmbauer M, Rodriguez IR (2014) Developing a versatile tool set for heterologous gene expression in Pichia pastoris. N Biotechnol 31:S190. doi: 10.1016/j.nbt.2014.05.933

Vogl T*, **Ahmad M***, Krainer FW, Schwab H, Glieder A (2015) Restriction site free cloning (RSFC) plasmid family for seamless, sequence independent cloning in Pichia pastoris. Microb Cell Fact 14:103. doi: 10.1186/s12934-015-0293-6 (*equal contribution)

Appendix:

#	IMBT Culture Collection #	Expression Vectors and Strains
1	3097	pAHBgl
2	3098	pAHSph
3	6413	pAaHBgl
4	6414	pAaHSph
5	6415	pAHSwa
6	6416	pAaHSwa
7	6417	pAABgl
8	6418	pAaABgl
9	6419	pAASph
10	6420	pAaASph
11	6421	pAASwa
12	6422	pAaASwa
13	6423	pAZBgl
14	6424	pAaZBgl
15	6425	pAZSph
16	6426	pAaZSph
17	6427	pAZSwa
18	6428	pAaZSwa
19	6429	pAKBgl
20	6430	pAaKBgl
21	6431	pAKSph
22	6432	pAaKSph
23	6433	pAKSwa
24	6434	pAaKSwa
25	6435	pGaHSwa
26	6434	pAaKSwa
27	6435	pGaHSwa
28	7038	CBS7435 Δ*his4* transformed with pGaHSwaLev (Mut^S)
29	7039	CBS7435 Δ*his4* transformed with pAHBglLev (Mut^S)
30	7040	CBS7435 Δ*his4* transformed with pAaHBglLev (Mut^S)
31	7041	CBS7435 Δ*his4* transformed with pAaHBgl2EALev (Mut^S)
32	7042	CBS7435 Δ*his4* transformed with pAaHBgl5EALev (Mut^S)
33	7043	CBS7435 Δ*his4* transformed with pAHBglPpaLev (Mut^S)
34	7044	CBS7435 Δ*his4* transformed with pAHBglPpaEALev (Mut^S)
35	7045	CBS7435 Δ*his4* transformed with pAHBglPpa5EALev (Mut^S)
36	6602	pPpKC1 (Zeocin™)
37	7591	pPpKC2 (G418)
38	7592	pPpKC3 (*HIS4*)
39	7593	pPpKC4 (*ARG4*)
40	6603	pPpKC1 S2*Amp*
41	6604	pPpKC1 Aro7
42	6605	pPpKC1 Trp3

#	IMBT Culture Collection #	Expression Vectors and Strains
43	6639	pPpKC1 Lys2
44	6640	pPpKC1 Met2
45	6641	pPpKC1 Trp5
46	6642	pPpKC1 Pha2
47	6643	pPpKC1 Tyr1
48	6644	pPpKC1 Pro3
49	6668	CBS7435 Δsub2
50	6669	CBS7435 Δkpx1
51	6670	CBS7435 Δkpx8
52	6671	CBS7435 Δctse1
53	6672	CBS7435 Δkpx9
54	6673	CBS7435 ΔprtP
55	6674	CBS7435 Δkpx13
56	6675	CBS7435 Δkpx14
57	6676	CBS7435 Δprc1
58	6677	CBS7435 Δkpx17
59	6678	CBS7435 Δkpx18
60	6679	CBS7435 Δkpx19
61	6680	CBS7435 Δkpx20
62	6681	CBS7435 Δkpx21
63	6682	CBS7435 Δkpx22
64	6683	CBS7435 Δkpx23
65	6684	CBS7435 Δkpx24
66	6685	CBS7435 Δkpx25
67	6686	CBS7435 Δyps1
68	6687	CBS7435 Δctse2
69	6768	pPpKC1_sub2
70	6769	pPpKC1_kpx1
71	6770	pPpKC1_kpx2
72	6771	pPpKC1_kpx3
73	6772	pPpKC1_kpx4
74	6773	pPpKC1_pep4
75	6774	pPpKC1_kpx5
76	6775	pPpKC1_kpx6
77	6776	pPpKC1_prb1
78	6777	pPpKC1_kpx7
79	6778	pPpKC1_kpx8
80	6779	pPpKC1_ctse1
81	6782	pPpKC1_ctsd
82	6783	pPpKC1_kpx9
83	6784	pPpKC1_kpx10
84	6785	pPpKC1_kpx11
85	6786	pPpKC1_prtP

#	**IMBT Culture Collection #**	**Expression Vectors and Strains**
86	6787	pPpKC1_kpx12
87	6788	pPpKC1_kpx13
88	6789	pPpKC1_kpx14
89	6790	pPpKC1_kpx15
90	6791	pPpKC1_prc1
91	6792	pPpKC1_kpx17
92	6793	pPpKC1_kpx18
93	6794	pPpKC1_kpx19
94	6795	pPpKC1_kpx20
95	6796	pPpKC1_kpx21
96	6825	pPpKC1_kpx22
97	6826	pPpKC1_kpx23
98	6827	pPpKC1_kpx24
99	6828	pPpKC1_yps1
100	6829	pPpKC1_ctse2
101	6830	pPpKC1_kex1
102	6831	pPpKC1_kex2
103	6832	pPpKC1_kpx25
104	6906	CBS7435 *Δkpx4*
105	6907	CBS7435 *Δctsd*
106	6908	CBS7435 *Δkpx10*
107	6909	CBS7435 *Δkex1*
108	6910	CBS7435 *Δkex2*
109	6911	CBS7435 *Δpep4*
110	6912	CBS7435 *Δprb1*
111	6913	CBS7435 *Δhis4 Δkpx5*
112	6914	CBS7435 *Δhis4 Δkpx7*
113	7013	CBS7435 *Δhis4 Δpep4 Δprb1*
114	7014	CBS7435 *Δhis4 Δpep4 Δprb1 Δaox1*
115	7015	CBS7435 *Δhis4 Δpep4 Δaox1*
116	7016	CBS7435 *Δprb1Δaox1*
117	7017	CBS7435 *Δyps1 Δaox1*
118	7018	CBS7435 *Δyps7 Δaox1*
119	7019	CBS7435 *Δprc1 Δaox1*
120	7020	CBS7435 *Δkex1 Δaox1*
121	7021	CBS7435 *Δmet2*
122	7022	CBS7435 *Δmet2 Δarg4*
123	7023	CBS7435 *Δmet2 Δhis4*
124	7024	CBS7435 *Δlys2*
125	7025	CBS7435 *Δlys2 Δarg4*
126	7026	CBS7435 *Δlys2 Δhis4*
127	7027	CBS7435 *Δpro3*
128	7028	CBS7435 *Δtyr1*

#	**IMBT Culture Collection #**	**Expression Vectors and Strains**
129	7029	CBS7435 *Δpha2*
130	7030	CBS7435 *Δmet2*
131	7031	CBS7435 *Δmet2 Δarg4*
132	7032	CBS7435 *Δmet2 Δhis4*
133	7033	CBS7435 *Δlys2*
134	7034	CBS7435 *Δlys2 Δarg4*
135	7035	CBS7435 *Δlys2 Δarg4*
136	7036	CBS7435 *Δpro3*
137	7037	CBS7435 *Δtyr1*
138	7076	CBS7435 *Δyps2 Δyps1*
139	7077	CBS7435 *Δyps7 Δyps2*
140	7078	CBS7435 *Δkex2 Δyps1 Δhis4*
141	7079	CBS7435 *Δkex2 Δpep4 Δhis4*
142	7080	CBS7435 *Δkex2 Δpep4 Δprb1 Δhis4*
143	7081	CBS7435 *Δaox1 Δaox2*
144	7082	CBS7435 transformed with pAaHBgl2EAHRP0 (Mut^S)
145	7083	CBS7435 transformed with pAaHBgl5EAHRP0 (Mut^S)
146	7084	CBS7435 transformed with pAHBglppa5EAHRP0 (Mut^S)
147	7085	CBS7435 *iLvE* Mut^+ single copy
148	7086	CBS7435 *iLvE* Mut^S single copy
149	7087	CBS7435 *iLvE* Mut^S single copy
150	7088	CBS7435 *iLvE* Mut^S single copy
151	7089	CBS7435 *iLvE* Mut^S three copies
152	7090	CBS7435 *iLvE* Mut^S 4 copies
153	7091	CBS7435 *iLvE* Mut^S 7 copies
154	7092	CBS7435 *iLvE* intracellular Mut^S 1 copy
155	7093	CBS7435 *iLvE* intracellular Mut^S single copy
156	7094	CBS7435 *iLvE* intracellular Mut^S 4 copies
157	7095	CBS7435 *iLvE* intracellular Mut^S 14 copies
158	7096	CBS7435 *iLvE* intracellular Mut^S 22 copies
159	6480	pPpRSFC
160	6481	pPpRSFC_alpha
161	6482	pPpRSFC_alpha.noEAEA
162	6483	pPpRSFC_GAP
163	6484	pPpRSFC_GAP_alpha
164	6485	pPpRSFC_GAP_alpha.noEAEA
165	6486	pPpRSFC-Empty Vector
166	6487	pPpRSFC_GAP-Empty Vector
167	6712	pPpRSFC_N.EGFP [7P]
168	6713	pPpRSFC_C.EGFP [8P]
169	6714	pPpRSFC_alpha_N.EGFP [9P]
170	6715	pPpRSFC_alpha_C.EGFP [10P]
171	6716	pPpRSFC_N.MYC [11P]

#	IMBT Culture Collection #	Expression Vectors and Strains
172	6717	pPpRSFC_C.MYC [12P]
173	6718	pPpRSFC_alpha_N.MYC [13P]
174	6719	pPpRSFC_alpha_C.MYC [14P]
175	6720	pPpRSFC_N.FLAG [15P]
176	6721	pPpRSFC_C.FLAG [16P]
177	6722	pPpRSFC_alpha_N.FLAG [17P]
178	6723	pPpRSFC_alpha_C.FLAG [18P]
179	6724	pPpRSFC_N.HIS.ncs [19P]
180	6725	pPpRSFC_C.HIS.ncs [20P]
181	6726	pPpRSFC_alpha_N.HIS.ncs [21P]
182	6727	pPpRSFC_alpha_C.HIS.ncs [22P]
183	6728	pPpRSFC_N.HIS [23P]
184	6729	pPpRSFC_C.HIS [24P]
185	6730	pPpRSFC_alpha_N.HIS [25P]
186	6731	pPpRSFC_alpha_C.HIS [26P]
187	6732	pPpRSFC_N.MBP [27P]
188	6733	pPpRSFC_C.MBP [28P]
189	6734	pPpRSFC_alpha_N.MBP [29P]
190	6735	pPpRSFC_alpha_C.MBP [30P]
191	6736	pPpRSFC_N.STREP [31P]
192	6737	pPpRSFC_C.STREP [32P]
193	6738	pPpRSFC_alpha_N.STREP [33P]
194	6739	pPpRSFC_alpha_C.STREP [34P]
195	6740	pPpRSFC-HRPA2 [1P-HRPA2]
196	6741	pPpRSFC_alpha-HRPA2 [2P-HRPA2]
197	6742	pPpRSFC_alpha.noEAEA-HRPA2 [3P-HRPA2]
198	6743	pPpRSFC_GAP_alpha-HRPA2 [5P-HRPA2]
199	6744	pPpRSFC_GAP_alpha.noEAEA-HRPA2 [6P-HRPA2]
200	6745	pPpRSFC_alpha_N.EGFP-HRPA2 [9P-HRPA2]
201	6746	pPpRSFC_alpha_C.EGFP-HRP [10P-HRPA2]
202	6747	pPpRSFC_alpha_N.MYC-HRPA2 [13 P-HRPA2]
203	6748	pPpRSFC_alpha_C.MYC-HRP [14 P-HRPA2]
204	6749	pPpRSFC_alpha_N.FLAG-HRPA2 [17P-HRPA2]
205	6750	pPpRSFC_alpha_C.FLAG-HRPA2 [18P-HRPA2]
206	6751	pPpRSFC_alpha_N.HIS.ncs-HRPA2 [21P-HRPA2]
207	6752	pPpRSFC_alpha_C.HIS.ncs-HRPA2 [22P-HRPA2]
208	6753	pPpRSFC_alpha_N.HIS-HRPA2 [25P-HRPA2]
209	6754	pPpRSFC_alpha_C.HIS-HRPA2 [26P-HRPA2]
210	6755	pPpRSFC_alpha_N.MBP-HRP [29P-HRPA2]
211	6756	pPpRSFC_alpha_C.MBP-HRPA2 [30P-HRPA2]
212	6757	pPpRSFC_alpha_N.STREP-HRPA2 [33P-HRPA2]
213	6758	pPpRSFC_alpha_C.STREP-HRP [34P-HRPA2]
214	6759	pPpRSFC-EGFP [1P-EGFP]

#	**IMBT Culture Collection #**	**Expression Vectors and Strains**
215	6760	pPpRSFC_GAP-EGFP [4P-EGFP]
216	6761	pPpRSFC_alpha_N.EGFP_self ligated [9P self-ligated]
217	6762	pPpRSFC_HIS [1PHIS]
218	6763	pPpRSFC_HIS_alpha [2PHIS]
219	6764	pPpRSFC_HIS_alpha.noEAEA [3PHIS]
220	6765	pPpRSFC_HIS_GAP [4PHIS]
221	6766	pPpRSFC_HIS_GAP_alpha [5PHIS]
222	6767	pPpRSFC_HIS_GAP_alpha.noEAEA [6PHIS]