PROBING THE SECRETORY PATHWAY OF THE YEAST *PICHIA PASTORIS*

Dissertation

Dipl. Ing. Christine Winkler

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

Ass. Prof. Dipl.-Ing. Dr. techn. Harald Pichler

Institut für Molekulare Biotechnologie

Technische Universität Graz

2014

Acknowledgements

Öffnet man die Augen, wird jeder Tag zum Erlebnis. *Oskar Kokoschka*

I pinned this sentence on the board above my desk a while ago, to remember myself of the adventures that a regular day at the bench and in the office could offer. Of course, many days went by since the beginning of my work as a PhD student, and I won't remember all of them. But I will remember some, like the day when I sent the samples obtained from genome walking to the sequencing facility, and the thrill of going through the first sequencing results some days later.

And I will remember the good times I had with my colleagues in the lab, in the office, in Graz and in Berkeley, and in the, sometimes pretty long, nights spent chatting in the dining area. I want to thank all these people for the nice experiences I had during my PhD time, and also as a diploma student at the institute.

Special thanks go to my supervisors Dr. Harald Pichler and Prof. Helmut Schwab for giving support and advice, and being open for discussions, also in times were 'time' was limited. I wish to thank Mudassar Ahmad and my diploma and project students, Peter Neugebauer, Thomas Vogl, Ariane Fankl, Anna Winkler, Thomas Kickenweiz, Altijana Hromic and Peter Augustin for their support in the lab and constant encouragement.

To my parents and sisters and family – thanks a lot for all the support I got in every conceivable way!

Table of contents

Abstract

The yeast *Pichia pastoris* is widely used as a host for the production of recombinant protein. An evident benefit of the *Pichia* expression system is that it has been highly successful in the secretion of proteins to the culture medium, a prerequisite to simple purification of the desired product. Nevertheless, secretion is often the limiting factor in protein production, and many of the determinants affecting secretion are still unknown. The aim of this thesis was to unravel factors involved in the secretion of recombinant protein in *P. pastoris*.

Here, a novel insertion mutagenesis method for random targeting of genes in *P. pastoris* is presented. Mutant colonies were generated by integration of a linear DNA marker cassette that allowed for easy recovery of DNA regions flanking the insertion loci. The industrially important enzyme horseradish peroxidase (HRP) was employed as a reporter protein to screen for altered secretion behavior. Using this strategy, a set of genes was identified that, when knocked-out, positively or negatively affected the amount of HRP secreted to the culture medium. This strategy identified a number of highly interesting genome loci of which many had not been associated with recombinant protein secretion before.

To confirm the results of the random approach, selected genes were targeted with a novel vector-based system for the generation of knockouts in *P. pastoris*. A detailed description of this knockout vector system is also part of this thesis.

Kurzfassung

Die Hefe *Pichia pastoris* wird oft als Wirtsorganismus für die Expression von rekombinantem Protein herangezogen. Einer der größten Vorteile des *Pichia* Expressionssystems ist die effiziente Sekretion des produzierten Proteins in das Kulturmedium, da die folgenden Produktaufreinigungsschritte dadurch massiv erleichtert werden. Jedoch ist Sekretion aus der Zelle in vielen Fällen der limitierende Schritt in der Produktion von rekombinantem Protein und kann daher die tatsächliche Ausbeute schmälern. Über die Faktoren, die Proteinsekretion in *P. pastoris* beeinflussen, ist noch nicht viel bekannt. Ziel meiner Dissertationsarbeit war es, einige dieser Faktoren zu identifizieren und näher zu beschreiben. Die daraus gewonnenen Erkenntnisse sollen dazu dienen, rekombinante Proteinproduktion in *P. pastoris* effizienter zu machen.

Mit dem Ziel Einflussfaktoren auf genomischer Ebene zu identifizieren, entwickelten wir eine neue Methode zur zufallsbasierten Mutagenese von *P. pastoris* Zellen, die im zweiten Teil dieser Arbeit erstmalig beschrieben wird. Diese Methode verwendet lineare DNA Selektionskassetten, die an zufälligen Positionen in die genomische DNA der Hefe integrieren. Die integrierten DNA Kassetten ermöglichen im nächsten Schritt die einfache Identifizierung des mutierten Genlokus. Das Enzym Horseradish peroxidase (HRP) wurde als sekretiertes Reporterprotein verwendet, um den Einfluss des mutierten Gens auf die Sekretionsleistung der Zellen zu bestimmen. Mithilfe dieser Strategie konnten wir mehrere Gene identifizieren, deren Inaktivierung die Sekretion von HRP in den Kulturüberstand positiv oder negativ beeinflusste. Besonders interessant ist dabei, dass viele dieser Gene bis zu diesem Zeitpunkt noch nicht mit rekombinanter Proteinsekretion in Verbindung gebracht worden sind.

Die Ergebnisse der zufallsbasierten Methode wurden im weiteren Verlauf des Projekts durch gezieltes Inaktivieren der zuvor identifizierten Gene bestätigt. Das zu diesem Zweck entwickelte Vektorsystem wird in dieser Arbeit ebenfalls detailliert erläutert.

Introduction

A short summary of the project

Recombinant proteins produced in the methylotrophic yeast *P. pastoris* are often targeted to the secretory pathway. Secretion of proteins enables the addition of post-translational modifications and facilitates easier recovery of the recombinant product from the culture supernatant. However, the interplay between recombinant protein expression and the yeast's secretory machinery is poorly understood. As a consequence, the bottlenecks for the secretory production of a specific protein are hard to identify and even harder to circumvent. The aim of this thesis was to improve the understanding of the molecular machinery involved in protein secretion by random mutagenesis of *P. pastoris* strains secreting recombinant model proteins. In the beginning, we tried to establish a strategy for random mutagenesis of *P. pastoris* using transposons. While testing various transposon systems, we discovered that linear DNA cassettes, having no sequence homologous to the host, efficiently and randomly integrated into the *P. pastoris* genome. We, hence, proceeded with the generation of mutant libraries using this method. The effect of the introduced mutations was assessed by screening for altered secretion levels of the model proteins. We prepared to screen several thousands of mutant colonies in order to cover a significant part of the yeast's genome. The large screening effort demanded a fast and reliable assay. The colorimetric assays for the enzymes horseradish peroxidase (HRP) and alternative pig liver esterase (APLE) met these requirements. In an alternative approach, we quantified the secretion of recombinant *S. cerevisiae* α-mating factor pheromone from *P. pastoris* cells. However, the screening assay developed to measure α-mating factor turned out to be unsuitable for the required high throughput.

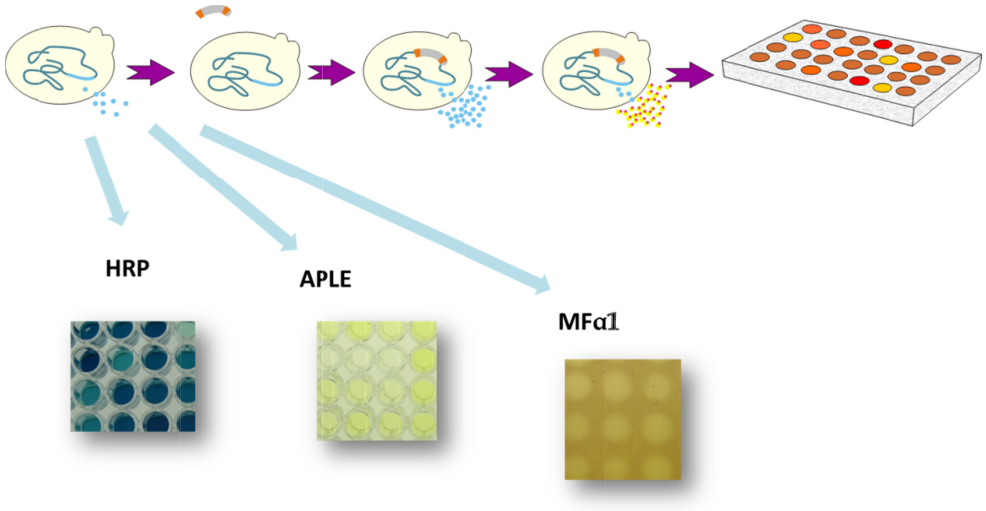


Figure 1 Scheme of random mutagenesis and screening. Strains secreting the model proteins HRP, APLE or α-mating factor (MFα1) were mutagenized by the random integration of a linear DNA marker cassette. Integration of the cassette can lead to gene disruption and, subsequently, to altered secretion levels of the model proteins. Significantly increased or decreased levels of secreted protein in the culture supernatants were detected in screening assays specific for the model proteins.

Subsequent genome walking allowed the identification of genes that, when disrupted, had a positive or negative effect on the secretion of HRP and APLE, respectively. In the next step, the previously observed secretion phenotypes were confirmed by constructing "clean knockout strains" lacking selected genes in the wild type strain background. Expression of secreted model proteins in these knockout strains led us to the highly interesting loci *RIM101*, *SGT2* and *KEP1*. Knockout of these genes reproducibly benefited the secretion of HRP, and in the case of *KEP1*, also other tested proteins. These three genes were initially identified in the screening for altered levels of HRP secretion. The genes selected in the screening using APLE as secreted model protein, on the other hand, yielded inconsistent and less reproducible results. Several of the affected genes were identified as coding for transcription factors, and possibly their molecular function led to the observed variations. To summarize, the work presented in this thesis shed light on effectors of recombinant protein secretion, which would probably not have been identified with conventional, non-random, approaches.

The secretory pathway

In eukaryotic cells, many newly synthesized proteins have to be delivered to different compartments of the cell. The process of transporting them to their ultimate destination at the cell surface or other subcellular compartments is known as protein secretion. Secretory proteins, also termed cargo, enter the endoplasmic reticulum (ER) where they are folded and receive first post-translational modifications. These primary modifications include signal sequence processing, disulfide bond formation, *N*-glycosylation, glycosyl-phosphatidyl-inositol addition, degradation, and sorting (Idiris et al. 2010). From the ER, secretory cargo is transported to the Golgi apparatus, which consists of an array of disk-shaped membranes called cisternae. Passing through the Golgi cisternae in anterograde (*cis* to *trans*) direction, the cargo proteins are modified and processed by Golgi-resident enzymes in an ordered manner. Posttranslational modifications taking place in the Golgi include proteolytic processing and addition of glycosyl residues. At the final Golgi compartment, the *trans*-Golgi network (TGN), proteins are packaged into vesicular carriers and transported to their ultimate localization in the cell or to the cell surface.

George Palade's pulse-chase autoradiographic tracing of newly synthesized secretory proteins in cells of the exocrine pancreas drew a basic chart of the secretory pathway (reviewed in Palade's Nobel lecture 1974). His efforts demonstrated that proteins are synthesized in the ER, from where they travel onwards through the Golgi apparatus and storage granules, before being released to the extracellular space. Palade further proposed that secretory cargo is conveyed between membrane-enclosed compartments in vesicular carriers. His crude description of the secretory pathway framed new questions about how these vesicular carriers are formed and how they fuse with the destination compartment, and, subsequently, Randy Schekman and James Rothman independently set out to unravel the molecular machinery that drives protein secretion.

Randy Schekman initially chose *Saccharomyces cerevisiae* and a genetic approach to dissect components associated with protein trafficking. Together with his graduate student Peter Novick he identified the first temperature-sensitive (*ts*) secretion mutant, *sec1*, which was defective in the last step of secretion and hence accumulated intracellular secretory vesicles at the restrictive temperature (Novick and Schekman, 1979). The observation that this accumulation of cargo rendered the cells denser led to the identification of another 23 *ts*-mutants (Novick et al., 1980). During the following years, dozens of *sec* mutants defective in different steps of the secretory

pathway were isolated in the Schekman laboratory. Among them were the components of the COPII coat complex, trafficking between ER and Golgi (reviewed in Jensen and Schekman, 2011). Functionality of this coat was ultimately proven by reconstituting vesicle formation with purified protein components in a biochemical assay (Barlowe et al., 1994).

James Rothman decided to use biochemical methods from the start. He developed a cell-free assay to reconstitute the transport of secretory cargo between isolated mammalian Golgi compartments. A central component of this system was the viral protein VSV-G, whose transfer between donor and acceptor Golgi-compartments was measured by the incorporation of a radiolabeled sugar residue at its destination (Balch et al., 1984). Investigations on the mechanism of vesicle formation and fusion led to the discovery of the proteins NSF (N-ethylmaleimide-sensitive factor) and SNAP (soluble NSF-attachment protein) (Block et al., 1988; Malhotra et al., 1988; Clary et al., 1990). Remarkably, the two different approaches of Rothman and Schekman converged when it became evident that NSF and α-SNAP were orthologous to Sec18 and Sec17, respectively (Eakle et al., 1988; Wilson et al., 1989; Griff et al., 1992). This finding highlighted the evolutionary conservation of the secretory components between yeast and mammals, and underlined the strengths of the two genetic and biochemical approaches. Finally, Rothman's lab identified the proteins that mediate vesicle fusion with a specific target membrane, the SNAREs (SNAP receptors) (Söllner et al., 1993). SNAREs are organelle-specific receptors that interact in the so-called "SNARE complex" to promote fusion of transport vesicles with acceptor organelles.

Organisation of the secretory pathway in *P. pastoris*

Yeasts are an attractive system to study secretory mechanisms. The yeast secretory pathway is relatively simple, but the basic molecular aspects are conserved between yeasts and mammals. Moreover, efficient genetic and biochemical tools exist in yeast to study protein traffic. The group of Benjamin Glick uses *P. pastoris* as a model organism to study biogenesis of the Golgi apparatus (reviewed in Glick and Nakano, 2009; Papanikou and Glick, 2009; Suda and Nakano, 2012). The most obvious difference in the secretory machinery between the two budding yeasts, *P. pastoris* and *S. cerevisiae*, is that *P. pastoris* contains a Golgi arranged into stacks, resembling the mammalian Golgi, while the Golgi cisternae of *S. cerevisiae* appear scattered throughout the cytoplasm (Rossanese et al., 1999; Mogelsvang et al., 2003). A typical *P. pastoris* cell contains 2–

5 Golgi stacks that localize close to discrete transitional endoplasmic reticulum (tER) sites, also known as ER-exit sites (ERES). Each stack consists of about 4 cisternae, which are biochemically polarized as *cis*, *medial*, *trans*, or TGN (Mogelsvang et al. 2003). The reason why *S. cerevisiae* shows a different Golgi organization is unknown. However, more recent studies indicated that differences in the Golgi organization between the two yeasts are superficial. Connerly et al. (2005) identified a temperature-sensitive mutation of Sec16, which at the restrictive temperature caused tER fragmentation and disruption of Golgi stacking in *P. pastoris* similar to the pattern seen in *S. cerevisiae*. On the other hand, slowed ER export in *S. cerevisiae* induced by glucose deprivation produced a tER pattern similar to that in *P. pastoris* (Levi et al., 2010). These findings drew a general picture of the secretory pathway organization in *P. pastoris*. They also made clear that further experiments are required to dissect the mechanism of protein secretion in this yeast in more detail. We hope that the results presented here can contribute to the understanding of protein secretion in *P. pastoris*.

P. pastoris as expression host for recombinant protein

P. pastoris has many features which qualify it as an ideal production host for recombinant protein expression, i.e. easy genetic manipulation, growth to high cell densities, the availability of strong inducible promoters and the possibility of post-translational modifications of expressed proteins. Various recombinant, heterologous proteins from all kingdoms of life have been expressed successfully in *P. pastoris,* many of them in large-scale industrial processes (Ahmad et al., in revision).

In the 1970s, Phillips Petroleum Company evaluated *P. pastoris* as source for single-cell protein, due to its ability to grow on methanol as sole carbon source. The oil crisis in 1973, however, increased prices for the source of methanol, methane, and the process became uneconomical (Cregg et al., 2000). Discovery of the strong, inducible *AOX1* promoter system in the 1980s led to the return of *P. pastoris* as a host for recombinant protein production (Cregg et al., 1985). The characteristics of the *AOX1* promoter justifying its extensive use in recombinant protein production are the high titers of soluble protein that can be achieved, and the strong regulation at the transcriptional level. The promoter is tightly repressed in the presence of glucose, glycerol or ethanol. Depletion of these carbon sources leads to de-repression, but substantial transcription is

only achieved after induction with methanol (reviewed in Cregg et al., 2000). In a typical *AOX1* promoter driven expression protocol the cells are first grown on a repressing carbon source for accumulation of biomass prior to induction with methanol. Using this two-step protocol, even toxic proteins can be expressed from P_{AOX1}.

The genes *AOX1* and *AOX2* encode alcohol oxidase, the enzyme catalyzing the first step in the methanol utilization pathway, namely oxidation of methanol to formaldehyde. A side product of this step, hydrogen peroxide, is potentially toxic to the cell. To avoid deleterious effects on the cell, the reaction takes place in specialized organelles, the peroxisomes. Upon shift to methanol, peroxisomes proliferate rapidly, but are degraded after the shift to another carbon source. This feature makes *P. pastoris* a popular system for studying the biogenesis and degradation of this organelle (Dunn et al., 2005; Joshi et al., 2011). The vast majority of alcohol oxidase enzyme is expressed from *AOX1*, while expression from *AOX2* is much weaker (Cregg et al., 1989). As a result, *aox1* knockout strains can grow on methanol, although at a reduced rate. Strains with this phenotype, termed methanol utilization slow (Mut^S), were reported to be better producers for some recombinant proteins, possibly due to their slower growth rate (reviewed in Macauley-Patrick et al., 2005). An obvious advantage of Mut^S cells is that they require less methanol for induction.

Aside from P_{AOX1}, we employed P_{GAP}, derived from the *P. pastoris* glyceraldehyde-3-phosphate dehydrogenase gene, for the expression of recombinant secreted proteins (Waterham et al., 1997). Expression from this promoter is constitutive and does not require shifting the cells to another carbon source, which makes this protocol more convenient to handle. However, this feature also renders it unsuitable for the expression of proteins potentially toxic to the host.

Strain engineering to improve secretion of recombinant protein in *P. pastoris*

Secretion of the produced protein is an attractive alternative to intracellular expression, as *P. pastoris* secretes only low levels of endogenous protein (Mattanovich et al., 2009). Secretion from the cell can therefore be seen as a first purification step, as the secreted recombinant product accounts for the majority of protein in the culture supernatant. Targeting of the recombinant product to the secretory pathway permits posttranslational modifications such as the formation of disulfide bridges or glycosylation to take place. Moreover, it prevents the accumulation of

potentially toxic proteins inside the cell. For many recombinant proteins secretion from the cell is still one of the limiting factors in high-yield production and a challenging step to manipulate as many cross-reacting components are involved. Cell engineering strategies are largely aimed at the main bottlenecks of the secretory pathway: protein folding, trafficking and glycosylation (reviewed in Idiris et al. 2010; Damasceno et al. 2011).

The folding capacity of the ER can potentially be increased by overexpressing ER-resident chaperones and isomerases, e.g. immunoglobulin-binding protein (BiP/Kar2p) and protein disulfide isomerase (PDI). Damasceno et al. (2007), Inan et al. (2006) and Zhang et al. (2006) reported a positive effect of chaperone co-overexpression on product yields. However, the effect appears to depend on the nature of the recombinant product and copy numbers of the integrated expression cassette (Butz et al., 2003; Wu et al., 2014). Overexpression of BiP was even reported to exhibit a negative effect on protein secretion in some cases. A possible explanation for this effect is that excess of one chaperone could mess up the cascade of folding helpers acting later on the nascent protein (reviewed in Idiris et al., 2010).

Accumulation of unfolded protein in the ER can trigger distinct stress-response pathways. Unfolded protein response (UPR) slows down general translation, while it enhances chaperone expression (Bernales et al., 2006). ER-associated protein degradation (ERAD) eliminates misfolded protein (Wu and Kaufman, 2006). Activation of UPR by overexpression of the involved transcription factor Hac1increased titers of recombinant protein for Guerfal et al. (2010) and Gasser et al. (2007). However, Hac1-overexpression did not benefit the secretion of all the recombinant proteins that were tested, as reviewed in Idiris et al. (2010). The interaction of UPR with other regulatory pathways might be too complex to guarantee universal positive effects on protein secretion.

To summarize, overexpression of proteins involved in the secretory pathway yielded conflicting results benefiting the secretion of some recombinant model proteins, while not affecting or even decreasing the secretion of others. This effect could be explained by the additional burden put on the host cells by overexpressing two proteins at the same time. In addition, the decision to overexpress a proposed secretion helper is mostly based on knowledge derived from its function in *S. cerevisiae*. This reasoning does not take potential functional or regulatory differences between the two yeasts, *S. cerevisiae* and *P. pastoris*, into account. At this point, it is worth mentioning that

the reported titers of secreted recombinant protein are much higher for *P. pastoris* than for *S. cerevisiae* (Schmidt, 2004). This suggests that *P. pastoris* has certain additional secretion enhancing features which are not present in baker's yeast. A suitable approach to identify factors that are unique to *P. pastoris* is random mutagenesis paired with the subsequent screening for interesting mutant phenotypes, as it was performed in our lab.

Random and targeted mutagenesis in *P. pastoris*

During the last years, the genome sequences of the important *P. pastoris* production strains GS115, DSMZ 70382 and CBS7435, have become available (De Schutter et al., 2009; Mattanovich et al., 2009; Küberl et al., 2011). The newly gained information makes *P. pastoris* accessible for strain engineering by targeted and random mutagenesis methods.

Targeted gene knockouts play an important role in the study of gene function. Genes are disrupted by linear DNA cassettes that replace the targeted locus *in vivo* by homologous recombination. However, the targeting of genes in *P. pastoris* has proven to be problematic. The efficiency of gene deletion is extremely low, due to the prevailing non-homologous end joining (NHEJ) mechanism for DNA repair, which requires little to no sequence homology to operate. A detailed discussion of this mechanism and various strategies for gene targeting in *P. pastoris* are presented in Chapter 3.

Random mutations of single nucleotides can be induced by mutagenic chemicals or UV light. A serious obstacle of this method is that, in the following step, the affected locus has to be mapped on the genome, which can only be achieved by complementation analysis or whole-genome sequencing. Both are time- and cost-intensive procedures.

The mutated locus can be readily identified if genes are disrupted by insertion of a transposable element. Transposable elements are DNA sequences which can move to different positions in the genome, thereby enhancing genetic diversity. Transposons were identified as mobile genetic elements over sixty years ago and, subsequently, became powerful tools for molecular and genetic studies (McClintock, 1950). Movement of transposons from one location to another is facilitated by the enzyme transposase, which catalyzes the DNA breakage and rejoining reactions required for transposition. Transposase specifically binds to terminal repeat sequences flanking the

transposon cassette, thereby facilitating its excision and reintegration at a new location. In many cases, the enzyme is itself encoded on the transposon, between the terminal repeat structures.

Transposon mutagenesis has been successfully used to study gene function in the yeasts *S. cerevisiae*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica* (Ross-MacDonald, 1999; Kumar et al., 2004; Seringhaus et al., 2006; Evertts et al., 2007; Park et al., 2009; Richard et al., 2001; Mauersberger et al., 2001). Applying transposon mutagenesis in *P. pastoris* is primarily hampered by the fact that no native transposon had been identified in this organism (Küberl et al., 2011, *personal communication*). Our initial attempts to establish the Hermes transposon system of *Musca domestica*, which had been used to mutagenize *S. pombe* (Evertts et al., 2007; Park et al., 2009), and the *zeta*-transposon of *Y. lipolytica* (Mauersberger et al., 2001) in *P. pastoris* were not successful.

Secondly, the shuttle mutagenesis approach, a well-established protocol in *S. cerevisiae* (Seifert et al., 1986), where fragments of yeast DNA are mutagenized by transposons in *E. coli* and are then used to replace the native alleles in the yeast genome, is not applicable in *P. pastoris* due to the yeast's low frequency of homologous recombination. Even when using extensive homology regions of more than 1000 bp the rate of correct replacement events rarely exceeds 30 % (Näätsaari et al. 2012). In comparison, in the model yeast *S. cerevisiae* targeted knockouts can be achieved with short flanking homology regions of only 40 bp (Brachmann et al., 1998). We reasoned that we could employ the preference of *P. pastoris* for non-homologous recombination to develop a protocol for random insertion mutagenesis.

The rationale behind the design of our mutagenesis cassette was that it should fulfill the following demands. It should 1) allow selection for a positive integration event after transformation, 2) have no homologous regions to the genome sequence of *P. pastoris* to exclude biased integration events and 3) be as short as possible to enhance transformation rates. The latter two requirements can only be fulfilled by antibiotic resistance markers, as they are obtained from unrelated species and usually are less than 2000 bp in length.

In Chapters 1 and 2, we show that a Zeocin resistance cassette, amplified by PCR, integrates randomly into the *P. pastoris* genome. The same holds true for a *KanMX6* cassette, but for this marker we observed a much higher transformation background, rendering it unattractive for large-

scale mutagenesis approaches. The mutants produced by insertion mutagenesis were subsequently screened for increased or decreased secretion levels of recombinant reporter proteins.

Reporter proteins and screening

Horseradish peroxidase (HRP)

HRP is a prominent example for an industrially important enzyme. Not only is HRP extensively used as a reporter enzyme in diagnostic assays; its various other applications include organic synthesis, bioremediation, and biosensors (reviewed in Veitch, 2004, and Ryan et al., 2006). Originally isolated from the horseradish *Armoricia rusticana*, reactions catalyzed by HRP have been described as early as in the beginning of the 19th century (Planche 1810). However, enzyme preparations from plant comprise a mixture of different isoenzymes displaying different biochemical properties. As many as 42 different isoenzymes were separated by isoelectric focusing of commercial enzyme preparations (Hoyle 1977). Despite the diversity and seasonal variation of plant extracts, horseradish roots are still the main source for HRP.

The most abundant of the isoenzymes, HRP C, is also the best studied one. Structural and functional analysis of HRP C revealed that the 308 amino acid protein contains four disulfide bonds and one salt bridge. Additionally, HRP incorporates a heme group and two calcium ions, both essential for structural and functional integrity. Nine potential *N*-glycosylation sites (N-X-S/T) were detected in the sequence, of which eight were found to be occupied in plant isolates. In total, HRP comprises a carbohydrate content of approximately 20 % (Veitch, 2004).

Recombinant production of HRP in *E. coli* is problematic, as it requires the subsequent recovery of the enzyme from inclusion bodies. Morawski et al. (2000) expressed the isoenzyme C1A in *S. cerevisiae* and *P. pastoris*. More recently, Krainer et al. (2013) reported the recombinant expression and characterization of 19 HRP isoforms in *P. pastoris*. Notably, despite the significant industrial interest, HRP is still secreted at relatively low levels (Krainer et al., 2013).

In the work presented here, we report the random mutagenesis of HRP-secreting strains, with the aim of identifying gene products involved in the secretion of this enzyme. The method used for random mutagenesis is described in detail in Chapter 2. Alterations in the secretion of HRP were

detected in a colorimetric screening assay using the classical peroxidase substrate ABTS [2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid)]. In the HRP-catalyzed reaction, hydrogen peroxide oxidizes the colorless substrate via an intermediate radical cation step to the green-colored end product with an absorbance maximum at 420 nm (Childs and Bardsley, 1975). Genes identified in this approach were selectively targeted with knockout cassettes to confirm the results of the random strategy. More information on the strategy for targeted knockouts can be found in Chapter 3. The resulting knockout strains were further characterized as described in Chapter 1 and in the supplementary information.

Alternative pig liver esterase (APLE)

APLE, one of the isoenzymes of pig liver esterase (PLE), was discovered by our group in the search for an esterase that can hydrolyze methyl-(2R,4E)-5-chloro-2-isopropyl-4-pentenoate in a highly stereoselective manner (Hermann et al., 2008). The catalytic activity of PLE was first described in 1903 (Dakin, 1903) and, since then, the enzyme has played a valuable role in the synthesis of chiral compounds (De Maria et al., 2007). Originally, PLE was employed as a crude extract isolated from the liver of the pig *Sus scrofa*. One of the initial drawbacks in the use of this crude extract was that it is in fact a mixture of several esterases comprising a sequence identity of >95%, which display different kinetics and substrate specificities (Hummel et al., 2007). To overcome this problem, attempts were undertaken to characterize all of the existing isoenzymes and to obtain significant amounts by heterologous overexpression. One option is expression of PLE in *E. coli*. However, APLE contains disulfide bridges and can therefore not fold into its biologically active form in the reducing environment of the bacterial cytoplasm. Possible alternatives are expression in the periplasm of *E. coli* or in the more oxidizing cytoplasm of special *E. coli* Origami cells. Our group obtained high yields of APLE in *E. coli* Origami B cells through co-overexpression of a bacterial chaperone (own unpublished results).

Lange et al. (2001) and Musidlowska et al. (2001) described for the first time the expression of the γ-PLE isoenzyme in the eukaryotic host *P. pastoris*. Similarly, our group reported the successful expression of APLE in this yeast (Hermann et al., 2008). Unexpectedly, despite deleting the putative ER-retention signal –HAEL and directing the protein for secretion by fusing it to the α-factor secretion signal, little enzyme activity was detected in the culture supernatant. Indeed, APLE and γ-PLE were found to be localized in the yeast periplasm instead (Hermann et al., 2008).

PLE was shown to be active as a trimer, with a total molecular weight of approximately 180 kD (Lange et al., 2001). We speculated that the large multimer might be retained in the cell due to insufficient passage through the yeast cell wall.

Its low secretion levels made APLE an interesting target for our random mutagenesis experiments, with the ultimate goal to find factors involved in the secretion of recombinant proteins in *P. pastoris*. Strains secreting APLE were mutagenized employing the strategy described in Chapter 2. Initially, we planned to detect changes in APLE secretion with a colorimetric assay basically as described in Hermann et al. (2008). In this assay, filter paper is soaked in buffered screening solution containing esterase substrate and pH-indicator. APLE catalyzes hydrolysis of the ester to the corresponding alcohol and acid. The carboxylic acid liberated in this step causes the pH-indicator to change its color, indicating the amount of APLE secreted by the *P. pastoris* strains. Unfortunately, the intrinsically low pH of *P. pastoris* cultures complicated read-out of the screen. In a second approach, we quantified secreted APLE in culture supernatants via a dot-blot with primary antibody specific for the protein. Limitations of this method were the time-consuming incubation steps in addition to low reproducibility. Consequently, we established a colorimetric screening assay based on the hydrolysis of para-nitrophenyl acetate (p-NPA), a standard esterase substrate. In total, 2100 mutants secreting APLE were screened with this assay. Results of this screening and the subsequently generated targeted knockouts of selected genes are presented in the supplementary information and discussed in the conclusion.

Alpha-mating factor (MFα1)

Recombinant proteins are targeted to the secretory pathway by addition of an N-terminal secretion signal sequence. In *P. pastoris* and other yeasts, the most commonly used signal sequence is the mating pheromone α-factor signal MFα1 from *S. cerevisiae* (Idiris et al., 2010). *S. cerevisiae* α-factor is only 13 amino acids in length, and it is released by cells of the α-mating type to induce mating with cells of the opposite *a*-mating type, subsequently leading to the formation of *a*α diploid cells. At genome level, the α-factor gene can contain three to five repeats of the mature pheromone peptide (Brake et al., 1983). First, the prepro-precursor protein comprising 165 amino acid residues in length is expressed in the cytosol. After transport into the ER, a 146 residue pro-α-factor is produced by signal peptidase cleavage. Localized to the late Golgi department, the protease Kex2 cleaves between the pro-domain and the first α-factor repeat, and at three sites

between the α-factor repeats. Further steps required for the production of mature mating pheromone are exoproteolysis by Ste13p, dipeptidyl aminopeptidase, to remove N-terminal GluAla and AspAla dipeptides from each repeat and by Kex1p, carboxypeptidase, to remove C-terminal Lys and Arg residues from the internal repeats (Bevan et al., 1998).

In an *in vitro* experiment aimed to reconstruct translocation into the ER, prepro-α-factor was demonstrated to translocate post-translationally into yeast microsomes (Hansen et al., 1986). Like in co-translational insertion, proteins that translocate post-translationally traverse the ER membrane through the Sec61 channel. However, the post-translational delivery mechanism additionally needs Sec62, Sec63, Sec71 and Sec72 for insertion (Johnson, 2013). The driving force for translocation through the channel is provided by the ER-resident chaperone Kar2p (BiP), which acts as a "molecular ratchet", effectively preventing backward movement of the translocating polypeptide chain (Matlack et al., 1999).

Brake et al. (1984) showed for the first time that the signal sequence of α-mating factor can be used to drive secretion of recombinant proteins. They fused the α-factor prepro-signal sequence to mature human epidermal growth factor (hEGF) and observed that *S. cerevisiae* cells transformed with the construct efficiently synthesized, processed, and secreted the mature foreign protein into the medium. Later, Rothblatt et al. (1987) tested if fusion to the α-mating factor signal sequence leads to post-translational translocation of recombinant proteins. Surprisingly, they found that the secretion-targeted model protein, chimpanzee α-globulin, crossed microsomal membranes by co-translational insertion. This finding is in contrast to the translocation mechanism found for the native α-mating factor peptide. In our experiments, the model proteins HRP and APLE were targeted to the secretory pathway by fusing them to the signal sequence of α-mating factor. Translocation to the ER could potentially be a limiting factor for high-yield production. However, whether these proteins translocate in a co- or post-translational fashion when fused to the signal sequence of α-mating factor is unknown.

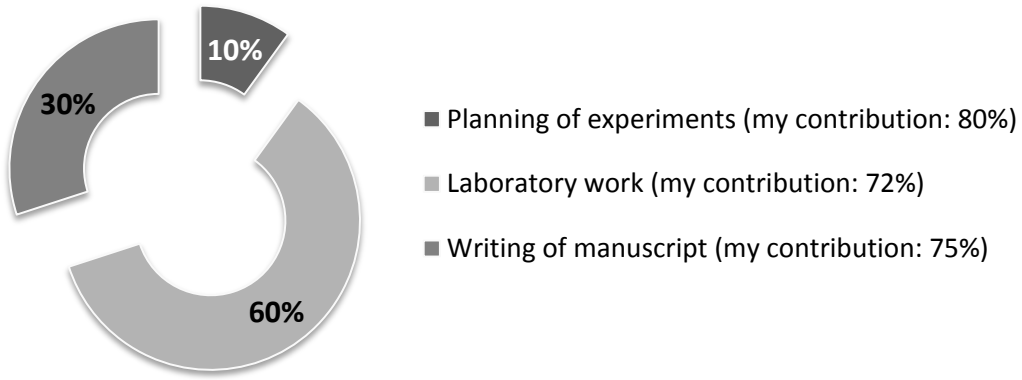
Apart from using the prepro-α-factor signal sequence as a leader to drive secretion of HRP and APLE, we tried to establish the mating pheromone itself as a third model protein. We reasoned that the small size of 13 amino acids should allow α-factor to be secreted efficiently, and that random mutagenesis of secreting strains could lead to the identification of genes that are essential for secretion. Methods to quantify secreted α-factor were published already more than 30 years

ago. Chan and Otte (1982) described that *S. cerevisiae* Mat*a sst1 sst2* cells are arrested in the G1 phase and subsequently growth inhibited in the presence of α-mating factor. Manney (1983) developed a screening assay to quantify α-factor pheromone based on these findings. He related the size of a halo of growth inhibition produced in a lawn of sensitive *sst1 (bar1)* mutant cells to the quantity of α-factor added to wells in an agar plate. We considered employing a variation of this assay to quantify recombinant *S. cerevisiae* α-mating factor secreted from *P. pastoris* cells. Therefore, we constructed *P. pastoris* strains expressing the *S. cerevisiae* α-mating factor gene. Moreover, we aimed to modify the halo assay developed by Manney (1983) to allow higher screening throughput. A typical result for our newly developed halo assay is shown in the supplementary information. Unfortunately, we found it difficult to quantify the size of the halos formed around pinning spots. Another drawback of the method was the low throughput of mutant colonies, since the formed halos were too big to allow screening in 96-well format. Consequently, we abandoned this screening approach.

CHAPTER 1

Random mutagenesis to improve secretion in *Pichia pastoris*

% Contribution



ENHANCED PROTEIN SECRETION IN *PICHIA PASTORIS* UPON KNOCKOUT OF *KEP1* ORPHAN GENE

Christine M. Winkler[1], Mudassar Ahmad[1], Peter Neugebauer[1], Britta Obrist[2], Barbara Darnhofer[2], Ruth Birner-Grünberger[2], Helmut Schwab[1,3] and Harald Pichler[1,3*]

[1]*Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria*
[2]*Institute of Pathology & Core Facility Mass Spectrometry, Medical University of Graz, Stiftingtalstrasse 24, 8010 Graz, Austria*
[3]*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

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 vivo*. Applying this method and performing medium throughput screening for enhanced 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 an endogenous protein in the culture supernatant, which was identified as a homolog of the *S. cerevisiae* flocculins (Flo proteins) by mass spectrometry. This phenotype was reversed by restoring *KEP1* expression in the respective knockout strain. Most interestingly, deletion of *KEP1* also benefits secretion of model proteins structurally unrelated to HRP, including alternative pig liver esterase (APLE) and human growth hormone (hGH). Thus, the knockout strain *kep1*Δ has the potential to be 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. For many proteins of plant and mammalian origin these modifications are prerequisites 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 is still 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). As the factors involved underlie complex interactions, engineering of these bottlenecks has proven to be extremely challenging. A common strategy to improve recombinant protein secretion has been to manipulate the organism based on knowledge obtained from the model yeast *S. cerevisiae*. Damasceno et al. (2011) and Idiris et al. (2010) reviewed the co-overexpression of ER-resident chaperones, e.g. BiP/Kar2 and PDI, and other proteins like the ERAD inducing transcription factor Hac1 in *P. pastoris* and related yeasts. In most of the cases, the co-overexpression of secretion helpers and 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 obvious 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; Stadlmayer et al. 2009; Baumann et al. 2011; Pfeffer et al. 2012). The massive 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 so far 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 horseradish peroxidase (HRP) as a reporter protein. We subsequently screened for gene deletions affecting the amount of HRP secreted to the culture medium positively and negatively. 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. We found that the *kep1*Δ knockout strain releases an endogenous protein to the culture supernatant to a higher extent than the wild type. Mass spectrometry identified this protein as Flo9, a lectin-like protein with similarity to the flocculin (Flo) protein family of *S. cerevisiae*. Restoring *KEP1* expression in the respective knockout strain reversed the phenotype of enhanced Flo9 release. To our surprise, the phenotype was enforced by disruption of the two genes encoding *P. pastoris* lysyl oxidase (PPLO), a protein also found at increased quantities in the *kep1*Δ culture supernatant.

Aside from enhancing HRP secretion, deletion of *KEP1* is equally beneficial for the secretion of two other structurally and functionally unrelated model 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.

Materials and methods

All *P. pastoris* strains constructed during this study are described in more detail in supplementary table 1. An overview of all employed primers is given in supplementary table 2. Unless otherwise mentioned, we used standard cloning techniques to construct the plasmids listed in supplementary 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 mutant 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 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. 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. 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 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. We used 0.6 µg of DNA per 80 µL of cells for transformation. 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 mutant 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).

Restored wild type secretion behavior through expression of flag-tagged proteins in knockout strains

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-Cl 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 anti-mouse antibody produced in goat (both Sigma-Aldrich).

Mass spectrometry

Cells were grown for 72 h on BYPD. Proteins in culture supernatants were deglycosylated with EndoH (New England Biolabs) prior to TCA precipitation and SDS-PAGE, performed as described above. The unique band in the *kep1Δ* strain and the corresponding position in the wild type control lane were excised from the Coomassie blue-stained gel and were reduced, alkylated and digested with Promega modified trypsin according to the method of Shevchenko et al. (1996). Digests were separated by nano-HPLC (Dionex Ultimate 3000) equipped with a µ-precolumn (C18, 5 µm, 100 Å, 5 x 0.3 mm) and an Acclaim PepMap RSLC nanocolumn (C18, 2 µm, 100 Å, 150 x 0.075 mm) (all Thermo Fisher Scientific, Vienna, Austria). Fifteen of 25 µL of digested protein samples were injected and concentrated on the enrichment column for 2 min at a flow rate of 20 µL/min with 0.5 % trifluoroacetic acid as isocratic solvent. Separation was carried out on the nanocolumn at a flow rate of 300 nL/min using the following gradient, where solvent A is 0.3 % formic acid in water and solvent B is a mixture of 80 % acetonitrile in water containing 0.3 % formic acid: 0-2 min: 4 % B; 2-70 min: 4-28 % B; 70-94 min: 28-50 % B, 94-96 min: 50-95 % B; 96-116 min: 95 % B; 116-116.1 min: 95-4 % B; 116.1-140 min: re-equilibration at 4 % B. The sample was ionized in the nanospray source equipped with stainless steel emitters (ES528, Thermo Fisher Scientific, Vienna, Austria). It was analysed in a Thermo LTQ-FT mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) operated in positive ion mode, applying alternating full scan MS (m/z 300 to 2000) in the ion cyclotron and MS/MS by collision induced dissociation of the 5 most intense peaks in the ion trap with dynamic exclusion enabled.

The LC-MS/MS data were analyzed by searching the NCBI *Pichia* (*Komagataella) pastoris* public database (15621 sequences; 7561633 residues) downloaded on Sep 4th, 2013, with Proteome Discoverer 1.4 (Thermo Fisher Scientific, Waltham, MA, USA) and Mascot 2.4 (MatrixScience, London, UK). Carbamidomethylation on Cys was entered as fixed modification. Oxidation on methionine was entered as variable modification. A maximum false discovery rate of 5 % using decoy database search, a Mascot ion score cutoff of 20 and a minimum of 2 identified unique peptides were chosen as identification criteria.

Results

Mutant library construction

Many researchers working with *P. pastoris* have struggled with this yeast's preference for non-homologous end-joining 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 DNA fragments lacking sequence homology into the genome sequence of *P. pastoris*.

The Zeocin resistance cassette of plasmid pGAPZ is 1172 bp in length (Fig. 1). It 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. 1988; Drocourt et al. 1990; Calmels et al. 1991) and the *S. cerevisiae CYC1* transcription termination region (GenBank accession number M34014). No significant homology of the cassette to the *P. pastoris* genome sequence was detected by 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) for secretory expression of HRP from P_{GAP}. The mutagenesis cassette was amplified from the template pGAPZ by PCR and the purified product was used to transform the HRP-secreting strain. We observed a transformation rate of ~200 CFU per microgram of transformed PCR product. 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. HRP enzyme activity in the culture supernatants was measured by a kinetic assay with chromogenic 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) diammonium salt (ABTS) as substrate. Color development was observed by eye over a time period of 2-5 min. Hits of the primary screen with higher or lower color intensity than the reference, the mutagenesis starting strain, were subjected 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.

Genome walking and bioinformatics search

Genomic regions flanking the insertion loci of the resistance cassette were amplified and sequenced following the template-blocking PCR protocol of Bae and Sohn (2010). The resulting nucleotide sequences were blasted 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 are 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. They were disrupted in more than one mutant subjected to genome walking. Mapping of the insertion locus showed that the 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.

Based on these 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 has no significant homology to any described protein of *S. cerevisiae*. We did not further investigate gene deletions that reduced HRP activity.

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. Following successful gene disruption, the marker and other elements of the vector were excised by recombination. The vector elements were looped out by recombinase, 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 transformants. Plasmids constructed for the secreted expression of HRP are depicted in supplementary figures 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 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. Instead, HRP secretion in the *kep1*Δ strain seemed to skyrocket as soon as the cells reached the stationary phase after ~40 h of cultivation time. This sudden increase 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, while 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. 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.

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 corresponding to the strength of the promoter.

Speculating that the higher level of HRP in the culture supernatant of knockout strains stemmed from a general effect on the protein secretion machinery, we expected the same favorable effect on the secretion of other recombinant proteins. 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. Both, activity assays with the esterase substrate p-NPA and Western blot analysis with primary antibody recognizing APLE 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. Taken together, the results suggest that *kep1*Δ has a universal effect on protein secretion, while the beneficial effect of *rim101*Δ and *sgt2*Δ depends on the secreted protein.

Restored wild type secretion behavior through expression of flag-tagged proteins in knockout strains

In our initial screening we found the ORFs 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, endogenous protein into the culture supernatant. This release resulted in a prominent band of ~130 kDa on an SDS-PAGE loaded with TCA-precipitated culture supernatants of the *kep1*Δ strain (Fig. 5b). 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.

Identification of the endogenous protein released by the *kep1*Δ strain

We used mass spectrometry to shed light on the identity of the endogenous protein released by the *kep1*Δ strain. Culture supernatants were treated with EndoH to remove high mannose *N*-glycans from glycoproteins. As a result, the prominent protein band with an apparent molecular weight of ~130 kDa shifted to ~110 kDa on a Coomassie-stained SDS-PAGE. We excised this band and the corresponding fragment in the wild type control lane, in order to compare the protein content of the two gel slices. LC/MS analysis identified a list of proteins that were present at different levels in the *kep1*Δ and wild type samples. The most prominent difference between the samples was

detected for CCA37505.1, encoding Flo9, a homolog of the *S. cerevisiae* Flo proteins, and for the two proteins CCA38674.1 and CCA40518.1. The latter two proteins share 78 % identity on the amino acid level, and 77 % on the DNA level. Both are annotated as copper-containing primary-amine oxidase. This enzyme was described in literature to possess lysyl oxidase activity (Tur and Lerch 1988; Kuchar and Dooley 2001; Duff et al. 2003), and is therefore referred to as *P. pastoris* lysyl oxidase (PPLO). Kuchar and Dooley (2001) reported glycosylated PPLO to migrate at 120 kDa, and at 107 kDa when deglycosylated - a migration pattern similar to the one we observed. Given that the predicted molecular weight of *P. pastoris* Flo9 is 51 kDa, we initially reasoned that the interesting band must stem from increased PPLO release. However, disruption of the two genes encoding PPLO, referred to as "lysoxA" (CCA38674.1) and "lysoxB" (CCA40518.1) throughout this paper, did not abolish the additional band at 130 kDa in the *kep1*Δ strain, but rather enforced it (Fig. 6a, 6b). Repeated mass spectrometric analysis of the interesting band, this time excised from the supernatant of *kep1*Δ *lysoxA*Δ *lysoxB*Δ, ran on a SDS-PAGE, confirmed the depletion of PPLO proteins in this triple knockout. On the other hand, Flo9 was unequivocally identified as the most prominent protein in the gel sample.

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 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 heme (Table 2). Hem3 catalyzes the third step of heme biosynthesis (Keng et al. 1992), while Fra1 is involved in the regulation of iron uptake in *S. cerevisiae* (Kumanovics 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).

The 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 transcription factor Rim101 is the ultimate target of a signal transduction pathway sensing extracellular pH. 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. 1999; 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 pathway in *P. pastoris* strongly indicated that this regulon is actually connected to HRP secretion.

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) in *P. pastoris*. 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 a different biological function in *P. pastoris*, which has to be identified yet.

As a member of the GET-complex, Sgt2 is involved in the insertion of tail-anchored (TA) proteins into the ER. TA proteins are 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 model 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.

For the hypothetical protein CCA40244.1 no close homolog in *S. cerevisiae* or any other related yeast was found. However, it shares a short stretch of homologous amino acids with other, so far uncharacterized, proteins of the yeasts *Ogataea parapolymorpha* and *Dekkera bruxellensis*. The 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), since 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 is the release of the endogenous protein Flo9 to the culture supernatant, a phenotype that can be reversed by restoring *KEP1* expression (Fig. 5b). *P. pastoris* Flo9 shows homology to several members of the *S. cerevisiae* Flo proteins, which are lectin-like adhesion proteins with a role in pseudohyphae development, invasive growth and flocculation (for a review, see Verstrepen and Klis 2006). The *P. pastoris* protein contains the mannose-binding PA14 domain, present in the *S. cerevisiae* proteins Flo1, Flo5, Flo9 and Flo10 (Rigden et al. 2004; Goossens and Willaert 2013), and the typical threonine- and serine-rich repeat sequences (Dranginis et al. 2007). However, the calculated molecular weight of the *S. cerevisiae* proteins is usually significantly above 100 kDa, while it is only 51 kDa for *P. pastoris* Flo9. We were surprised to find that the band for Flo9 runs at an apparent molecular weight of 130 kDa in SDS-PAGE (Fig. 5 and 6). Removal of selected carbohydrate chains with EndoH shifted the band to 110 kDa, though this endoglycosidase removes only high mannose and some hybrid types of *N*-linked carbohydrates, and the removal of *N*-glycosylations could therefore be incomplete. Moreover, Flo proteins in *S. cerevisiae* were reported to be massively O-glycosylated (reviewed in Dranginis et al. 2007). These additional carbohydrate modifications, or unknown covalent linkages to other proteins, could be responsible for the unexpectedly high apparent molecular weight of the Flo9 protein. Another surprising finding was that the Flo9 band observed in the *kep1Δ* strain became even more pronounced upon additional disruption of the two genes encoding PPLO (Fig. 6).

PPLO had first been identified by Green et al. (1983) in their search for yeasts that can live with small organic amines as their sole source of nitrogen. The enzyme was initially termed benzylamine oxidase, until Tur and Lerch (1988) reclassified it as lysyl oxidase, due to its broad substrate specificity that indicated similarities to mammalian lysyl oxidases. In vertebrates, lysyl oxidases catalyze the oxidation of lysine residues in collagen and elastin during biogenesis of connective tissue (Smith-Mungo and Kagan 1998). Interestingly, PPLO was shown to catalyze the crosslinking of tropoelastin at a rate comparable to the true mammalian enzyme (Duff et al. 2003). Crystallization studies indicated that the enzyme is present as homodimer in solution (Duff et al. 2003; Duff et al. 2006). It is worth mentioning, however, that all functional and structural studies on PPLO were published before the complete genome sequence was first made available in 2009 (De Schutter et al. 2009). Only then it was revealed that the enzyme is encoded by two highly

homologous genes on chromosome 2 and 4, corresponding to the proteins CCA38674.1 and CCA40518.1, respectively.

The present findings suggest a functional link between Kep1, Flo9 and PPLO, but the nature of their molecular interaction remains elusive. Both Flo9 and PPLO are secreted proteins and most likely localized at the surface of the yeast cell, while no potential secretion signal sequence was detected for Kep1, as determined by the SignalP 4.1 prediction tool (Emanuelsson et al. 2007; Petersen et al. 2011). This observation raises the question where in the cell the proposed interaction takes place, and since little is known about the function of Kep1, if it has a structural or regulatory base. Moreover, the enhanced secretion of HRP after the cells entered the stationary phase (Fig. 2b), indicated an influence of the cell's physiological state on protein regulation. Further work will focus on elucidating the function of Kep1 in the cell, and, especially, its effect on Flo9 and PPLO localization.

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. The most striking result was the identification of 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

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

Dranginis AM, Rauceo JM, Coronado JE, Lipke PN (2007) A biochemical guide to yeast adhesins: glycoproteins for social and antisocial occasions. Microbiol Mol Biol Rev 71:282–94. doi: 10.1128/MMBR.00037-06

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

Duff AP, Cohen AE, Ellis PJ, Kuchar JA, Langley DB, Shepard EM, Dooley DM, Freeman HC, Guss JM (2003) The crystal structure of *Pichia pastoris* lysyl oxidase. Biochemistry 42:15148–57. doi: 10.1021/bi035338v

Duff AP, Cohen AE, Ellis PJ, Hilmer K, Langley DB, Dooley DM, Freeman HC, Guss JM (2006) The 1.23 Angstrom structure of *Pichia pastoris* lysyl oxidase reveals a lysine-lysine cross-link. Acta Crystallogr D Biol Crystallogr 62:1073–84. doi: 10.1107/S0907444906026333

Ecamilla-Trevino 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

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

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*. Molecular and General Genetics 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

Goossens KVY, De Greve H, Willaert RG (2013) Cloning, expression, and purification of the N-terminal domain of the Flo1 flocculation protein from *Saccharomyces cerevisiae* in *Pichia pastoris*. Protein Expr Purif 88:114–9. doi: 10.1016/j.pep.2012.12.001

Green J, Haywood GW, Large PJ (1983) Serological differences between the multiple amine oxidases of yeasts and comparison of the specificities of the purified enzymes from *Candida utilis* and *Pichia pastoris*. Biochem J 211:481–93.

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, Riezmann 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 Des Sel 14:705–710. 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

Kuchar JA, Dooley DM (2001) Cloning, sequence analysis, and characterization of the "Lysyl Oxidase" from *Pichia pastoris*. J Inorg Biochem 83:193–204.

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.

Petersen TN, Brunak S, von Heijne G, Nielsen H (2001) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785-786. doi:10.1038/nmeth.1701

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

Rigden DJ, Mello LV, Galperin MY (2004) The PA14 domain, a conserved all-beta domain in bacterial toxins, enzymes, adhesins and signaling molecules. Trends Biochem Sci 29: 335–339.

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

Shevchenko A, Jensen ON, Podtelejnikov AV, Sagliocco F, Wilm M, Vorm O, Mortensen P, Shevchenko A, Boucherie H, Mann M (1996) Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. Proc Natl Acad Sci U S A 93:14440–14445.

Smith-Mungo LI, Kagan HM (1998) Lysyl oxidase: properties, regulation and multiple functions in biology. Matrix Biol 16:387–98.

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

Tur SS, Lerch K (1988) Unprecedented lysyloxidase activity of *Pichia pastoris* benzylamine oxidase. FEBS Lett 238:74–76.

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

Verstrepen KJ, Klis FM (2006) Flocculation, adhesion and biofilm formation in yeasts. Mol Microbiol 60:5–15. doi: 10.1111/j.1365-2958.2006.05072.x

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

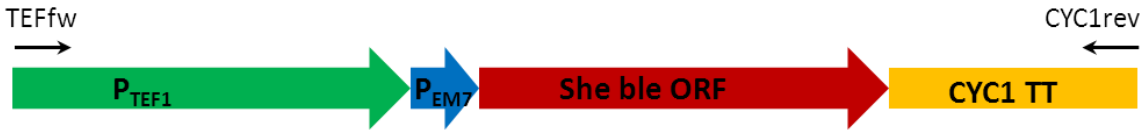
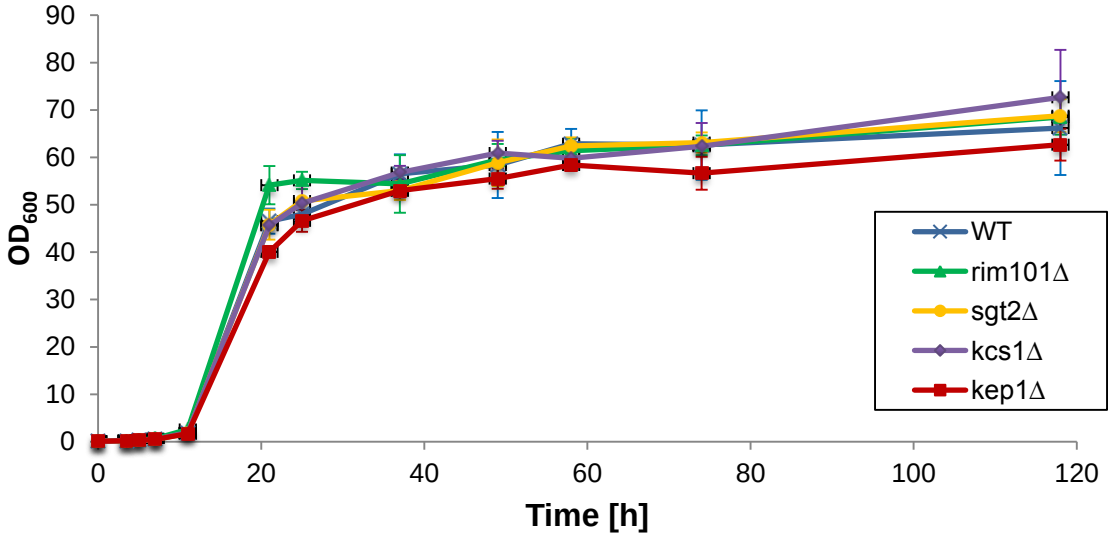


Fig. 1 Schematic drawing of Zeocin resistance mutagenesis cassette. The 1172 bp fragment composed of eukaryotic promoter P_{TEF1}, prokaryotic promoter P_{EM7}, *She ble* open reading frame and *CYC1* transcription terminator was amplified by PCR with the indicated primers.

a



b

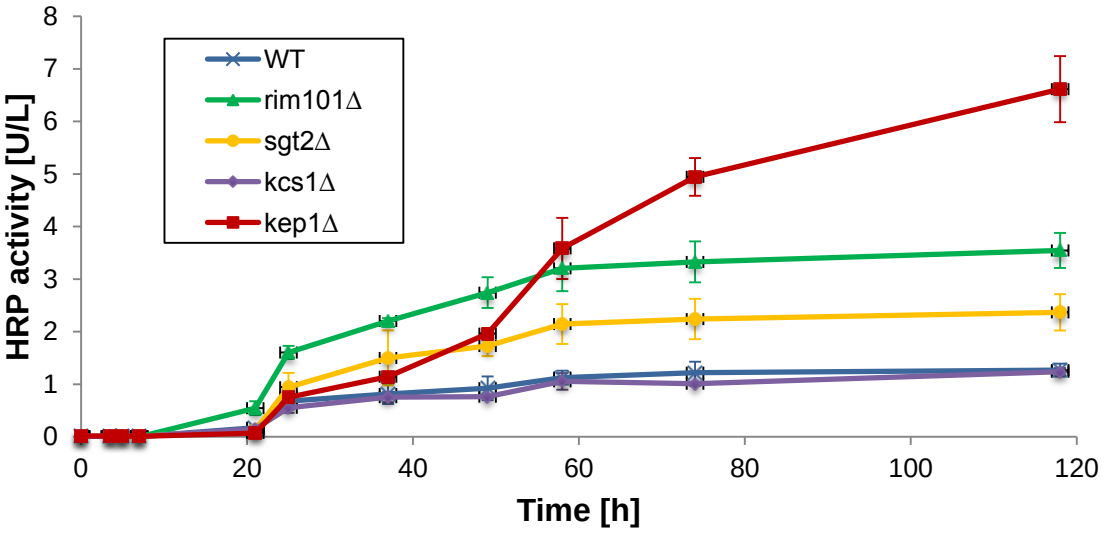
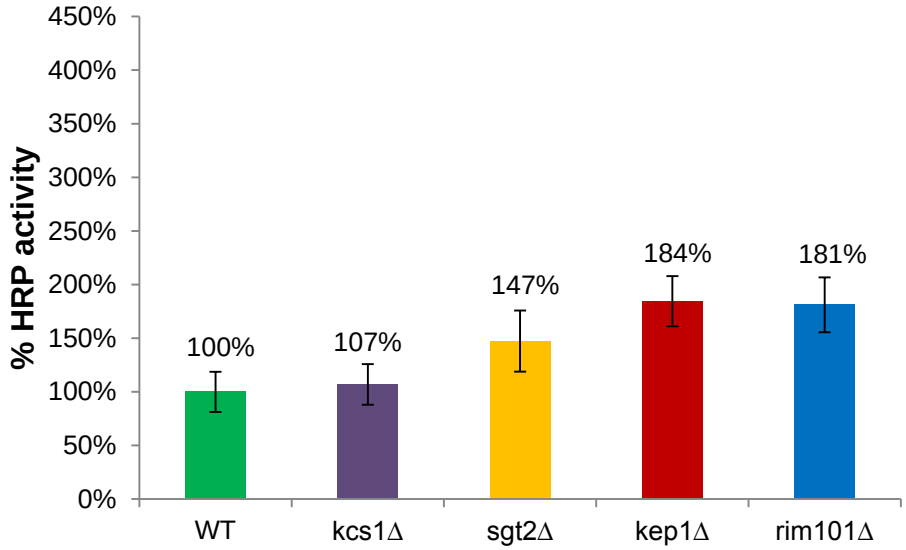
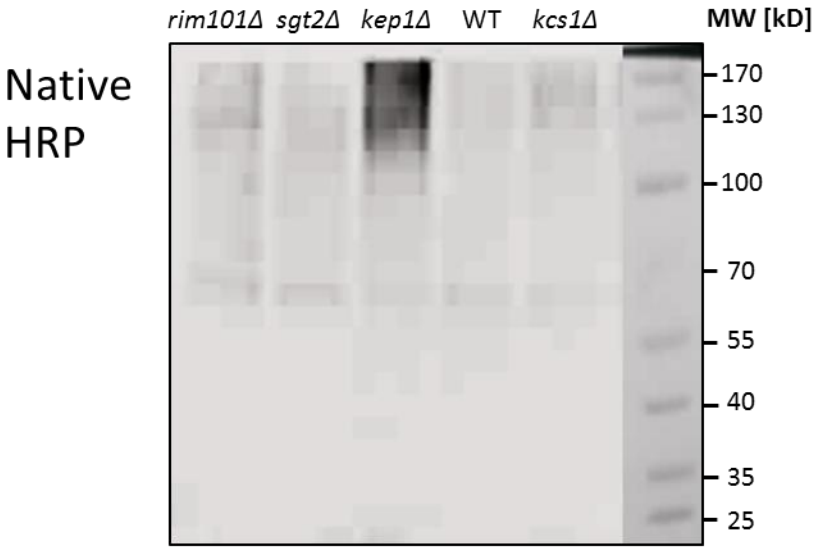


Fig. 2 HRP activity in supernatants of *kep1Δ* does not correlate with growth. Growth curve analysis of control and knockout strains secreting HRP from *GAP* promoter (a). Cells were grown in BYPD (2% glucose) in baffled shake flasks at 28°C. Experiments were performed in biological triplicate. HRP activity was detected in culture supernatants at the same time points (b).

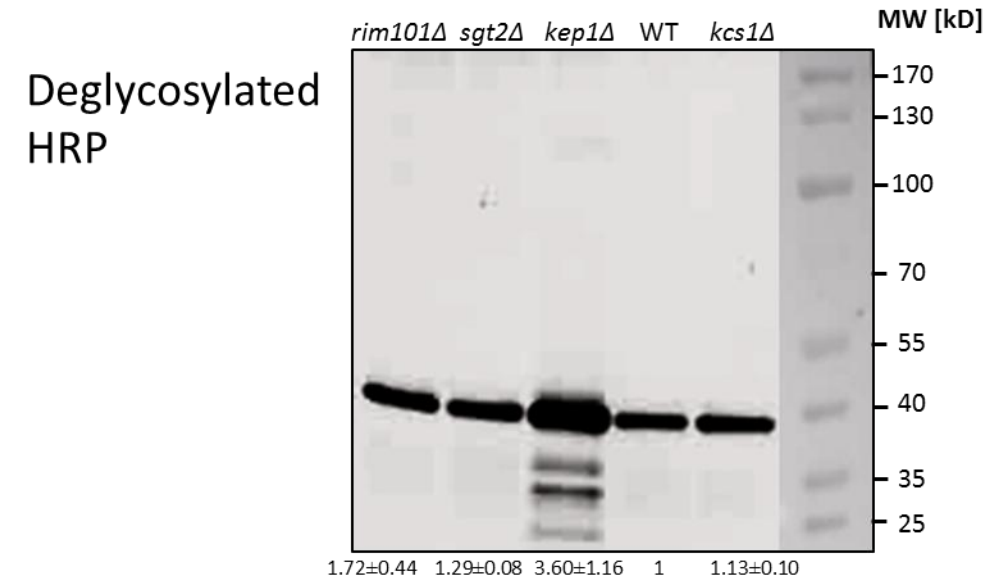
a


% HRP activity
450%
400%
350%
300%
250%
200%
150%
100%
50%
0%
100%
107%
147%
184%
181%
WT
kcs1Δ
sgt2Δ
kep1Δ
rim101Δ

b


Native HRP
rim101Δ
sgt2Δ
kep1Δ
WT
kcs1Δ
MW [kD]
170
130
100
70
55
40
35
25

c



d

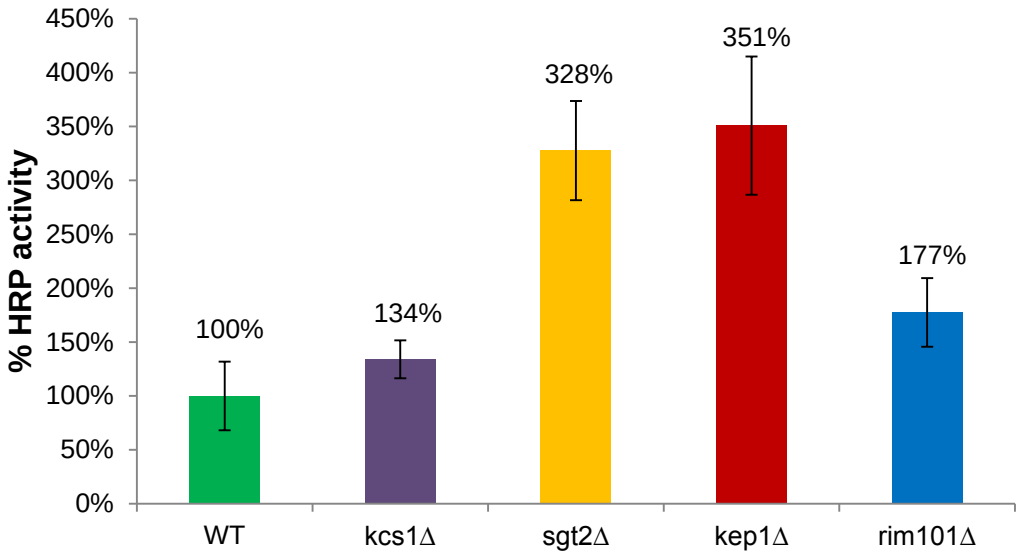
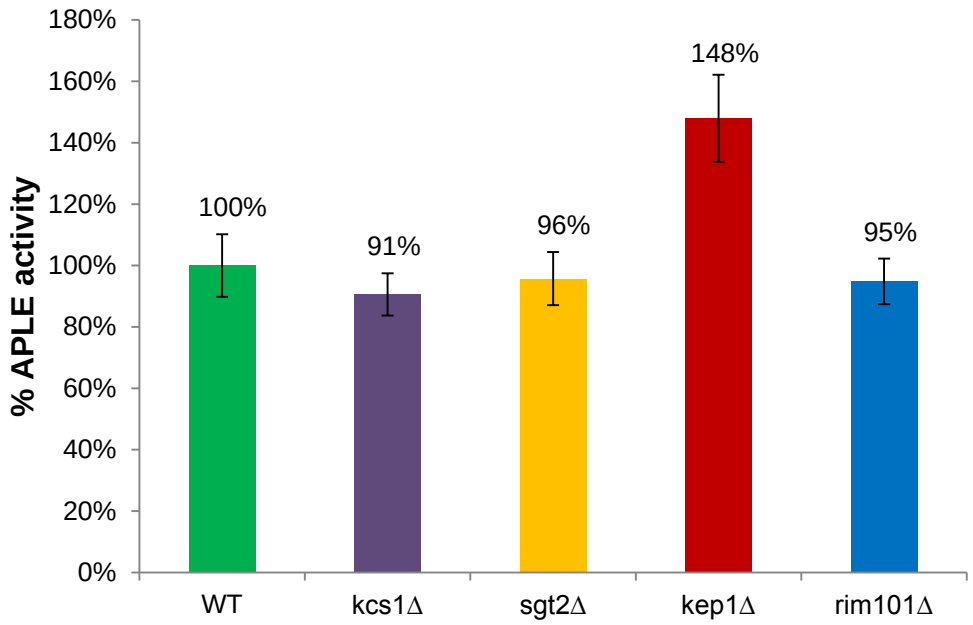


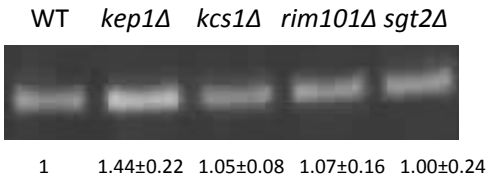
Fig. 3 Quantification of secreted HRP in deep-well plate culture supernatants. Relative HRP activity upon expression from *GAP* promoter (a). WT control and mutant strains expressing HRP from P_{GAP} were grown on BYPD (2% glucose) in deep-well plates for 35 h before analysis. Peroxidase activity in 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 HRP 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.

a



b



c

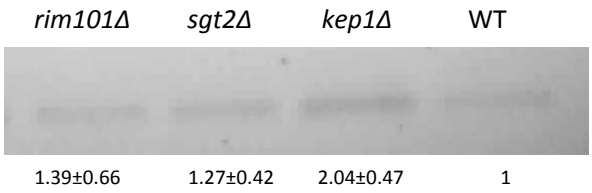
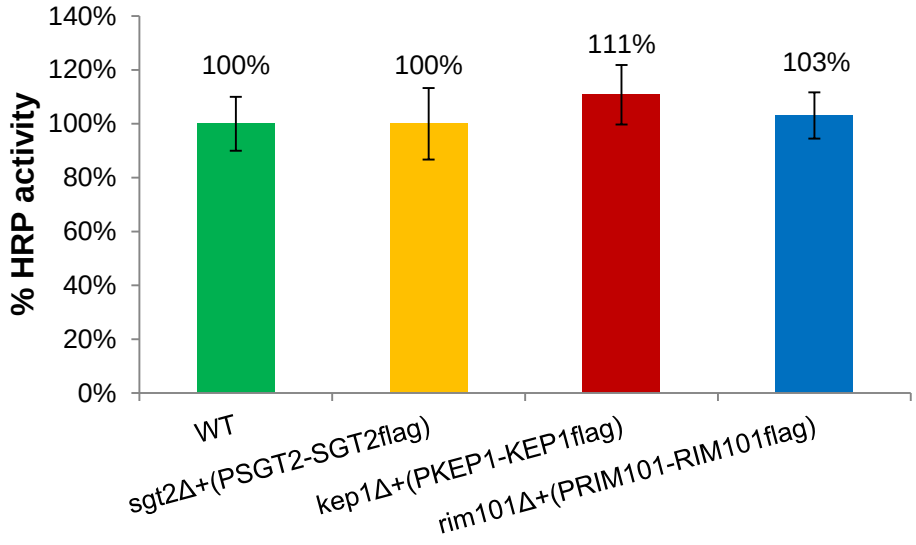


Fig. 4 Quantification of APLE and hGH secretion in deep-well plate cultivations. Strains expressing APLE from P_{GAP} were grown on BYPD (2% glucose) in deep-well plates for 35 h before analysis. Relative activity of APLE in culture supernatants was measured with pNPA assay (a). Results represent the 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



b

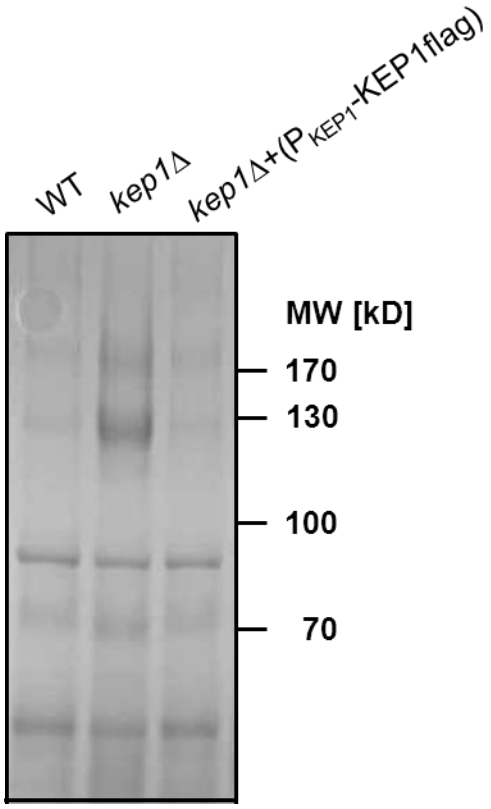


Fig. 5 Expression of targeted proteins in knockout strains restores wild-type HRP secretion levels. Relative HRP activity upon expression from *GAP* promoter (a). The mutant strains were transformed with expression constructs coding for flag-tagged versions of the respective target gene under control of its native promoter to reconstitute wild type-like expression. All strains express 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 approximately 130 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.

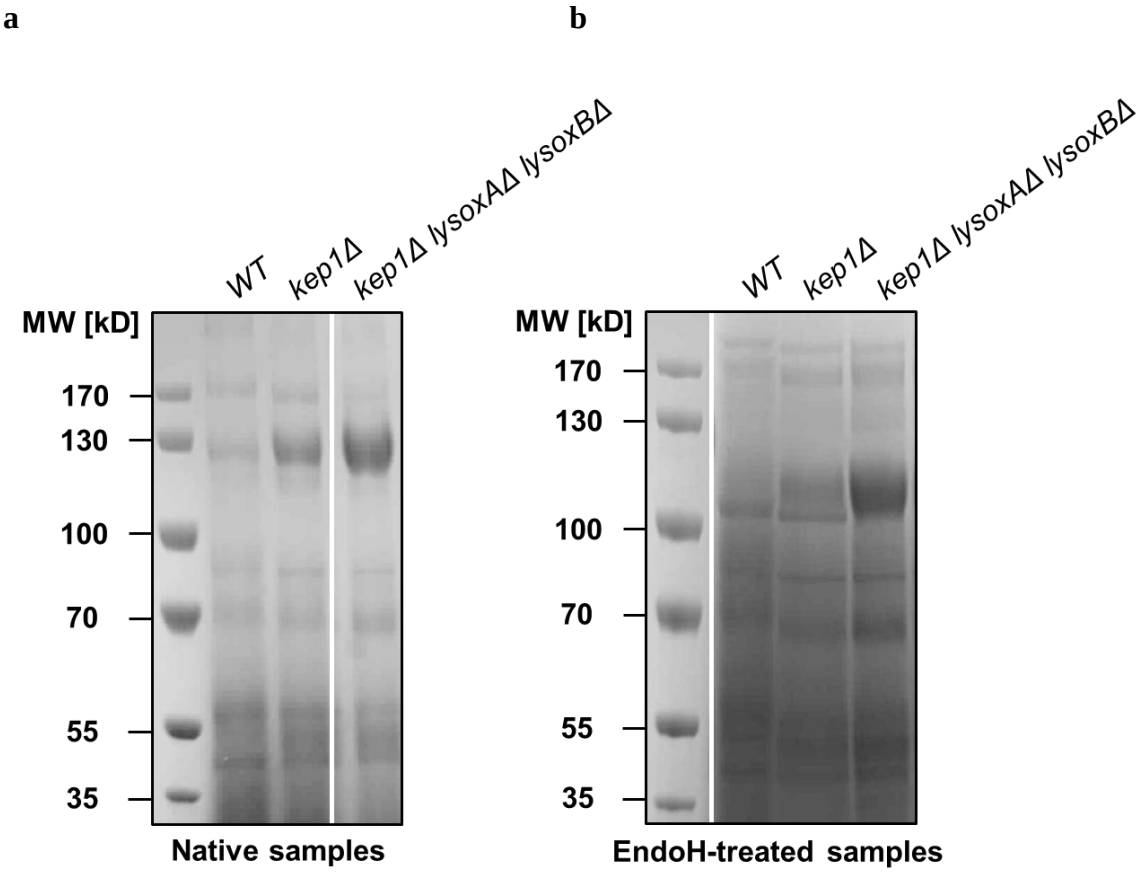


Fig. 6 Disruption of the two lysyl oxidase genes in *P. pastoris* does not abolish the 130 kDa band found in the culture supernatant of *kep1Δ* knockout strain, but enforces it instead. The interesting band runs at ~130 kDa in native supernatant samples (a), and at ~110 kDa in samples treated with EndoH to remove high mannose *N*-glycans from glycoproteins (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. Ladder: PageRuler™ Prestained Protein Ladder (Thermo Scientific).

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
WT+GAPEGFP	CBS7435 *Δhis4 Δarg4::pPT4-GAP-[EGFP]-Arg4-Zeocin*	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
kep1Δ+GAPEGFP	CBS7435 *Δhis4 Δkep1 Δarg4::pPT4-GAP-[EGFP]-Arg4-Zeocin*	This study
KEP1flag	kep1Δ+GAPαHRP *pPpT4_KEP1_C.FLAG*	This study
kep1ΔLysOxAΔ	CBS7435 *Δhis4 Δkep1 Δaoc1*	This study
kep1ΔLysOxAΔLysOxBΔ	CBS7435 *Δhis4 Δkep1 Δaoc1 Δaoc2*	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
kcs1Δ+GAPEGFP	CBS7435 *Δhis4 Δkcs1 Δarg4::pPT4-GAP-[EGFP]-Arg4-Zeocin*	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
rim101Δ+GAPEGFP	CBS7435 *Δhis4 Δrim101 Δarg4::pPT4-GAP-[EGFP]-Arg4-Zeocin*	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
sgt2Δ+GAPEGFP	CBS7435 *Δhis4 Δsgt2 Δarg4::pPT4-GAP-[EGFP]-Arg4-Zeocin*	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
pPpT4-GAP-EGFP-Arg4-Zeocin	pPpGAP-EGFP-ARG4-Zeocin	Schroer et al. (2010)
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
pPpKC1_LysOxA	5'LysOxA-FRT-Zeocin-Flippase-FRT-3'LysOxA	This study
pPpKC1_LysOxB	5'LysOxB-FRT-Zeocin-Flippase-FRT-3'LysOxB	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

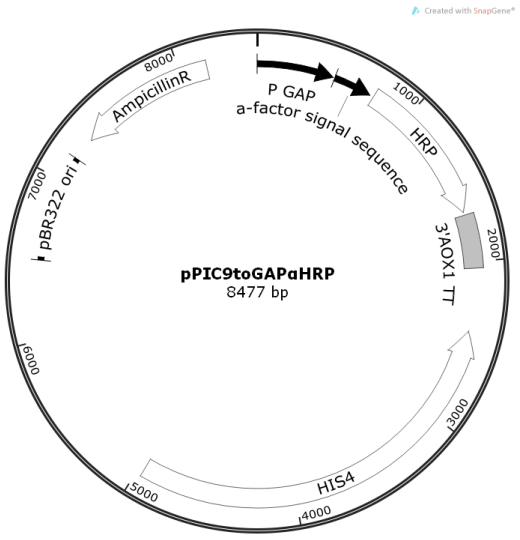
Suppl. table 3 All primers used during this study (During the course of the project, the names we used to refer to certain gene loci, changed: KEP1=H8=FLO11=MOEP; H7=KCS1)

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
3UTRLysylF	CGGCCGATCAGGCCAACCACCATTATAACCGCTCGCTTAG

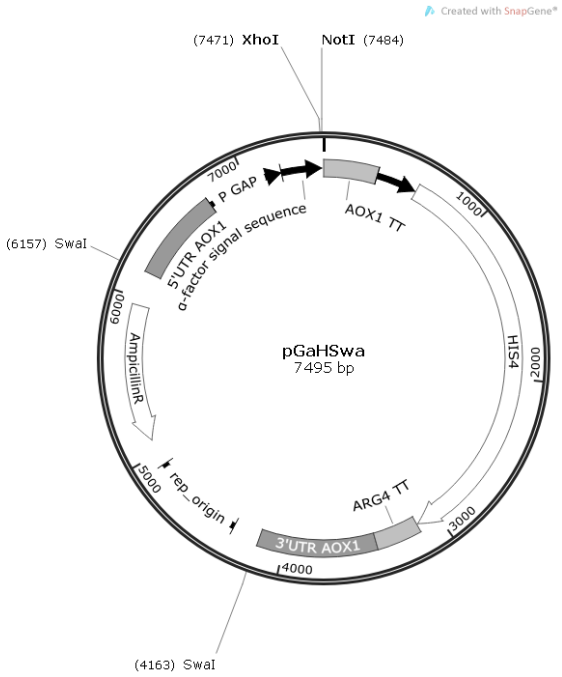
3UTRLysylR	GCGTCCATAATTTAAATGGTGCAGACCAACATTGACC
5UTRLysylF	GTCTGCACCATTTAAATTATGGACGCTGATTTCTACTCCAAC
5UTRLysylR	CGGCCCTAGTGGCCTACTCCTCGTCAGCATCGAATG
InsideLysOxAF	TGAGGGAGAAGAAGGCTACTTTC
InsideLysOxAR	CTGGCATGATGTTGTAAGCTCTTG
SeqLysOxfw	TCGGCAATCTTTTGAGTTTCACC
SeqLysOxrev	AAGCGTTGGTTGTCTTTTTGTTG
seqLysylAF	GCTGTTATCAATTCAGGTGTCGTTG
seqLysylAR	CTAATCCACCATCTACGTACTCTGG
3UTRLysylBF	CGGCCGATCAGGCCAGCTGCTCTTCTGAGAAGAGAAGCTAC
3UTRLysylBR	CCACATCCCCATTTAAATATCTGAGTAGAAAGTGTGCTGGTTGAATGG
5UTRLysylBF	TCTACTCAGATATTTAAATGGGGATGTGGGAACATACCATGTTACTCG
5UTRLysylBR	CGGCCCTAGTGGCCGGCAAACACATCGTCGCATTGAAAAC
SeqLysOxBF	CGAGCCTTACTATGGTGAATTGTG
SeqLysOxBR	CACACTCTCTCTCCATTGACAAG
seqLysylBnew	CAATTGCAACTGCAGACGAC
InLysOxBF	CTCTTTGGTGGCCTTAGCTGTG
InLysOxBR	GACTTCGTCCGAGTTGGTCATC
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
gapRIMfw	GTCCCTATTTCAATCAATTGAACAACTATCAAAACACAATGTTTGGAAACACAGGCTTGC

gapSGTfw	CCTATTTCAATCAATTGAACAACTATCAAAACACAATGAGTGCCACCAATAAAGAAGTAG
gapMOEPfw	TATTTCAATCAATTGAACAACTATCAAAACACAATGATCCCAAATTTATCATCTGGCATG
flagTTrev	AAATGGCATTCTGACATCCTCTTGAGCGGCCGCTTACTTATCGTCGTCATCCTTATAGTC
Construction of plasmid pGaHSwa	
OePGapAlphaR	GAAGGAAATCTCATTGTGTTTT
OePAox1GapR	CTACAAAAATTATTAGAGATTA
OePGapAlphaF	CAAAACACAATGAGATTTCCTT
OePAox1GapF	AAAATAATCTCTAATAATTTTT

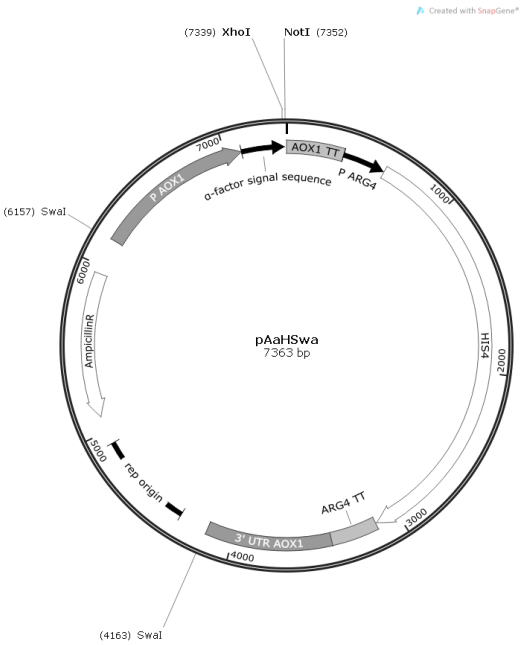


Supp. Fig. 1 Plasmid pPIC9toGAP with HRP-C1A CDS integrated between *Xho*I and *Not*I sites.

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.

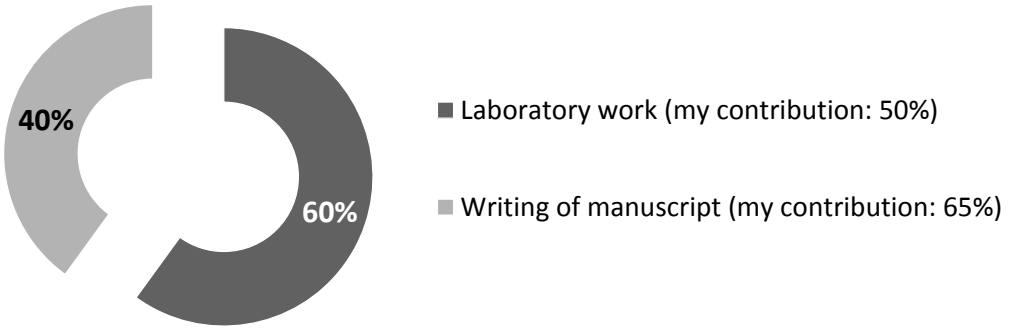


Suppl. Fig. 3 Expression plasmid pAaHSwa. The plasmids can be targeted to the *AOX1* locus by restriction with *Swa*I. The target gene is inserted downstream of P_{AOX1} and the α-mating factor signal sequence by cloning with *Xho*I and *Not*I. The *HIS4* gene acts as selectable marker.

CHAPTER 2

Random mutagenesis and screening

% Contribution



RANDOM MUTAGENESIS AND SCREENING

Christine M. Winkler[1], Joan Lin-Cereghino[2], and Geoff P. Lin-Cereghino[2]

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

[2] *Department of Biological Sciences, University of the Pacific, Stockton, CA, USA*

Summary: Random mutagenesis provides a valuable tool to elucidate gene function and to draw connections between gene products and molecular pathways. This chapter provides two protocols for the random mutagenesis of *Pichia pastoris*, along with methods to screen for the resulting mutant phenotypes. Genes are disrupted by insertion of a non-homologous DNA cassette. In the next step, primers designed to anneal to the cassette enable the identification of the mutagenized locus. We applied our system in *P. pastoris* to enhance recombinant protein secretion of horseradish peroxidase (HRP) and β-galactosidase, respectively. The subsequent screening identified unexpected genes that were later confirmed to benefit recombinant protein secretion, validating the substantive potential of the methodology.

Key words: *Pichia pastoris*, insertional mutagenesis, genome walking, restriction enzyme mediated integration (REMI), protein secretion

1. Introduction

Generating mutants is a strategy often used to identify gene products involved in cellular mechanisms. The process can be divided into three steps: 1) generating a mutation in the genomic DNA; 2) using a screen or selection to recognize those cells with a phenotypic change in the desired molecular mechanism; and 3) determining the locus with the mutation. Although chemical agents have been used successfully as mutagens in *P. pastoris (1)*, these are problematic for several reasons. These agents are usually strong carcinogens, posing a threat to the user. In addition, even if the desired phenotype is acquired, the identification of the mutated locus is difficult and time-consuming, especially if cloning by complementation is not straightforward. Therefore, new methods of mutagenesis are needed for *P. pastoris*, especially to make the recovery of the mutant gene easier. This chapter provides two protocols for generation of mutant strains along with ways to rapidly screen or select these cells for the desired phenotype and identify the responsible gene.

Both protocols involve insertional mutagenesis, in which the mutation of interest is physically marked by the integration of a non-homologous sequence of DNA. The advantages of insertional mutagenesis include that these integrations are usually 1) single copy so that only one locus in the cell is mutated; 2) stable, meaning that there is little rearrangement of the DNA in and around the insertion site after many generations; and 3) conservative in that none of the genomic DNA is deleted or duplicated around the insertion site *(2)*. Both protocols described in this chapter begin with random integration of a selectable marker gene into the *P. pastoris* genome to generate strains with enhanced secretion efficiency. However, the first method involves cultivation in multi-well format and a colorimetric assay to identify colonies with the desired mutant phenotype, while the second method singles out beneficial mutations through growth on selective media. The gene associated with the anticipated phenotype is then identified by PCR-based genome walking in the first and by plasmid rescue in the second method. Because the *P. pastoris* genome has been sequenced and is readily available through BOGAS and Genbank (http://bioinformatics.psb.ugent.be/orcae/ and http://www.ncbi.nlm.nih.gov/genbank/), these methodologies have allowed for the facile identification of genes involved in the regulation of protein secretion. The users should select the protocol that suits their needs the best. The

knowledge obtained from these techniques can be used to understand basic biology of *P. pastoris* and make this yeast an even better system for recombinant protein production.

2. Materials

2.1. Preparation of competent cells for electroporation

P. pastoris strain (choice depends on the marker gene located in the transformed DNA)

YPD: 1% yeast extract, 2% peptone, 1% dextrose

YPD/0.02 M HEPES

1 M dithiothreitol (DTT)

1 M sorbitol (sterile filtered)

Sterile water (ice cold)

Method 1: Mutagenesis of HRP-secreting strain by random integration of Zeocin resistance cassette

2.2. Construction of mutagenesis starting strain

P. pastoris strain GS115 *his4* (Life Technologies, Carlsbad, CA)

Vector pPICtoGAPαHRP (*see* **Note 1**)

*Swa*I restriction enzyme

Promega Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI)

Minimal dextrose (MD) agar: 1.34% YNB (w/o amino acids), 4×10^{-5}% biotin, 2% glucose and 2% agar.

2.3. Amplification and transformation of mutagenesis cassette

Vector pGAPZA (Life Technologies, Carlsbad, CA)

Primer TEFfw, 5'-CCCACACACCATAGCTTCAAAATG-3'

Primer CYC1rev, 5'-AGCTTGCAAATTAAAGCCTTCGAG-3'

Phusion High-Fidelity polymerase (Thermo Scientific, Waltham, MA)

Promega Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI)

BioRad MicroPulser electroporation apparatus

2 mm gap electroporation cuvettes

1 M sorbitol

YPD

YPD + Zeocin agar: 1% yeast extract, 2% peptone, 2% dextrose, 2% agar and 100 µg/mL Zeocin (InvivoGen, San Diego, CA)

2.4. Glycerol stocks of mutant library

96-well deep well plates (DWP) (Bel-Art Scienceware, Pequannock, NJ)

YPD + Zeocin: 1% yeast extract, 2% peptone, 2% dextrose and 100 µg/mL Zeocin (InvivoGen, San Diego, CA)

50% (w/v) glycerol

2.5. Pre- and main culture

DWP (Bel-Art Scienceware, Pequannock, NJ)

Multi-channel pipette

DWP shakers (Infors HT Multitron, Bottmingen-Basel, Switzerland)

BYPD: 1% yeast extract, 2% peptone, 2% dextrose, 200 mM potassium phosphate buffer, pH 6.0

2.6. HRP activity screen in culture supernatant

50 mM sodium acetate buffer, pH 4.5

20x ABTS stock: 550 mg ABTS dissolved in 50 mL of 50 mM sodium acetate buffer, pH 4.5 (to be stored at 4°C)

30% (w/w) hydrogen peroxide

96-well microtiter plates, polystyrene, flat-bottom

Spectrophotometer for 96-well plates

Multi-channel pipette

2.7. Isolation of genomic DNA

DWP (Bel-Art Scienceware, Pequannock, NJ)

DWP shaker (Infors HT Multitron, Bottmingen-Basel, Switzerland)

YPD

Glass beads, 0.25-0.5 mm

Lysis buffer: 2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris, pH 8.0, 1 mM EDTA

Phenol:chloroform:isoamylalcohol (25:24:1)

Disruptor Genie SI-D238 (Scientific Industries, Bohemia, NY)

TE buffer: 10 mM Tris, 1 mM EDTA, pH 8.0

EtOH, 95%

EtOH, 70%

RNAse A, 10 mg/mL

4 M ammonium acetate

Nuclease-free water

2.8. Template-blocking PCR

Primer CSF27, 5'-GACGCGTAATACGACTCACTATAGGGA-3'

Primer CSR30, 5'-ATCTCCCTATAGTGAGTCGTATTACGCGTC-3'

Primer CP, 5'-ACGCGTAATACGACTCACTATAGGGAGATC-3'

Primer GSPTEFa, 5'-TTCCAAACCTTTAGTACGGGTAATTAACGACAC-3'

Primer GSPTEFb, 5'-GCTGTGCTTGGGTGTTTTGAAGTGGT-3'

Primer GSPCYC1a, 5'-GAGTTAGACAACCTGAAGTCTAGGTCCCTA-3'

Primer GSPCYC1b, 5'-GTACAGACGCGTGTACGCATGTAACATTATAC-3'

Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI)

50 mM $MgCl_2$

1 M Tris-HCl, pH8

*Bgl*II, *Bam*HI or *Sau*3AI restriction enzymes

ddGTP, 10 mM

Klenow fragment, 10 U/µL

T4 DNA ligase (5 U/µL)

Maxima Hot Start Green PCR Master Mix (Thermo Scientific, Waltham, MA)

Method 2: Mutagenesis of β-galactosidase-secreting strain by restriction enzyme-mediated integration (REMI)

2.9. Random mutagenesis by REMI

pGAPZα B (Life Technologies, Carlsbad, CA)

pREMI-Z (NCBI accession number AF282723)

*Bam*HI restriction enzyme

Zymo Research Clean and Concentrator Kit (Irvine, CA)

Electroporator (model ECM 630, BTX Harvard Apparatus, Holliston, MA)

2 mm gap electroporation cuvettes

1 M sorbitol

YPD

YPD + Zeocin plates: 1% yeast extract, 2% peptone, 1% dextrose, 2% agar and 100 μg/mL Zeocin (Life Technologies, Carlsbad, CA)

2.10. Screening/selection of mutants

YN + lactose + histidine plates: 0.34% yeast nitrogen base, 1% ammonium sulfate, 2% agar supplemented with 1.0% lactose and 50 μg/mL histidine, adjusted to pH 6.7 with 1 M sodium phosphate

YN + methanol + histidine plates: 0.34% yeast nitrogen base, 1% ammonium sulfate, 2% agar supplemented with 0.5% methanol and 50 μg/mL histidine

X-Gal

2.11. Isolation of genomic DNA

YPD

Yeast Geno-DNA-Template DNA Extraction Kit (G-Biosciences, St. Louis, MO)

Chloroform

Ethanol, 95%

Ethanol, 70%

TE buffer: 10 mM Tris, 1 mM EDTA, pH 8.0

2.12. Plasmid rescue from mutant strains

*Eco*RI or *Hin*dIII restriction enzymes

QIAquick PCR Cleanup Kit (Qiagen,Valencia, CA)

T4 DNA ligase

One Shot MAX Efficiency DH10B competent *E. coli* cells (Life Technologies, Carlsbad, CA)

Lennox Broth (LB) liquid medium: 0.5% yeast extract, 1% glucose, 0.5% NaCl

Lennox Broth (LB) + Zeocin solidified medium: 0.5% yeast extract, 1% glucose, 0.5% NaCl, 2% agar supplemented with 25 µg/mL Zeocin

Zeocin (Life Technologies, Carlsbad, CA)

QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA)

*Nco*I and *Eco*RV restriction enzymes

2.13. Sequencing and bioinformatic analysis

M13 Forward (-20) primer: 5'- GTAAAACGACGGCCAG-3'

M13 Reverse primer: 5'- GGAAACAGCTATGAC-3'

3. Methods

3.1. Preparation of competent cells for electroporation

Both methods for mutagenesis require *P. pastoris* cells with a high transformation efficiency. Although a fairly rapid "condensed" procedure for this purpose has been developed *(3)*, a more lengthy, conventional method, modified from a protocol described by Cregg *(4)*, is used in order to produce cells that will give rise to a greater number of transformants. The following protocol should generate cells with a transformation efficiency of 10^4-10^5/µg DNA (*see* **Note 2**).

1. Inoculate 10 mL of YPD with a fresh *P. pastoris* colony from a plate and grow overnight with shaking at 30ºC (*see* **Note 3**).
2. The next morning, use the overnight culture to inoculate 500 mL YPD in a 2.8 L baffled flask to an OD_{600} of approximately 0.1.
3. Grow the 500 mL of cells to an OD_{600} of approximately 1.0 in a 30ºC shaker (*see* **Note 4**).
4. Pour the culture into sterile bottles and pellet by centrifugation for 10 min (perform this and all further centrifugation steps at 4,000 x g and 4ºC).
5. Resuspend the cells gently in 100 mL YPD/0.02 M HEPES.
6. Add 2.5 mL of 1 M DTT dropwise while swirling the cells.
7. Incubate 15 min with shaking at 30ºC.
8. Add ice cold water to 500 mL.
9. Pellet by centrifugation for 10 min.
10. Resuspend the cells in 500 mL of ice cold water.
11. Pellet by centrifugation for 10 min.
12. Resuspend cells in 250 mL of ice cold water.
13. Pellet cells by centrifugation for 10 min.
14. Resuspend cells in 20 mL of ice cold 1 M sorbitol and transfer them to a 50 mL conical tube.
15. Pellet by centrifugation for 10 min.
16. Resuspend the cells gently in 1.5-2.0 mL of ice cold 1 M sorbitol by flicking and inverting the tube. Do not vortex.

17. Aliquot the cells in convenient volumes (50 or 100 µL) into chilled 1.5 mL centrifuge tubes.
18. Wrap tubes of cells in several layers of paper towels and place inside a small styrofoam box. Place box in -80ºC freezer to allow the cells to freeze slowly.
19. When needed, thaw tubes of cells on ice.

Method 1: Mutagenesis of HRP-secreting strain by random integration of Zeocin resistance cassette

This protocol describes an efficient insertion mutagenesis method for random targeting of genes in *P. pastoris*. Mutant colonies are generated by integration of a linear Zeocin resistance cassette, previously amplified from plasmid pGAPZA (Life Technologies) by PCR (*see* **Note 5**). We demonstrate the feasibility of this method by mutagenizing strains secreting recombinant horseradish peroxidase (HRP), with subsequent screening for improved secretion to the culture supernatant. To simplify deep-well plate cultivation in the screening step, we chose to express HRP from the constitutive *GAP* promoter. However, in our experience, the results obtained from screening with P_{GAP} could be successfully transferred to P_{AOX1} *(5)*. The cassette insertion locus in mutants with enhanced secretory capacity is later identified by PCR-based genome walking.

3.2. Random mutagenesis by integration of Zeocin resistance cassette

3.2.1. Construction of mutagenesis starting strain

1. Linearize plasmid pPICtoGAPαHRP (**Fig. 1**) by restriction with *Sal*I, which cuts at a unique restriction site in the *HIS4* gene (*see* **Note 1**).
2. Confirm complete digestion on agarose gel before purifying the linear plasmid with the Promega Wizard SV Gel and PCR Clean-Up System.
3. Transform 600 ng of purified DNA into electrocompetent GS115 cells. Use a BioRad MicroPulser electroporation apparatus with “Pic” settings (2 mm gap cuvette, 2.0 kV pulse) for transformation.
4. Immediately after transformation add 500 µL of 1 M sorbitol and 500 µL of YPD and allow cells to recover at 28°C for 2 h.

5. Select transformants by plating the transformation mixture on MD plates for selection of histidine prototrophy. Streak to obtain single colonies.
6. Check several transformants for secretion of HRP employing the ABTS assay described in section ***3.3.3.***
7. Pick one HRP-secreting strain, which displays an HRP secretion level that is representative of a majority of strains, as the starting strain for mutagenesis.

3.2.2. Amplification and transformation of mutagenesis cassette

1. Amplify mutagenesis cassette from template pGAPZA with primers TEFfw and CYC1rev (**Fig. 2**). For Phusion High-Fidelity DNA polymerase use the following cycling parameters: 98°C/30 s – (98°C/5 s – 70°C/20 s – 72°C/20 s) x 35 – 72°C/7 min – 4°C/∞
2. Load PCR product on agarose gel. Excise and purify the 1172 bp band using the Promega Wizard SV Gel and PCR Clean-Up System.
3. Use 1-2 µg of purified PCR product to transform 80 µL of electrocompetent cells of the mutagenesis starting strain. Prepare the electrocompetent cells as described in section ***3.1.*** Use a BioRad MicroPulser electroporation apparatus with "Pic" settings (2 mm gap cuvette, 2.0 kV pulse) for transformation.
4. Immediately after electroporation, add 500 µL of 1 M sorbitol and 500 µL of YPD and allow cells to recover at 28°C for 2 h.
5. Plate different aliquots of transformation mixture on YPD + Zeocin agar plates, starting with 50 µL portions (*see* **Note 6**). It is preferred to have a low density of transformants on the plate to enable picking of single colonies (*see* **Note 7**). Incubate cells for 3 days at 28°C.
6. Repeat steps 1- 5 until a sufficient number of transformants is obtained. We screened around 3000 colonies; this number can be adjusted depending on the anticipated secretion enhancement.

3.2.3. Glycerol stocks of mutant library

1. Fill an adequate number of 96-well DWPs with 250 µL YPD + Zeocin per well, except for column 4. This column is reserved for reference strains and should be filled with 250 µL YPD w/o Zeocin per well (*see* **Note 8**). The Zeocin is added to prevent contamination and

to confirm the presence of the marker cassette in the genome. Using multi-channel pipettes is highly recommended for all pipetting steps involving 96-well plates.

2. Inoculate wells using sterile toothpicks. Column 4 is reserved for sterile controls, negative control GS115 (*his4* Zeo$^-$) and mutagenesis starting strain GS115 + pPICtoGAPαHRP (*HIS4* Zeo$^-$). All other wells are inoculated with mutagenesis cassette transformants (*HIS4* Zeo$^+$).
3. Incubate the DWPs for 48 h at 28°C, 320 rpm, 80% humidity on shaker (*see* **Note 9**).
4. Fill two 96-well microtiter plates per DWP with 100 µL of 50% glycerol per well. One will be your working plate while the other will serve as the backup to retrieve identified hits. Transfer 100 µL of cell culture per well from the DWP to each microtiter plate and mix carefully by pipetting up and down. Freeze at -80°C.

3.3. Deep-well plate cultivation and HRP activity screening

Mutants are cultivated on buffered YPD (BYPD), since the reporter protein, HRP, is expressed from P_{GAP}. Alternatively, an induction protocol for expression from P_{AOX1} is described in the "Notes" section. The HRP activity screen is based on the protocol described in Morawski et al. *(6)*.

3.3.1. Pre-culture

1. Thaw "working plates" of the glycerol stocks prepared in section ***3.2.3.***
2. Fill 96-well DWPs with 250 µL BYPD per well.
3. Inoculate BYPD with 10 µL of glycerol stock.
4. Grow cells for 48 h (*see* **Note 9**) at 28°C, 320 rpm, 80% humidity.

3.3.2. Main-culture

1. Fill 96-well DWPs with 500 µL BYPD per well (*see* **Note 10**).
2. Inoculate wells with 10 µL of pre-culture.
3. Grow cells for 48 h (*see* **Note 11**) at 28°C, 320 rpm, 80% humidity.
4. Prior to activity screen, measure OD_{600} after diluting 1:20 (*see* **Note 12**).

3.3.3. HRP activity screen in culture supernatant

1. Prepare 1x ABTS screening solution: per screened DWP, mix 19 mL sodium acetate buffer, 1 mL of 20x ABTS stock and 1.75 µL of 30% (w/w) H_2O_2 (*see* **Note 13**).
2. Centrifuge DWPs for 10 min at 1,500 x g, 22°C.
3. Transfer 15 µL of supernatant into 96-well microtiter plates; be careful not to touch the cell pellet.
4. Immediately before analysis, add 140 µL of 1x ABTS screening solution per well to start reaction. The uncolored solution will instantly start to turn turquoise.
5. Measure absorption kinetics at 405 nm in a spectrophotometer for 2-5 min.
6. Calculate activity/OD_{600} and retrieve the best hits from the "backup plate" of the glycerol stocks prepared in section ***3.2.3.*** We selected all the strains that had at least double the activity of the wild type control.
7. Repeat cultivation and screening steps for best hits in subsequent rounds of rescreening.

3.4. PCR-based genome walking

The described protocol for the isolation of genomic DNA (gDNA) is based on the method devised by Hoffman and Winston *(7)*. To avoid nonspecific amplification in the genome walking step, we adapted the template-blocking PCR protocol of Bae and Sohn *(8)*. In this protocol the 3' ends of the restricted genomic DNA fragments are blocked with dideoxynucleoside triphosphate (ddNTPs) prior to the ligation of appropriate adaptor cassettes (**Fig. 3**). This step prevents generation of additional binding sites for the adaptor primer, which could be formed by filling in the recessive ends of the adaptor-ligated restriction fragments with DNA polymerase, ultimately resulting in nonspecific amplification. The unknown gDNA sequence between adaptor and $Zeocin^R$-mutagenesis cassette is PCR-amplified and sequenced using primers that bind in these two elements. Subsequently, the resulting sequences, representing the mutagenesis cassette insertion loci, are mapped using NCBI BLAST.

3.4.1. Isolation of genomic DNA

1. Fill the DWP with 600 µL YPD per well. Inoculate wells with best hit *P. pastoris* mutant strains.
2. Grow cells 24 - 36 h at 28°C with 320 rpm, 80% humidity.
3. Transfer the culture volumes to 1.5 mL microcentrifuge tubes (*see* **Note 14**).

4. Spin for 1 min at max. speed (>13,000 x g) in a table top centrifuge.
5. Decant supernatant and add approximately 0.3 g of glass beads, 150 µL of lysis buffer and 150 µL of phenol/chloroform/isoamyl alcohol to the cell pellet.
6. Vortex for at least 8 min using the Disrupter Genie (*see* **Note 15**).
7. Add 150 µL of TE buffer and spin tubes for 5 min at max. speed.
8. Transfer the upper aqueous phase to a fresh 1.5 mL microcentrifuge tube. Precipitate by adding 1 mL of ice-cold EtOH and incubating at -20°C for 30 min.
9. Spin the tubes for 5 min at max. speed and decant the supernatants.
10. Dry the pellets at 60°C until the EtOH has completely evaporated.
11. Resuspend the pellets in 400 µL of TE buffer and add 5 µL of RNAse A to degrade the RNA. Incubate 2 h at 37°C.
12. Precipitate the gDNA by adding 10 µL of 4 M ammonium acetate and 1 mL of ice-cold EtOH.
13. Spin 5 min at max. speed to pellet the gDNA. Remove the supernatants and wash the pellets with 1 mL of 70% EtOH. Spin again.
14. Dry pellets to remove any residual EtOH.
15. Dissolve each pellet in 50 µL of distilled water.

3.4.2 Template-blocking PCR

1. A double-stranded adaptor cassette is generated by annealing the two primers CSF27 and CSR30. This cassette possesses a three bases (GAT) overhang at its 5' end that is complementary to one-base-filled *Bam*HI-, *Bgl*II-, and *Sau*3AI-digested fragments, respectively. To assemble the adaptor, mix 20 µL of 100 µM primer CSF27, 20 µL of 100 µM primer CSR30, 10 µL of 50 mM $MgCl_2$ and 1.25 µL of 1 M Tris-HCl, pH 8. Heat to 100°C for 5 min and allow to slowly cool down.
2. Digest 2 µg of gDNA with either *Bgl*II, *Bam*HI or *Sau*3AI at 37°C overnight in a final volume of 34 µL.
3. Add 0.8 µL ddGTP and 1 µL of Klenow fragment to the reaction mixture. Incubate for 3 h at 37°C.
4. Purify the Klenow treated DNA using the Promega Wizard SV Gel and PCR Clean-Up System, following the manufacturer's instructions. Elute in 55 µL of nuclease-free water.

5. Mix the eluted DNA with 1 µL of the adaptor (prepared in step 1), 6 µL of T4 ligase buffer (10x) and 1 µL T4 ligase (5 U/µL). Incubate overnight at 16°C for efficient ligation.
6. Purify the DNA again using the Promega Wizard SV Gel and PCR Clean-Up System. Elute in 40 µL TE buffer.
7. Perform template-blocking PCR in forward and reverse direction with primer pairs CP+GSPTEFa and CP+GSPCYC1a, respectively. As template, use the restricted gDNA with ligated adaptor produced in steps 5 and 6. We use Thermo Scientific's Maxima Hot Start Green PCR Master Mix and the following cycling parameters for the PCR: 95°C/4 min – (95°C/30 sec – 60°C/50 sec – 72°C/4 min) x 35 cycles – 72°C/10 min – 4°C/∞. An example result is shown in **Fig. 4**.
8. Load a fraction of the PCR on an agarose gel. If the PCR resulted in a band, the product of the template-blocking PCR can be used as template for a second, "nested" PCR reaction with the nested primer pairs CP+GSPTEFb and CP+GSPCYC1b, respectively (Cycling parameters: 95°C/4 min – (95°C/30 sec – 60°C/50 sec – 72°C/1 min per kb, depending on largest fragment to be amplified) x 40 cycles – 72°C/10 min – 4°C/∞).
9. Load the full volume of the nested PCR on an agarose gel. Excise the product band and purify DNA using the Promega Wizard SV Gel and PCR Clean-Up System.
10. Sequence the band employing the same primers that were used for amplification.

3.4.3. Bioinformatics analysis

1. Identify the integration locus of the mutagenesis cassette by using NCBI BLASTn. Compare the obtained sequencing results to the genome sequence of *P. pastoris* GS115 (*see* **Note 16**).
2. Use BLASTp to compare the corresponding amino acid sequence of the disrupted ORF (or ORFs in the vicinity of the integration site, if the cassette had integrated into an intergenic region) to proteins of other related yeasts. This step is recommended, since the annotation of proteins in *P. pastoris* is sometimes incomplete or incorrect.

We confirmed the beneficial effect of gene disruptions identified in the screening by generating targeted knockouts of selected genes *(5)*. The enhanced secretory performance of these knockout strains relative to the wild type control in repeated rounds of screening is depicted in **Fig 5.**

Method 2: Mutagenesis of β-galactosidase-secreting strain by restriction enzyme-mediated integration (REMI)

A second method for generating mutants in the electrocompetent *P. pastoris* cells is based on restriction enzyme-mediated integration (REMI), a technique which has been used in many organisms *(9, 10)*. This method involves 1) the random integration of a linearized plasmid containing a selectable marker into the *P. pastoris* genome; 2) the identification of mutant colonies with a desired phenotype; and 3) the retrieval of the circularized plasmid from the genomic DNA of the mutant strain and analysis of its DNA sequence to determine the disrupted gene (**Fig. 6**).

Although the details may differ depending on the microbe used, the integration event can be generally viewed in three steps: 1) during transformation, the restriction enzyme enters the nucleus along with a linearized plasmid whose ends have been generated by the same enzyme; 2) the restriction enzyme digests the host chromosome at its specific recognition site; and 3) the complementary ends of the chromosome and the linearized plasmid anneal and are consequently ligated by cellular enzymes to create a non-homologous integration *(2)* (*see* **Note 17**). The REMI protocol has been described previously *(11)*. The plasmid used in this procedure, pREMI-Z plasmid (a generous gift of Ben Glick, University of Chicago), is a 1,985 bp vector containing a Zeocin resistance gene under the control of the *Saccharomyces cerevisiae TEF* and bacterial *EM7* promoters, Col E1 origin of replication and a unique *Bam*HI site, allowing for transformation of *E. coli* and *P. pastoris*.

3.5. Random mutagenesis by REMI

1. Digest 5 µg of pREMI-Z with *Bam*HI. A small portion of the reaction mixture should be run on a 1% agarose gel to confirm linearization.
2. Purify the digested plasmid to remove salts and enzyme using a Zymo Research Clean and Concentrator Kit.
3. Quantify the purified plasmid by absorbance at 260 nm.
4. Mix approximately 1 µg of the *Bam*HI-linearized plasmid with 50 µL of freshly thawed, electrocompetent cells and 1 unit of *Bam*HI in a 2 mm gap cuvette (*see* **Note 18**).

5. After incubating the cuvette on ice for 5 min, electroporate using standard conditions (1.5 kV, 50 uF, 200Ω) with a BTX electroporation generator.
6. Immediately after electroporation, add a mixture of 500 µL of 1 M sorbitol and 500 µL of YPD to the cuvette. Transfer the cell suspension to a 1.5 mL microfuge tube allowing it to recover for 1 h in a shaking incubator at 30ºC.
7. Plate the transformation mixture in 100 µL aliquots on YPD + Zeocin plates. Plates should be incubated for 2-3 days until colonies appear (expect approximately 500-1,000 colonies/plate) (*see* **Note 19**).

3.6. Screening/selection of mutants

With the goal of isolating strains with enhanced secretion, a plasmid encoding an α-mating factor (α-MF)-β-galactosidase fusion was expressed from the *GAP* promoter and integrated into *P. pastoris* strain yDT39 *his4 met2 (12)*. The plasmid contained *MET2* as a selectable marker. While the α-MF acts as a secretion signal, β-galactosidase serves to convert lactose, which cannot be utilized by *P. pastoris*, into galactose and glucose, which can be used metabolically. In wild type cells, the protein is retained within the cell. Thus, when these cells are plated on medium containing lactose as a sole carbon source, they fail to grow because the enzyme is not efficiently secreted. These cells were made electrocompetent and mutagenized by the REMI method, as described above. If the strains harboring pREMI-Z insertion mutations allowed export of the β-galactosidase protein, they would be expected to grow on lactose **(Fig. 7).** This selection was performed as follows:

1. Replica-plate colonies from YPD+Zeocin plates to YN+lactose+histidine plates. A non-mutagenized, wild type strain containing an integrated pGAPZα B plasmid (which carries the Zeocin resistance gene) should be used as a negative control (*see* **Note 20**).
2. Incubate the YN+lactose+histidine plates at 30 ºC for 3-4 days. Supersecreting colonies should grow significantly better than the negative control strain (*see* **Note 21**).
3. Restreak the putative lactose-utilizing colonies on YN+lactose+histidine plates to confirm robust growth; however, mutant strains will grow at different rates.

4. Transfer colonies from lactose plates to YN+methanol+histidine plates supplemented with X-Gal (80 μg/mL) to confirm β-galactosidase secretion, as evidenced by the formation of a blue halo around the colony *(13)* (*see* **Note 22**).
5. Other reporter constructs (containing the G418 resistance gene or a prototrophic selectable marker gene if the strain carries auxotrophies) can be transformed into each mutant strain and assayed to see if the strain gives the expected phenotype. For instance, a α-MF-horseradish peroxidase (HRP) fusion can be expressed to see if its secretion is enhanced in a mutant strain compared to a wild type strain *(12)*.

3.7. Identification of mutagenized locus by plasmid rescue

As a first step, genomic DNA is isolated from each mutant using components of the Yeast Geno-DNA-Template DNA Extraction Kit (G-Biosciences, St. Louis, MO) with some modifications of the manufacturer's protocol. The genomic DNA of each strain should be analyzed by Southern blotting, using a probe generated from the pREMI-Z plasmid to confirm that the chromosomal DNA of each strain contains only one pREMI-Z insertion. A detailed description of the Southern blot method would go beyond the scope of this article, but standard protocols can be found in several sources (*15*). Strains displaying a single band on the Southern are then subjected to plasmid rescue. The genomic DNA of these mutant strains is digested with a single restriction enzyme to generate fragments with self-annealing ends. The fragments will be allowed to anneal and be ligated. Only those circularized fragments containing the pREMI-Z sequence and flanking genomic DNA should be able to transform *E. coli* and confer Zeocin resistance (**Fig. 6**).

3.7.1. Isolation of genomic DNA

1. Grow *P. pastoris* colonies in 3 mL of YPD in 50 mL conical tubes in a 30°C shaker overnight.
2. The next day, pellet 1.5 mL of the liquid cultures in a microfuge tube.
3. After the supernatant is removed, add 150 μL of Yeast Suspension Buffer, 5 μL of LongLife Zymolyase, and 1 μL of β-mercaptoethanol to each cell pellet. The reagents, as well as those in the later steps, are provided by the Yeast Geno-DNA-Template DNA Extraction Kit.

4. Resuspend each pellet and incubate in a 37°C water bath for 1 h, occasional flicking the tubes to mix the samples.
5. Centrifuge the samples at 16,000 x g for 5 min and discard the supernatant. Resuspend the pellets in 150 µL of sterile water.
6. Add 400 µL of Lysis Buffer and invert several times before the addition of 5 µL of LongLife Proteinase K solution.
7. Incubate at 60°C for 3 h with mixing by inversion every 30 min (*see* **Note 23**).
8. After the samples cool to room temperature, add 200 µL of chloroform and vortex well.
9. Centrifuge for 5 min at 16,000 x g.
10. Remove the upper aqueous phase and transfer it to a new microfuge tube.
11. Add 60 µL of DNA Stripping Solution. Invert several times and incubate at 60°C for 10 min.
12. Cool samples to room temperature and add 120 µL of Precipitation Solution. Mix by inversion.
13. Centrifuge samples for 6 min at 16,000 x g.
14. Transfer the supernatant to a new tube and add 700 µL of 95% ethanol.
15. Invert each tube 20 times to precipitate the genomic DNA (avoid vortexing which will shear the DNA) and then centrifuge for 1 min at 16,000 x g to pellet the DNA.
16. Remove the ethanol and add 1 mL of 70% ethanol to wash the pellet.
17. Invert and flick the tube to dislodge the pellet from the bottom.
18. After centrifugation at 16,000 x g for 1 min, discard the 70% ethanol and leave each tube open and inverted for 5 min in a 37ºC oven to remove any residual ethanol (*see* **Note 24**).
19. Add 200 µL of TE buffer and 1 µL of LongLife RNAse to each sample. The DNA should be allowed to rehydrate for 30 min at 37°C with intermittent inversion (no vortexing).
20. Determine the DNA concentration by measuring absorption at 260 nm. Most samples are usually 1-2 µg/µL.
21. Store at -80°C.

3.7.2. Plasmid rescue from mutant strains

1. Digest approximately 3 µg of genomic DNA with either *Eco*RI or *Hin*dIII for 3 h at 37°C.

2. Purify the fragments with the QIAGEN QIAquick PCR Cleanup Kit and elute in 30 µL (*see* **Note 25**).
3. Incubate 1 µg of each purified, digested DNA with T4 DNA ligase and appropriate reaction buffer in a total volume of 20 µL overnight at room temperature.
4. Transform 3 µL of the ligation mix into One Shot MAX Efficiency DH10B competent *E. coli* cells according to the manufacturer's instructions. These cells are especially suitable for transformation by large plasmids.
5. Plate the transformation reaction on LB (Lennox Broth, low salt) + 25 µg/mL Zeocin solidified medium and incubate overnight at 37 ºC.
6. Grow up at least six of the resulting colonies in liquid LB + 25 µg/mL Zeocin medium for 18 h at 37 ºC (*see* **Note 26**).
7. Isolate the plasmids using QIAprep Spin Miniprep Kit and follow the special directions for elution of large plasmids.
8. Digest the plasmids with a restriction enzyme(s) that recognizes a site in pREMI-Z (i. e. *Nco*I and *Eco*RV) and compare the resulting restriction patterns on an agarose gel (**Fig. 8 and** *see* **Note 27**).

3.7.3. Sequencing and bioinformatic analysis

1. Use the M13 Forward primer (-20) and M13 Reverse primer to sequence the flanking genomic DNA in the rescued plasmids. These primer sites flank both sides of the original *Bam*HI site that is located in pREMI-Z (**Fig. 6**). At least 4 plasmids originating from the plasmid rescue of each mutant strain should be sequenced.
2. The National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) and the Bioinformatics Online Genome Annotation System (BOGAS, http://bioinformaticspsb.ugent.be/webtools/bogas) should be used to determine the gene disrupted in the mutant strains by the pREMI-Z sequence based on the identity between the sequenced DNA and the published *P. pastoris* genome. The genes identified by this selection strategy have been previously described (*12*).
3. Amino acid sequence homology to other cloned genes in the database should be analyzed with the BLAST program. Expert Protein Analysis System (ExPASy) at http://www.expasy.org/ can be used to analyze the predicted amino acid sequences of the

identified genes for motifs, such as kinase, DNA binding domains, and other structural analysis, which will provide clues to the function of the *P. pastoris* gene product.

4. Notes

1. We constructed the plasmid pPICtoGAPαHRP (**Fig. 1**), a chimera of the two vectors pPIC9 and pGAPZαA, which combines the two features *GAP* promoter and *HIS4* selection marker. The vector has the HRP gene inserted between *Xho*I and *Not*I restriction sites *(5)*.
2. Electrocompetent cells can also be prepared by the condensed protocol, but the lower transformation efficiency of these cells will require many more electroporation reactions to achieve the same number of transformants.
3. Make sure that the strain is sensitive to Zeocin and has auxotrophic markers that will permit transformation with plasmids later.
4. Cells can be harvested at any OD_{600} between 0.7 and 1.3; however, an OD_{600} outside this range will lead to lower transformation efficiencies.
5. We also tested the applicability of a *KanMX6* marker cassette for mutagenesis, conferring resistance to Geneticin. We observed random integration into genomic DNA, similar to the Zeocin resistance cassette. However, elevated levels of transformation background on YPD plates containing 300 µg/mL Geneticin made us proceed with the Zeocin resistance marker instead.
6. We advise spreading the transformation mixture on plates by using glass beads, 2.85-3.3 mm (e.g. Roth).
7. If available, a colony picking robot can facilitate the picking procedure. Spread 500 µL of transformation mixture on agar prepared appropriate bioassay trays. Use the picking head recommended by the supplier for picking the yeast colonies to the DWP.
8. Several thousand of colonies have to be picked and screened to theoretically cover all nonessential genes of *P. pastoris*. Therefore, we recommend dividing the picking procedure into several working days, depending on the available capacities, e.g. automated pickers, DWP shakers etc. The transformation plates should be stored at 4°C during this time.
9. It is crucial for the outcome of the later screening that cells grow to similar cell density. Grow them to saturation and possibly control OD_{600} after diluting 20-fold. Eventually extend the incubation time until all wells show the same density.

10. In case of expression from P_{AOX1}, pre-grow cells for 32 h in 200 µL of $BMG_{lucose}Y$ (1% yeast extract, 2% peptone, 2% glucose, 0.2 M potassium phosphate buffer, pH 6.0, 1.34% YNB, 4×10^{-5}% biotin) per well. Start the induction by adding 200 µL of BMMY2 (1% yeast extract, 2% peptone, 2% methanol, 0.2 M potassium phosphate buffer, pH 6.0, 1.34% YNB, 4×10^{-5}% biotin). After another 12, 24 and 36 h, add 50 µL of BMMY10 (as BMMY2, but 10% methanol) to keep methanol concentration at ~1% final concentration. Harvest cells 48 h after the start of induction *(14)*.
11. Cultivation time can be adjusted depending on the expression level of the secreted protein.
12. Gene disruptions can cause reduced growth rates. Measuring OD_{600} is important, since altered HRP activity in the culture supernatant could stem from altered growth behavior.
13. 1x ABTS screening solution should always be prepared freshly. Keep on ice and protect from light until use in the assay.
14. Volumes of >200 µL are easily transferred from the DWPs by using a 200 µL-tip on top of a 1000 µL-tip.
15. Vortexing without the Disrupter Genie results in reduced cell lysis and, subsequently, reduced yields.
16. We noticed that, in some cases, the genome sequence and protein annotation differs between the sequences of GS115 and CBS7435 available in the NCBI database. It can therefore be beneficial to compare results for the two strains.
17. While the activity of the restriction enzyme has been shown to be required for this process in *Saccharomyces cerevisiae*, the restriction enzyme is not needed for REMI of some fungal species, such as *Ustilago maydis*. It is not clear if the addition of the restriction enzyme enhances the transformation efficiency with the linearized plasmid in *P. pastoris*.
18. Insufficient enzyme leads to lower transformation efficiency while excess enzyme has been associated with genetic damage and multiple insertion events *(2)*.
19. To analyze sufficient colonies, 5 electroporation reactions were initially performed, yielding approximately 50 plates in total.
20. If another *P. pastoris* strain is used which has other nutritional auxotrophies, make sure to include the necessary amino acids in the growth medium.

21. To save time and effort, electroporated cells (step # 6 above) can, instead, be allowed to recover in 1 mL of 1 M sorbitol for 1 h and then be plated directly onto YN+lactose+histidine plates.
22. β-galactosidase, which is secreted by cells, metabolizes the X-Gal in the medium to form a blue product on the periphery of the column (blue halo). Colonies that do not show the supersecretion phenotype appear blue but show no blue halo.
23. This is the most crucial step to attain high yields of DNA. The samples should become clear over time as the cells lyse. If not, vortex them mildly to accelerate the process. By the end of three hours, all solids should be gone, and the tube contents should be translucent and viscous. If not, continue incubation.
24. Excess ethanol will inhibit digestion of the genomic DNA in the later steps, but the pellet should not be completely dried or it will be tough to dissolve.
25. The fragments will be circularized and ligated, and then will be transformed into competent *E. coli*. If the resulting circularized pREMI-Z and flanking DNA are too large, the ligation product will be unable to transform the *E. coli*. Therefore, digesting the genomic DNA with each enzyme separately increases the chances that at least one recircularized plasmid will be small enough for transformation. Any restriction enzyme that does not cut within pREMI-Z can potentially be used for this purpose.
26. In some cases, presumably because the plasmid is so large, less than six transformant colonies will be produced.
27. One would expect that all plasmids originating from the same transformation reaction would be the same and therefore produce identical restriction patterns. However, extensive restriction analysis has indicated that a minority (<25%) of the plasmids may contain a random genomic fragment that inserted prior to circularization or show rearrangement of the flanking genomic and plasmid DNA (**Fig. 8**). These plasmids should be ignored. In rare cases, where the plasmid restriction patterns fall into two distinct groups, then representatives of both groups should be pursued for sequencing.

References

1. Tolstorukov I., Cregg J.M. (2007) Classical Genetics. In: Cregg JM (ed) Pichia Protocols, 2nd edn. Humana Press, New Jersey, pp. 189-201.

2. Riggle P.J., Kumamoto C.A. (1998) Genetic analysis in fungi using restriction-enzyme-mediated integration. *Current opinion in microbiology* 1(4):395-9. Epub 1999/03/06. PMID: 10066514.

3. Lin-Cereghino J., Wong W.W., Xiong S. et al. (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*. *BioTechniques* 38(1):44, 6, 8. Epub 2005/02/01. PMID: 15679083; PMCID: PMC2504082.

4. Cregg J.M., Russell K.A. (1998) Transformation. In: Higgins DR, Cregg JM (ed) Pichia Protocols, Humana Press, New Jersey, pp. 27-39.

5. Winkler C.M., Ahmad M., Neugebauer P. et al. *Manuscript in preparation*

6. Morawski B., Lin Z., Cirino P., et al. (2000) Functional expression of horseradish peroxidase in *Saccharomyces cerevisiae* and *Pichia pastoris*. *Protein Eng* 13:377–84. doi: 10.1093/protein/13.5.377

7. Hoffman C.S., Winston F. (1987) A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformaion of *Escherichia coli*. *Gene* 57:267–272. doi: 10.1016/0378-1119(87)90131-4

8. 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

9. Manivasakam P., Schiestl R.H. (1998) Nonhomologous end joining during restriction enzyme-mediated DNA integration in *Saccharomyces cerevisiae*. *Molecular and cellular biology* 18(3):1736-45. Epub 1998/03/06. PMID: 9488490; PMCID: PMC108888.

10. Kuspa A. (2006) Restriction enzyme-mediated integration (REMI) mutagenesis. *Methods Mol Biol* 346:201-9. Epub 2006/09/08. PMID: 16957292.

11. Schroder L.A., Glick B.S., Dunn W.A. (2007) Identification of pexophagy genes by restriction enzyme-mediated integration. *Methods Mol Biol* 389:203-18. Epub 2007/10/24. PMID: 17951645.

12. Larsen S., Weaver J., de Sa Campos K., et al. (2013) Mutant strains of *Pichia pastoris* with enhanced secretion of recombinant proteins. *Biotechnol Lett*. doi: 10.1007/s10529-013-1290-7

13. Staley C.A., Huang A., Nattestad M. 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(2):118-27. Epub 2012/01/31. PMID: 22285974; PMCID: PMC3464313.

14. Weis R., Luiten R., Skranc W., et al. (2004) Reliable high-throughput screening with *Pichia pastoris* by limiting yeast cell death phenomena. *FEMS Yeast Res* 5:179–89. doi: 10.1016/j.femsyr.2004.06.016

15. Sambrook, J, Fritsch, EF and Maniatis, T (1989). *Molecular cloning: A Laboratory Handbook*, 2nd edn. Cold Spring Harbor Laboratory Press.

Figures

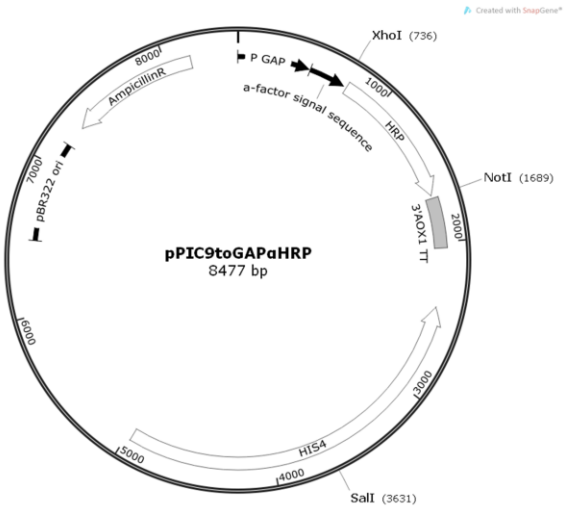


Fig. 1. Expression vector pPICtoGAPαHRP

Fig. 2. Schematic drawing of Zeocin resistance mutagenesis cassette. 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.

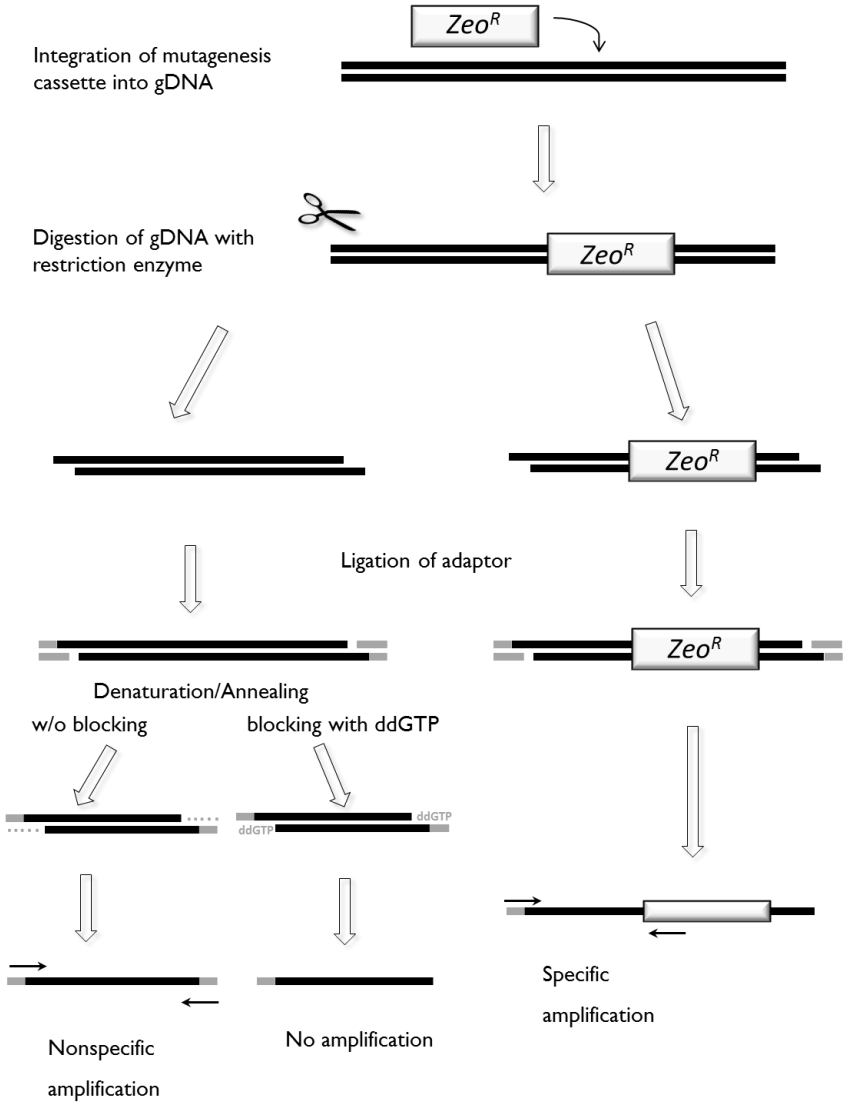


Fig. 3. Template-blocking PCR scheme: Upon transformation the mutagenesis cassette integrates randomly into the genome. Isolated gDNA is then cut with restriction enzymes, yielding fragments that contain the mutagenesis cassette and fragments that do not. In the next step, the adaptor is ligated to the complementary sticky ends of the restricted gDNA. If the fragment contains the mutagenesis cassette, the genomic sequence between adaptor and mutagenesis cassette can be amplified using primers that bind in these two elements. Through rounds of denaturation and annealing, DNA polymerase can potentially generate additional binding sites for the adaptor primer by filling in recessive ends, ultimately resulting in nonspecific amplification products of fragments that do not contain the mutagenesis cassette. This process is prevented

when 3' ends of the restricted gDNA are blocked with dideoxynucleoside triphosphate (ddNTPs) prior to the ligation of appropriate adaptor cassettes.

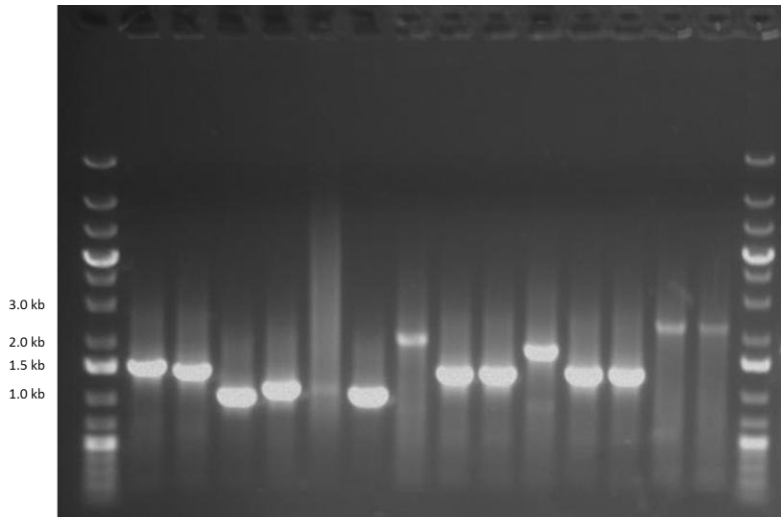


Fig. 4. A typical result of Template-blocking PCR

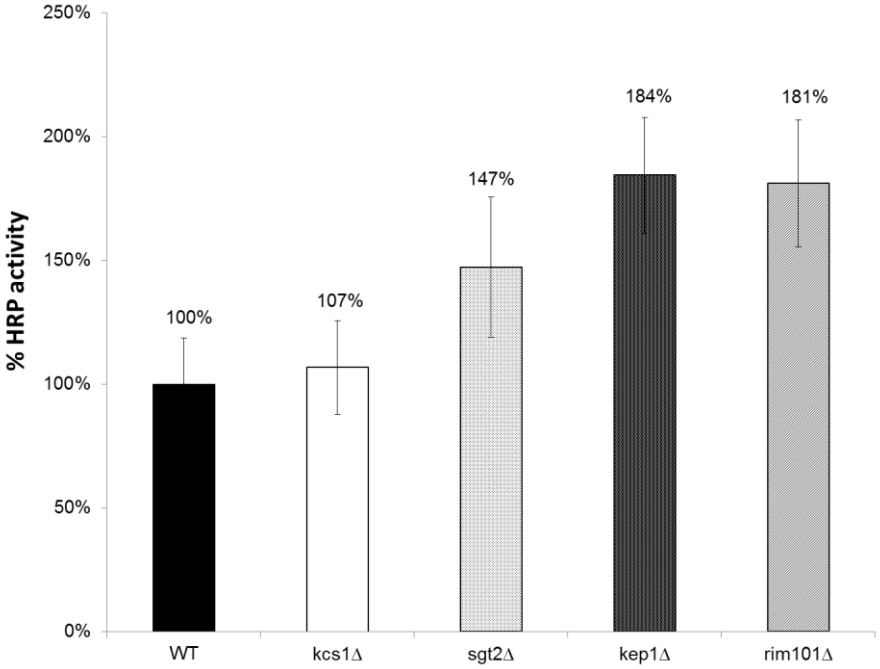


Fig. 5. Quantification of secreted HRP in culture supernatants of selected knockout strains. The genes *KCS1*, *SGT2*, *KEP1* and *RIM101* had previously been identified in a screening for enhanced secretion of HRP, following random mutagenesis of the expression strain. Relative HRP activity upon expression from *GAP* promoter. WT control and mutant strains expressing HRP from P_{GAP} were grown on BYPD (2% glucose) in DWP for 35 h before analysis. Peroxidase activity in supernatants was measured with ABTS assay. Results represent the mean of four biologically independent experiments with 12 technical replicates per experiment.

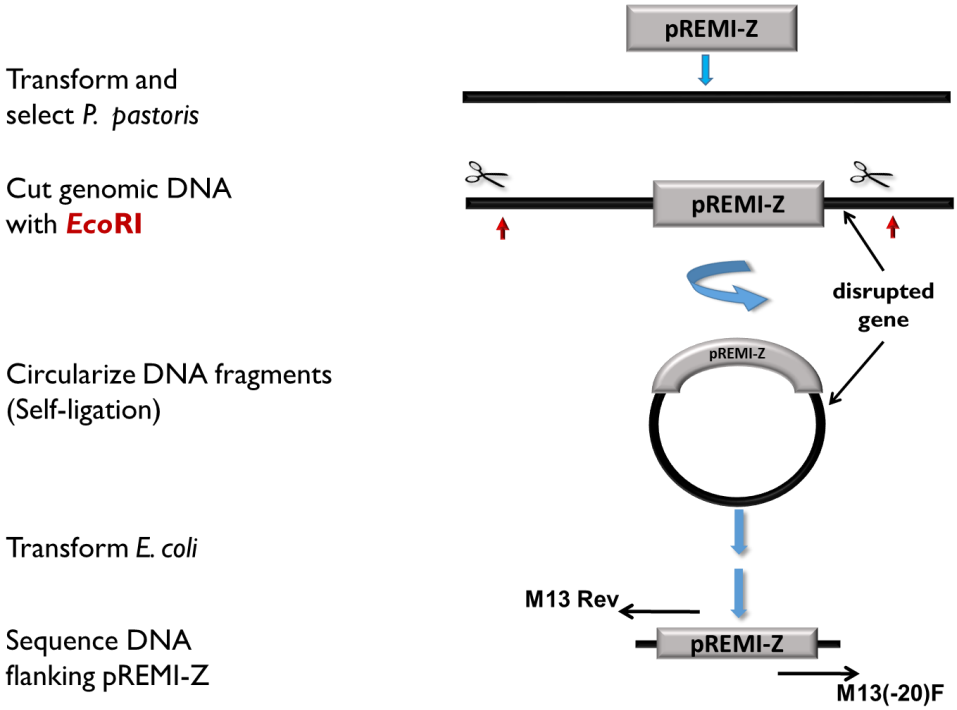


Fig. 6. Key steps in the REMI method

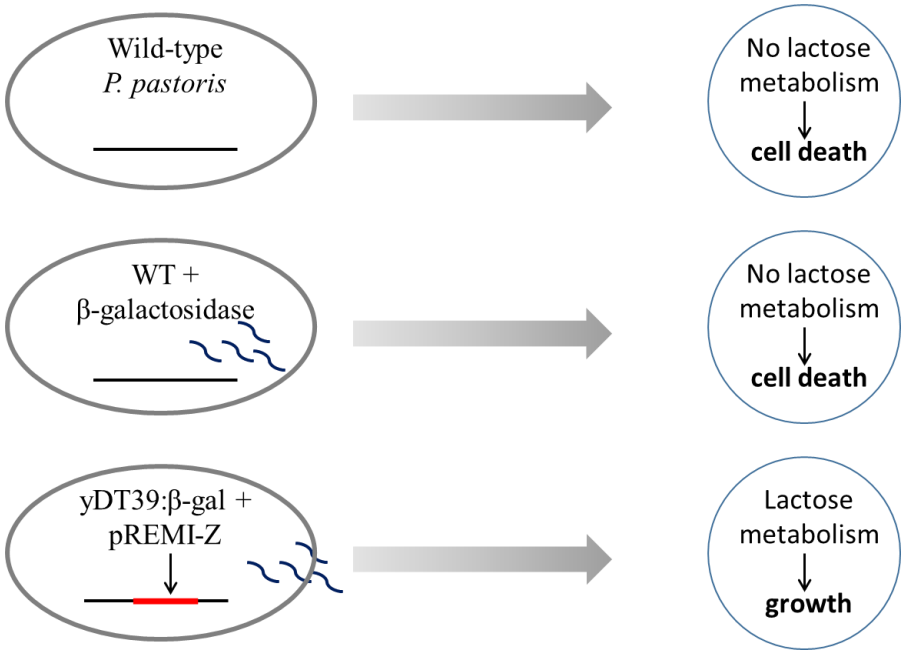


Fig. 7. Strategy to isolate strains that supersecrete β-galactosidase

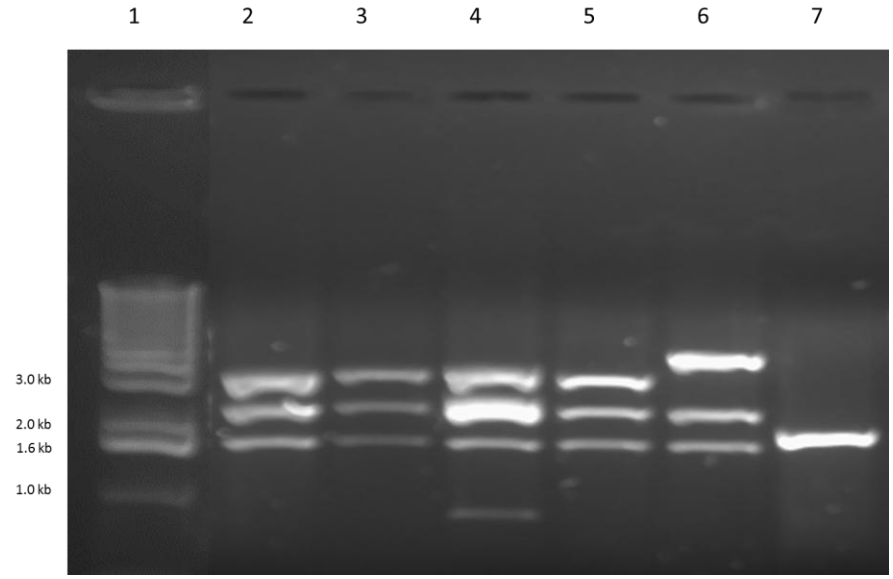
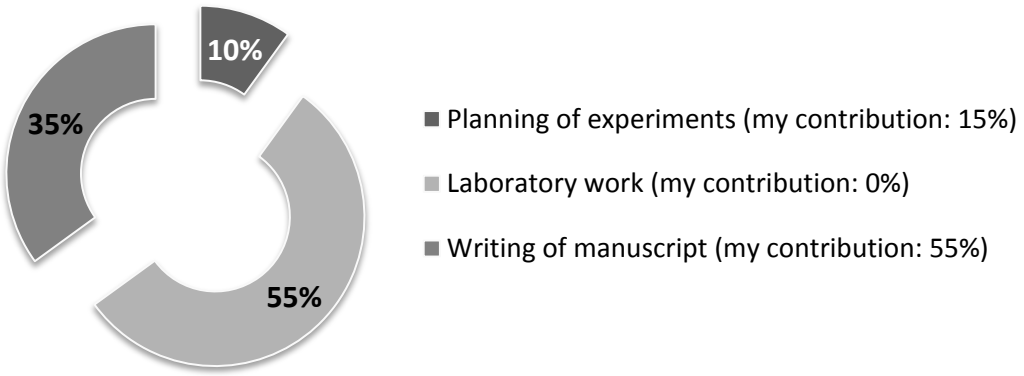


Fig. 8. Restriction analysis of plasmids isolated from colonies resulting from a plasmid rescue. Plasmids were digested with *Nco*I and *Eco*RV and then resolved on a 0.8% agarose gel. Lane 1: 1kb molecular weight marker; lanes 2-6: plasmids isolated from colonies; lane 7: pREMI-Z. Lanes 2, 3 and 5 contain the pREMI-Z plasmid with the same flanking genomic DNA while the plasmids in lanes 4 and 6 contain an additional insert along with this flanking genomic DNA.

CHAPTER 3

Strategy for targeted knockouts in *Pichia pastoris*

% Contribution



A NOVEL SET OF VECTORS FOR CREATING MULTIPLE GENE DELETIONS IN *PICHIA PASTORIS* WITH HIGH EFFICIENCY

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:

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:

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. The aim of this study was 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. After the successful gene knockout, the respective selection marker can be recycled. Excision of the marker is mediated by Flp recombinase and occurs at high frequency of ≥95%. As an application test of our vector system we reproduced the already described gene knockouts of *LYS2, MET2*, *TYR1*, *SUB*2, *PEP4*, *PRB1* and *PRC1*. Surprisingly, deletion of the protease encoding gene *PEP4* could only be achieved when using the *HIS4* gene for selection. 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 propose a fast pooling method for the parallel screening of multiple gene knockouts.

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. The only prerequisite for gene replacement by homologous recombination is that sequence information about the target locus must be accessible.

During the last years, the genome sequences of the important *P. pastoris* production 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 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 deletion is extremely low, with homologous targeting sequences of <500 bp leading to <1% of positive targeting events [4].

Transformed DNA fragments can get integrated into the genome by 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 [5]. On the other hand, NHEJ requires little to no sequence homology to operate [6]. The free DNA ends are first bound by the heterodimer Ku70/80, which in turn recruits the catalytic subunit of DNA protein kinase (DNA-PKcs) [5, 7]. HR is 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 [8]. This property of *S. cerevisiae*

allows construction of knockout cassettes by one-step PCR, which integrate with routinely 70% efficiency at the correct locus [9]. In order to foster HR in *P. pastoris*, a key player of NHEJ, Ku70p, was deleted by Näätsaari et al. [10]. 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 in literature. For the specific targeting of a locus, homologous flanking regions of ~1 kb are commonly used. Together with the selection marker, this requirement results in targeting cassettes of several thousand base pairs in length, which can be assembled either by cloning [11–15] or overlap-extension PCR (OE-PCR) [10, 16]. The published cloning methods usually require several sub-cloning steps and careful selection of appropriate restriction endonucleases, complicated by low restriction efficiencies and incompatibility of some enzymes. The main disadvantage of fusing long DNA fragments by PCR is obviously the risk to accumulate 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 is an important factor in gene targeting. Various auxotrophic and antibiotic resistance markers have already been described in *P. pastoris* [11, 13, 17–21]. Nevertheless, extensive genetic engineering projects, such as the manipulation of the yeast's glycosylation pathway [22], stress the need for marker recycling. Nett and co-workers [11, 18] 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 [21]. 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 puts 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* [11]. 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 [10] 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 for the targeting of genes 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 *SfiI* 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* [19], *TYR1* [13], *SUB2* [29], *PEP4*, *PRB1* [30] and *PRC1* [31]. To our knowledge, we describe for the first time the targeting of *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* [32], we argue that alternative biosynthesis routes for phenylalanine must exist in *P. pastoris*. Moreover, we introduce a fast and cleverly devised pooling method for the screening of 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. [33] 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. In between of these FRT repeats a stuffer fragment, bounded by two *SfiI* restriction sites (GGCCNNNN/NGGCC), was cloned to construct the knockout plasmid pPpKC1 (Figure 1A). The single-stranded overhangs generated by the *SfiI* 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 *SfiI 1* and *SfiI 2*. The special feature of *SfiI* restriction endonuclease, a type IIF restrictions enzyme, is that it interacts with two restriction sites simultaneously and cleaves them in a concerted manner, guaranteeing high restriction efficiencies [34]. We observed exceptionally high ligation efficiencies of more than 95% with *SfiI*-cut vectors and inserts (data not shown). To expand the versatility of the system, the Zeocin marker cassette of pPpKC1 was exchanged for the alternative *P. pastoris* markers *KanMX6, HIS4* and *ARG4*, in combination 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 *SfiI 1* and *SfiI 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. Furthermore, we examined selective concentrations of Zeocin (25, 50 and 100 µg/ml) and Geneticin (200 and 300 µg/ml). A concentration of 25 µg/ml Zeocin was found to be sufficient for identifying single copy transformants on YPD media. Higher concentrations of Zeocin 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 was needed to select positive transformants. When using the *KanMX6* selection marker, 300 µg/ml of Geneticin was found 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 four protease genes (*SUB2, PEP4, PRB1* and *PRC1*). To target the knockout cassettes to these loci, we amplified approximately 1000 bp of respective 5'- and 3'- regions from gDNA of wild-type CBS7435. Table 1 provides detailed information on the disrupted genes and the length of the amplified homology regions. During this PCR step the restriction sites *SfiI 1* and *SfiI 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. *SmaI*, between the fragments that could later be used for linearization of the vector. This unique restriction site was generated by choosing the binding position of the most 5' and most 3' primer, P1 and P4, on the genome sequence in a way that they reconstitute the recognition sequence for a blunt end restriction enzyme after fusion in the OE-PCR (Figure 2A). Following restriction with *SfiI*, 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 restriction site located between the homology regions. 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 to 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. However, Whittaker and Whittaker [13] 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 [35]. 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 - 47% 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 come 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 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 *SfiI* recognition sites. Marker recycling efficiencies for Flp-mediated recombination after 24 and 48 h of induction in 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 [19, 28]. 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 the 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 [32]. 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. For knockouts of the protease encoding genes *SUB2*, *PEP4*, *PRB1* and *PRC1* no clear knockout phenotype has been described. 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. Isolated gDNA is used as template to check for correct integration of the knockout cassettes by PCR as sketched in Figure 2C. We identified targeting events of *SUB2*, *PEP4*, *PRB1* and *PRC1*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 (Additional File 4) and sequencing.

Unexpectedly, we were not able to delete *PEP4* using a Zeocin^R-marker on the knockout cassette. Extensive screening of 400 transformants identified six clones with the cassette integrated at the target locus. However, all of the six transformants turned out to have the coding sequence of *PEP4* reintegration at another position in the genome (data not shown). We assume that gene targeting was complicated by the important role of Pep4 as a major vacuolar protease. Pep4 was described to activate itself as well as other proteases, such as proteinase B (Prb1) and carboxypeptidase Y (Prc1) [36]. Deletion of *PEP4* could therefore have a detrimental effect on cell viability. Additionally, the strong antibiotic Zeocin might put too much pressure on the weakened cells. Contrary to our observations, Pan et al. [16] reported the knockout of *PEP4* using Zeocin expressed from P_{TEF1} as a marker. This promoter is significantly stronger than the P_{ARG4} promoter we used for expression [37]. Lower expression levels of the resistance gene may explain why we did not succeed in obtaining *pep4* knockouts with Zeocin as a marker. In order to omit any

negative effect of Zeocin 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. Beside the single knockout strains *sub2*, *pep4*, *prb1* and *prc1*, we generated the protease double knockout *pep4prb1*.

Conclusions:

We trust that the method and tools presented here can contribute to the investigation of gene function in *P. pastoris* by making the creation of gene knockout strains faster and more efficient. 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 biosynthetic and four protease genes. The observed knockout efficiencies varied significantly (4 – 47%) 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* 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* [10] 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 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. *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 not able to grow in rich media [13, 35]. 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) [10] 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 *PciI and BglII* 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 (*PciI, BglII, and NcoI*). 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 *PvuII-AvrII*-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 *NdeI-PstI* 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 "*SfiI* 2" (5'-GGCCGATCAGGCC-3') and *"SfiI* 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. *SmaI,* 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 *SfiI* 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 *SfiI*. The insert was also sequenced by LGC Genomics (Berlin, Germany).

Pichia transformations

P. pastoris competent cells were prepared using the condensed protocol [38]. 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^{o}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^{o}C, 320 rpm. The cultures were pinned on 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 (*sub2, pep4, prb1* and *prc1*) were confirmed by PCR only. Genomic DNA of multiple clones was isolated in one step using our suggested 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 was used per PCR reaction.

Isolation of genomic DNA

P. pastoris gDNA was isolated using a modification of the protocol by Hoffman and Winston [39]. 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^{o}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 grams 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^{o}C for 30 min to increase the overall yield of gDNA. After incubation, tubes were centrifuged for 1 min to pellet gDNA and supernatant was discarded. The pellets were dried at 60^{o}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^{0}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

Genomic DNA – gDNA

Deep well plate – DWP

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

Homologous recombination – HR

Non-homologous end joining – NHEJ

Flippase - Flp

Flippase recombination target – FRT

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 and DK Molecular Enzymology 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. Higgins DR, Cregg J: *Pichia Protocols*. *Volume 103*. New Jersey: Humana Press; 1998.

5. Pastwa E, Błasiak J: **Non-homologous DNA end joining.** *Acta Biochim Pol* 2003, **50**:891–908.

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

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

8. 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.

9. 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.

10. 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**:e39720.

11. 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.

12. 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.

13. 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.

14. 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**:1241–1249.

15. 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.

16. Pan R, Zhang J, Shen W-L, Tao Z-Q, Li S-P, Yan X: **Sequential deletion of Pichia pastoris genes by a self-excisable cassette.** *FEMS Yeast Res* 2011, **11**:292–8.

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

18. 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.

19. Thor D, Xiong S, Orazem CC, Kwan A-C, 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.

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

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

22. 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.

23. 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.

24. Soderholm J, Bevis BJ, Glick BS: **Vector for pop-in/pop-out gene replacement in Pichia pastoris.** *Biotechniques* 2001, **31**:306–10, 312.

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

26. Güldener U, Heck S, 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.

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

28. 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.

29. 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.

30. 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.

31. 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.

32. 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.

33. 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**:856–865.

34. 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.

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

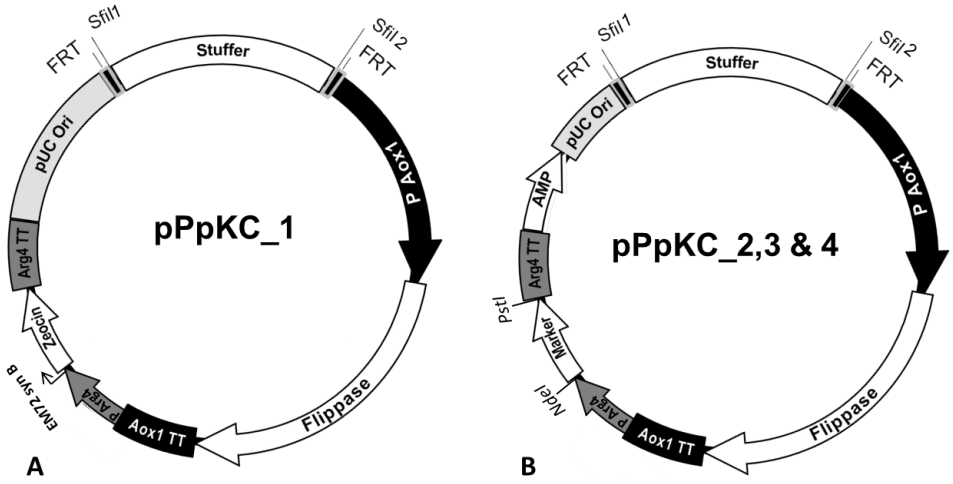
36. Rothman JH, Hunter CP, Valls L a, Stevens TH: **Overproduction-induced mislocalization of a yeast vacuolar protein allows isolation of its structural gene.** *Proc Natl Acad Sci U S A* 1986, **83**:3248–52.

37. 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.

38. 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.

39. 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 constructed during this study. (A) pPpKC1. (B) pPpKC2, 3 and 4. Indicated are the unique restriction sites *NdeI* and *PstI* to exchange the marker.

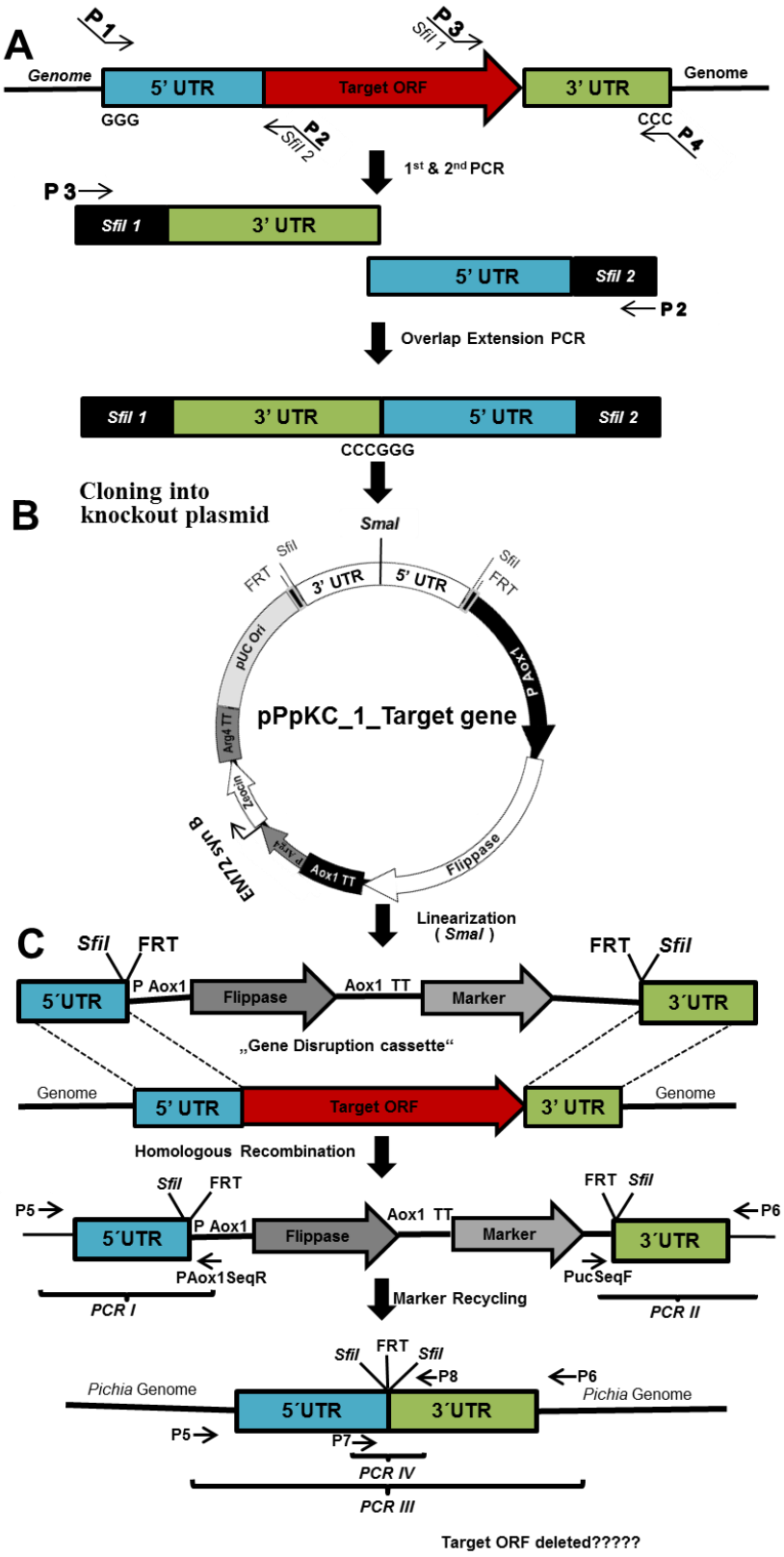

A
P1
P3
SfiI 1
Genome
5' UTR
Target ORF
3' UTR
Genome
GGG
CCC
P2
SfiI 2
P4
1st & 2nd PCR
P 3
SfiI 1
3' UTR
5' UTR
SfiI 2
P2
Overlap Extension PCR
SfiI 1
3' UTR
5' UTR
SfiI 2
CCCGGG
B
Cloning into knockout plasmid
SmaI
FRT
SfiI
3' UTR
5' UTR
SfiI
FRT
pUC Ori
P AOX1
pPpKC_1_Target gene
Arg4 TT
Zeocin
EM7L syn B
R Ampr
Aox1 TT
Flippase
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
SfiI
FRT
P5
5´UTR
P Aox1
Flippase
Aox1 TT
Marker
FRT
SfiI
3´UTR
P6
PAox1SeqR
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 the unique restriction site (*SmaI*) for subsequent linearization. (B) The *SfiI*-restricted 3´and 5´-homology regions were cloned into the knockout plasmid. The final knockout vector was linearized using *SmaI* 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 (P5, PAox1SeqR) and region II) containing the 3´-homology (PucSeqF, P6). Clones with positive results for both PCRs were selected for marker recycling. The removal of the integrated cassette was verified by amplification of region III) and region IV). P1 – P8: Primers; FRT: Flippase recombination target

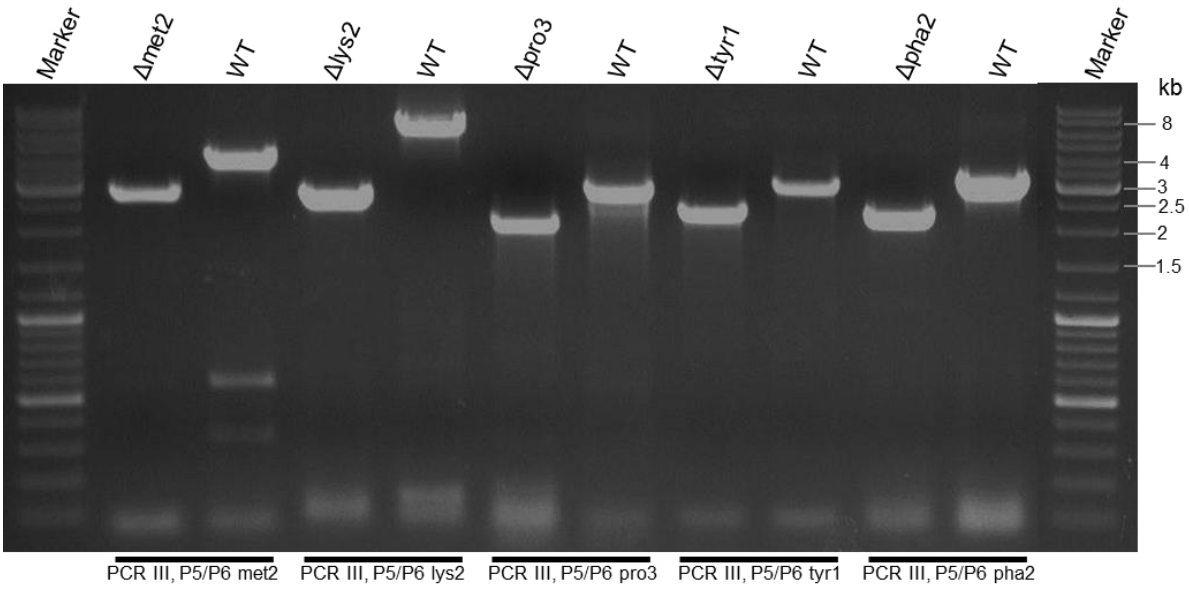


Figure 3: Verification of genomic loci by PCR analysis using knockout specific primer pairs (P5 and P6; PCR III). The results for wild type and gene knockout strains are shown. Marker: GeneRuler DNA Ladder Mix (Thermo Scientific); WT: wild type CBS7435

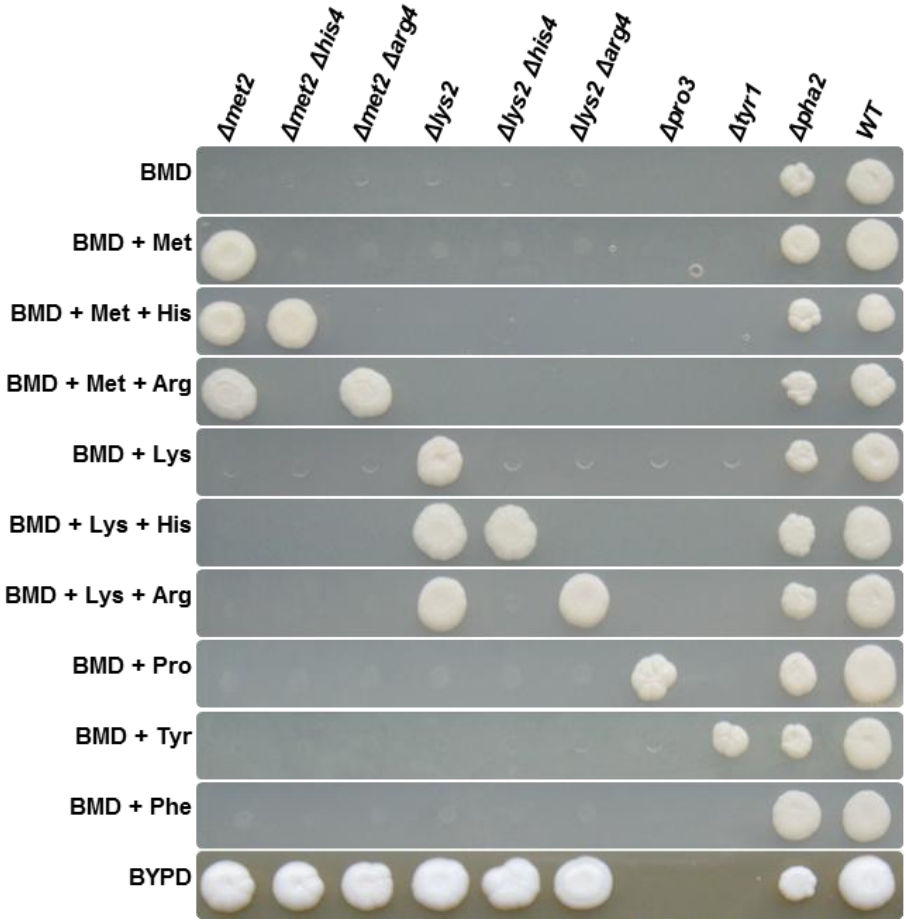


Figure 4: Growth behaviour of *P. pastoris* biosynthetic gene knockout strains. Upon cultivation in 96-well deep-well plates containing 250 µl BMD media supplemented with respective amino acids for 24 h at 28 ^{o}C, 320 rpm and 80% humidity, approximately equal number of cells (OD_{600} = 0.5) were pinned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

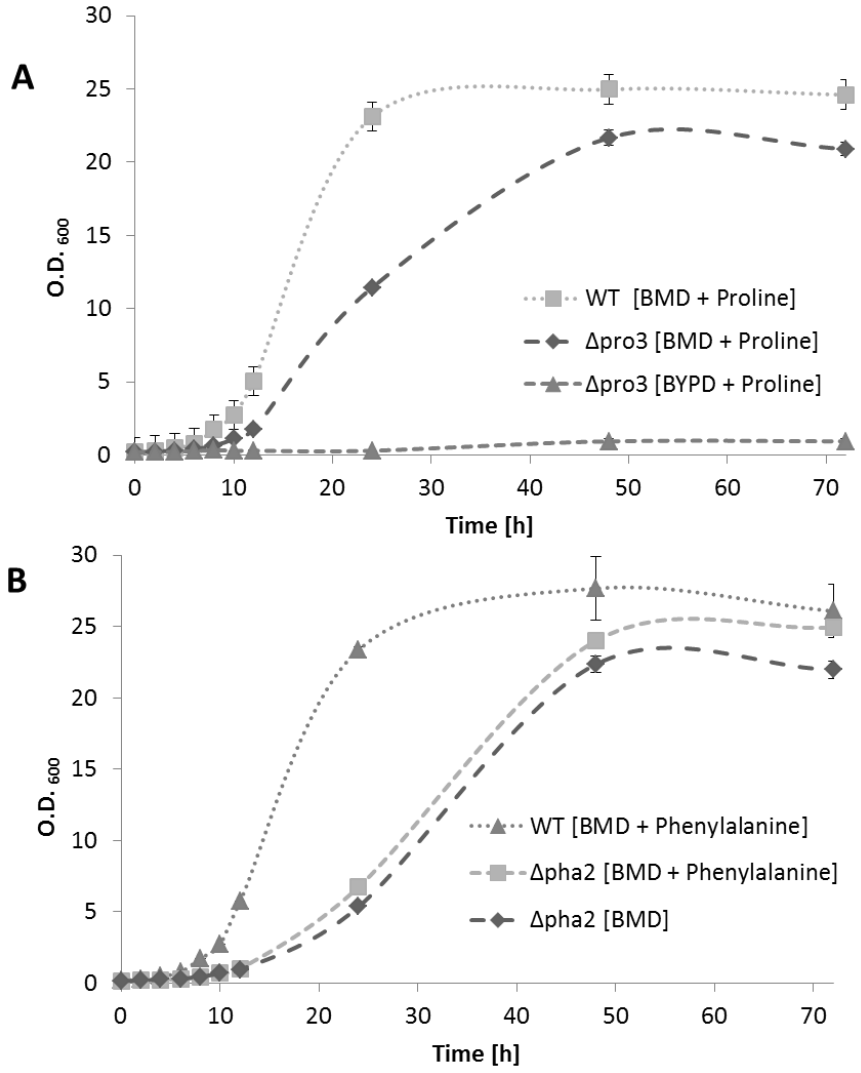


Figure 5: Growth rate analysis of *P. pastoris* wild type, *pro3* and *pha2* strains. The strains were cultivated in 300 ml baffled shake flasks at 28^{0}C and 120 rpm. Experiments were performed in triplicate.

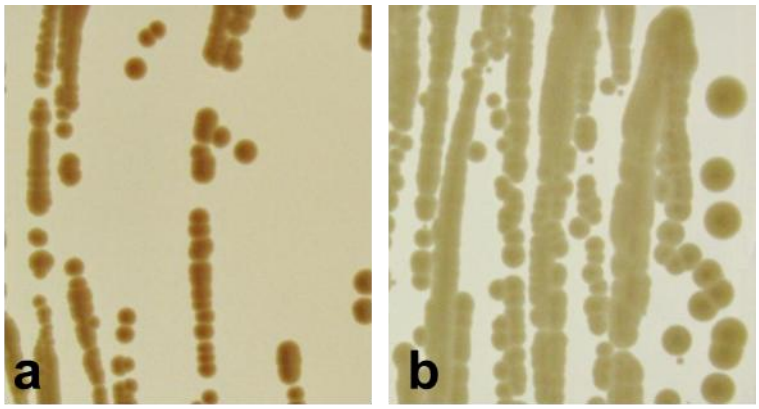


Figure 6: Growth phenotype of the *pha2* strain on (a) buffered minimal media, (b) buffered minimal media supplemented with phenylalanine (150 µg/ml) after ~10 days of incubation at 28^0C.

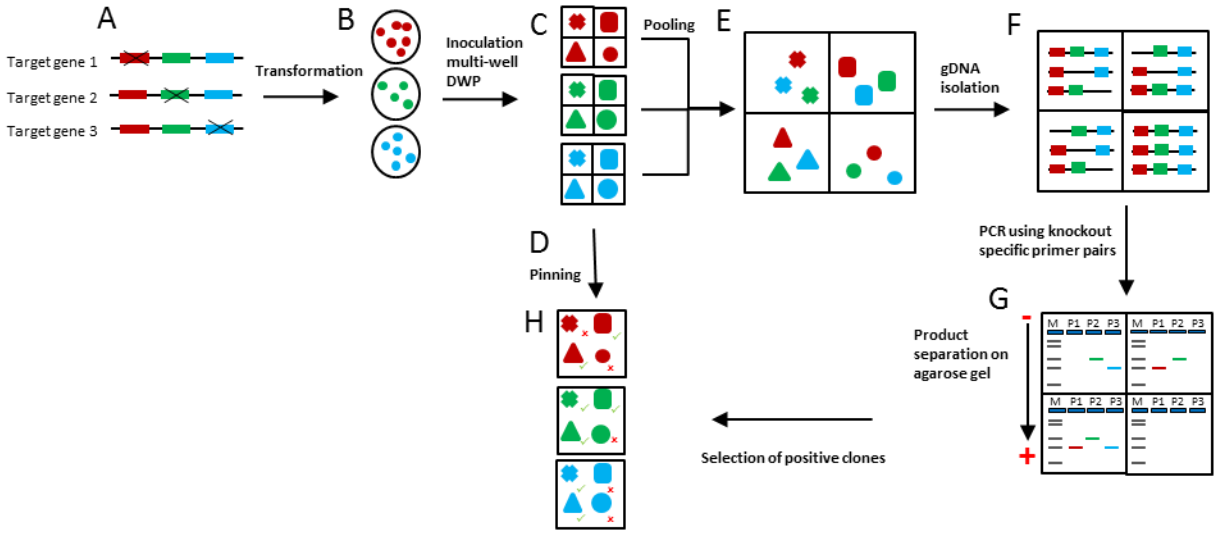


Figure 7: Strategy for fast identification of positive gene targeting events in multi-well format. (A) In separate DNA transformation reactions different genes are targeted with specific gene targeting cassettes. (B) After transformation, cells are plated on selective medium. (C) Single colonies are 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.

Tables:

Table 1: Knockout efficiencies for biosynthetic and protease genes

Target Loci	Protein ID CBS 7435	5'UTR/3'UTR [bp]	Selection marker	N° of clones screened	N° of positive clones	Efficiency	Efficiency [%]
Biosynthetic gene knockouts – phenotypic growth analysis on selective media							
Met2*	CCA40261.1	1249/1369	Zeocin^R	248	37	37/248	14.9
Lys2*	CCA37057.1	1495/1158	Zeocin^R	408	17	17/408	4.2
Pro3	CCA40748.1	951/957	Zeocin^R	264	89	89/264	33.7
Tyr1	CCA38031.1	1231/867	Zeocin^R	176	15	15/176	8.5
Pha2**	CCA40709.1	1033/1089	Zeocin^R	176	83	83/176	47.2
Protease gene knockouts – confirmation by PCR using gDNA isolated by pooling method							
Sub2	CCA37470.1	1014/1157	Zeocin^R	24	3	3/24	12.5
Pep4***	CCA39046.1	941/985	*HIS4*	24	5	5/24	20.8
Prb1	CCA36690.1	1039/1136	Zeocin^R	48	6	6/48	12.5
Prc1	CCA36928.1	1127/979	Zeocin^R	24	1	1/24	4.2

* The targeting efficiencies for *MET2* and *LYS2* loci were averaged from transformations into different strain backgrounds.

** 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.

*** Knockout of *PEP4* was not successful with Zeocin marker

Table 2: Strains used and constructed during this study

Strain	Genotype	Knockout vector used	Reference
CBS7435	WT		[10]
Pp3520	CBS7435 *his4::FRT*		[10]
Pp3521	CBS7435 *arg4::FRT*		[10]
Pp7030	CBS7435 *met2::FRT*	pPpKC1_met2	This study
Pp7031	Pp3521 *met2::SfiI1 FRT SfiI2*	pPpKC1_met2	This study
Pp7032	Pp3520 *met2::SfiI1 FRT SfiI2*	pPpKC1_met2	This study
Pp7033	CBS7435 *lys2::SfiI1 FRT SfiI2*	pPpKC1_lys2	This study
Pp7034	Pp3521 *lys2::SfiI1 FRT SfiI2*	pPpKC1_lys2	This study
Pp7035	Pp3520 *lys2::SfiI1 FRT SfiI2*	pPpKC1_lys2	This study
Pp7036	CBS7435 *pro3::SfiI1 FRT SfiI2*	pPpKC1_pro3	This study
Pp7037	CBS7435 *tyr1::SfiI1 FRT SfiI2*	pPpKC1_tyr1	This study
Pp7029	CBS7435 *pha2::SfiI1 FRT SfiI2*	pPpKC1_pha2	This study
Pp6668	CBS7435 *sub2::SfiI1 FRT SfiI2*	pPpKC1_sub2	This study
Pp6911	Pp3520 *pep4::SfiI1 FRT SfiI2*	pPpKC3_pep4	This study
Pp6912	CBS7435 *prb1::SfiI1 FRT SfiI2*	pPpKC1_prb1	This study
Pp6676	CBS7435 prc1::*SfiI1FRT SfiI2*	pPpKC1_prc1	This study
Pp7013	Pp6911 prb1::*SfiI1 FRT SfiI2*	pPpKC3_prb1	This study

Supporting information

Additional File 1: Primers used in the current study.

Primer List	5'-3'	Purpose
PciIFRTSfiI1F	TCACATGTGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGCCGATCAGGCCCAACTTACTCCAACCTTCTACGA	Primers used to construct the basic knockout plasmids
BglIIFRTSfiI2R	TTAGATCTGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCGGCCCTAGTGGCCGAGTTAGAGTTGACAACACGGCAC	
P(AOX1)forw	AAGGTACCAGATCTAACATCCAAAGACGAAAG	
PAox1R	CGTTTCGAATAATTAGTTGTTTTTTGATCTTC	
FLPF	TTATTCGAAACGATGCCACAATTTGATATATTATG	
FLPR	TTATATGCGTCTATTTATGTAGGATGAAAGG	
Aox1TTF	GACGCATATAAGTTTTAGCCTTAGACATGACTG	
Aox1TTR	CGTTCCGTTCCGCACAAACGAAGGTCTCAC	
PAGR4F	CTTCGTTTGTGCGGAACGGAACGTATCTTAG	
PARG4R	GTAACAACACTAGCTGGTAATAAGTTTAGAAC	
EM72F	CTTATTACCAGCTAGTGTTGTTACTTTATACTTCCG	
ZeoR	CAAACTCAGTATATTAGTCCTGCTCTTCTGCGAC	
Arg4TTF	GAGCAGGACTAATATACTGAGTTTGTTAATGATA	
Arg4TTKpnIR	GTGGTACCAATGCGAGGATGCTGCTGGAGAC	
PucOriPciIR	ACATGTGAGCAAAAGGCCAGCAAAAGG	
PucOriKpnIF	CCAGCAGCATCCTCGCATTGGTACCACTGAGCGTCAGAC	
3UTRMET2F	TCGGCCGATCAGGCCGCTACAAAGTAAAAATTTGCCTCACGC	to create CBS 7435 Δmet2 Gene ID: FR839631 REGION: 159701..161128
3UTRMET2R	GGACTTATGGTAGTTGGATTTAAATTCTAGTTGGGCTTGTGTACCTTTG	
5UTRMET2F	GCCCAACTAGAATTTAAATCCAACTACCATAAGTCCTAGCTC	
5UTRMET2R	TCGGCCCTAGTGGCCAATTCAATATGTTGAGGCG	
Up5UTRMET2F	CCACCAGTAACTTTCCCTATATACTCAGC	
Down3UTRMET2R	GATACGTAGTCTGCTCTTTGCTTTCTG	
3'UTRLys2F	TCGGCCGATCAGGCCTTCTATATGTAAGTGATATTAAAC	
3'UTRLys2R	CGCCCAAGTTGAAATATTCAGAAGTACGGGGTAGAAGGCC	
5'UTRLys2F	TGAATATTTCAACTTGGGCGTCAGC	
5'UTRLys2R	TCGGCCCTAGTGGCCGGTTCGTTCTTAAGAGTGCG	
5'Lys2OutsideF	GAGAAGAAGAGGAAACTGCC	
3'Lys2OutsideR	CACATGTGGACATACTCCCATG	
3UTRPRO3F	TTTCTAGAGAATAGGAACTTCGGTAAATTCACTGACTGCCTCTTTCTTTC	for PRO3 Knockout Gene ID: FR839631 REGION: 991404..992228
3UTRPRO3R	CTTGGAATGGTTAATTTAAATATTTTCAGTATCAAACCCGTTGAACTTG	
5UTRPRO3F	GATACTGAAAATATTTAAATTAACCATTCCAAGAAATGCATCTTTCCG	
5UTRPRO3R	TCTCTAGAAAGTATAGGAACTTCCTTTATTAGTGATTAAGCAAACTAAAGTGGGAG	
Up5UTRPRO3F	GGCTTGTCAGAATGTTCAGCTTCGGC	
Down3UTRPRO3R	GTAAGGTCCGCTGGGTCCATAAAACTGTC	
N3UTRTYR1F	TCGGCCGATCAGGCCCATACTTCAGCAATTTTCATTGAGCAAGG	to create CBS 7435 Δtyr1 Gene ID: FR839629 REGION: 613763..615058
N3UTRTYR1R	CTGCTGCATTTAAATAACAGCCGCTGTGTCCG	
N5UTRTYR1F	GCGGCTGTTATTTAAATGCAGCAGATCAGTATAGTTTGAACTTG	
N5UTRTYR1R	TCGGCCCTAGTGGCCTCTCTTCTCTATCACATCCGATCACC	
Up5UTRTYR1F	GATCACGTTCAAGAGAGGTTTGGATTCC	
NDown3UTRTYR1R	CAGCATTGTATGTAGTTCATCCCTAGC	
3UTRPHA2F	TCGGCCGATCAGGCCAAAAGGGTTAAGTGTAAGATGTAAATATATTAATTTCG	for PHA2 Knockout Gene ID: FR839631 REGION: 929530..930480
3UTRPHA2R	CTGCAATGGCTGGATATTTAAATGATCGATATGACTCCCCTTCTGG	
5UTRPHA2F	GTCATATCGATCATTTAAATATCCAGCCATTGCAGTTTGGATTC	
5UTRPHA2R	TCGGCCCTAGTGGCCGGTTAGGTTATCCTATATGGGGGAACG	
Up5UTRPHA2F	GTCGTTCATGAAAGACCTGCGC	
Down3UTRPHA2R	GTTGAATTCCAGAAGCCTTGAGATCTATG	
3UTRsub2F	TCGGCCGATCAGGCCTGACTCATTGACCCCAGCTCAAC	For Sub2 knockout cassette construction and verification of integration at correct locus in combination with Paox1SeqR and PucSeqF
3UTRsub2R	GGGACTGACCCGGGTGAGGAAAACACTCATTGAAATTCCTG	
5UTRsub2F	CCTCACCCGGGTCAGTCCCAACTTGTTGG	
5UTRsub2R	TCGGCCCTAGTGGCCGATCCCTGTAATTTCAGCGATGGAG	
Up5UTRsub2F	ACGATTAAGGCAAATCTTCCGGTTC	
Down3UTRsub2R	GAAACAAATCAGTGACGGCGATGTC	
3UTRapr1F	TCGGCCGATCAGGCCCTCAGTTTATGACCTAGGCAAAGATGC	For pep4 knockout cassette construction and verification of integration at correct locus in combination with Paox1SeqR and PucSeqF
3UTRapr1R	GATAAAGGTCCCCGGGACCTCGGTTGTAAGCGGTAATTC	
5UTRapr1F	CCGAGGTCCCGGGGACCTTTATCACGTTGAATCTAGTTG	
5UTRapr1R	TCGGCCCTAGTGGCCGCTTGTGTATCTTAGCAGAATGAACTTTGG	
Up5UTRapr1F	GAAAATAGTGTATCACTGCCAGCATC	
Down3UTRapr1R	CTCATCTATACCCCAGGACCAG	

3UTRprb2F	TCGGCCGATCAGGCCCACTGTCACCATTAGCACCAAACTG	For Prb1 knockout cassette construction and verification of integration at correct locus in combination with Paox1SeqR and PucSeqF
3UTRprb2R	GCCTCTAATCCCGGGAAAGTTTAACTTCATACAGAATAACTTCATG	
5UTRprb2F	AAACTTTCCCGGGATTAGAGGCGGTTGAACTCTG	
5UTRprb2R	TCGGCCCTAGTGGCCGTTGCTTCCTCCGACGATACTG	
Up5UTRprb2F	GCAGTATCCTGCTCATCTTCCCGTAC	
Down3UTRprb2R	CATGAACGTGTTGAACTTGGACGCC	
3UTRkpx16F	TCGGCCGATCAGGCCGGCATCTGCAAGGACAGACC	For Prc1 knockout cassette construction and verification of integration at correct locus in combination with Paox1SeqR and PucSeqF
3UTRkpx16R	CACCTATCCCGGGAAAAGGCACATAAAGCAATCAATC	
5UTRkpx16F	GCCTTTTCCCGGGATAGGTGATCCCTCAAAGAAGG	
5UTRkpx16R	TCGGCCCTAGTGGCCGGCCCCATATGATCAGCCAG	
Up5UTRkpx16F	CAAGTTCAAATGGCTCCATGGAGC	
Down3UTRkpx16R	GCATTGAGGAAGTACATGGTCACG	
nCDSsub2F	GCACACTCGCTTTTGATACCATCTC	To rule out reintegration of excised CDS of Sub2
nCDSsub2R	ATCCGAGTCATCAAGTACATCCTTGG	
nCDSpep4F	CTCTCTACTCTAGGTATTGGTGCTGAAG	To rule out reintegration of excised CDS of Pep4
nCDSpep4R	ACCTACTGCATCTTTGCCTAGGTC	
nCDSprb1F	AAACTCTTGGGCCAAGTTTTCAACAG	To rule out reintegration of excised CDS of Prb1
nCDSprb1R	GATTGGCTATCTTATCTGCCATAGCAG	
nCDSprc1F	ATGAGAATTCTCTGGCTGATCATATGGG	To rule out reintegration of excised CDS of Prc1
nCDSprc1R	TCCTAAAGCTATTGGTCTGTCCTTGC	
PAox1SeqR	GGTTTCATTCAACCTTTCGTCTTTGGATG	For confirmation of cloning of homologous recombination sequences into the knockout plasmid to construct gene disruption cassettes and confirmation of site of integration in pichia pastoris in combination with P5 and P6 Figure
PucSeqF	CTTTTTACGGTTCCTGGCCTTTTGC	

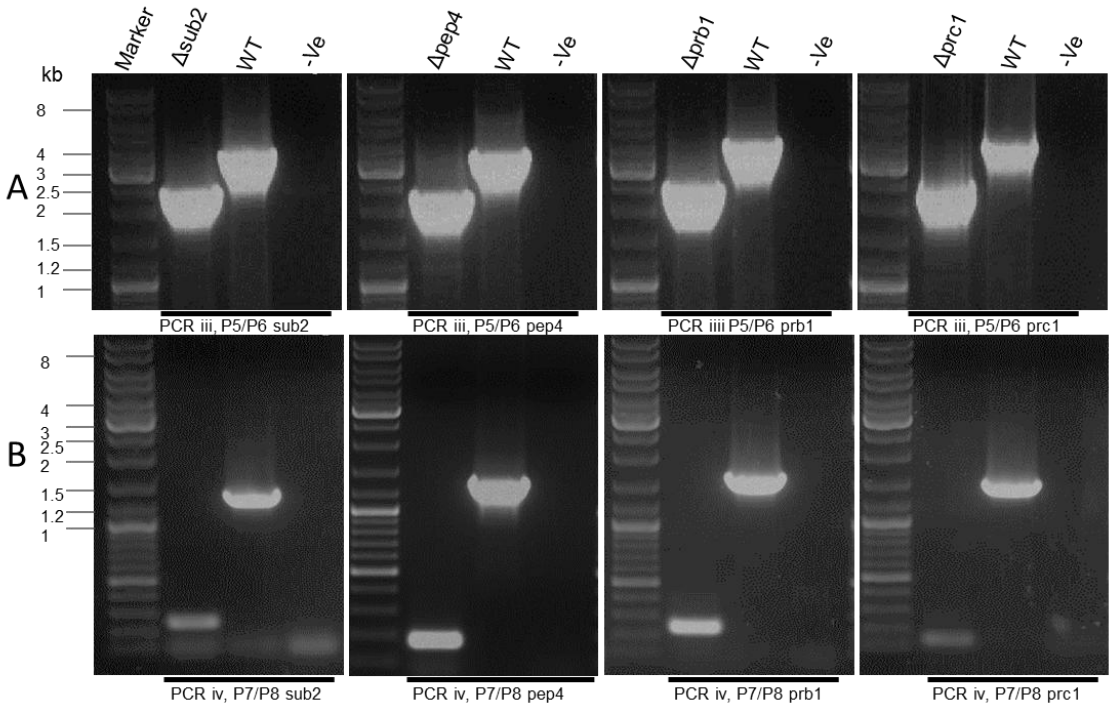
Additional File 2: Elements of *E. coli*/*P. pastoris* gene knockout shuttle vectors and their function.

Elements	Origin	Function
P_Aox1	*Pichia pastoris* CBS7435	*P. pastoris AOX1* promoter for expression of Flippase
Flippase[a]	*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

(a) Three restriction sites, namely *SwaI*, *EcoRI* and *NdeI* were mutated by creating silent mutations with overlap extension PCR.

Additional file 3: Plasmid sequences of the constructs used in this study in Genbank format. *(not included in this thesis)*

Additional File 4: PCR based characterization of genomic loci for protease deficient strains using specific primers pairs.



Verification of genomic loci for protease deficient strains using specific primers pairs. Marker, GeneRuler DNA Ladder Mix (SM0331,Thermo Scientific). WT, *P. pastoris* CBS7435 wild type. –Ve, negative control for PCR. A, Analysis of genomic loci for excision of target gene coding sequence with outer primers. B, PCR IV, using primers binding inside the target locus to rule out reintegration events (for details see figure 2C).

CHAPTER 4

Report about research stay

Report about research stay at the Department of Molecular and Cell Biology, University of California at Berkeley, California, USA

July 2012 – January 2013

Topic: Investigation of post-Golgi protein trafficking pathways in yeast

PI abroad: Randy Schekman

During my work in Graz that focuses on the secretory pathway in the yeast *Pichia pastoris* I discovered a set of genes, which possibly have an influence on the secretion efficiency of recombinantly expressed proteins in this yeast. The aim of my research was to acquire new techniques to further study the role of these genes in the secretion of cargo proteins. The Schekman lab has profound knowledge in studying the secretory pathway in both yeast and mammalian cells. The *sec* proteins, a compilation of proteins involved in various steps of secretion, were first identified and described in Randy Schekman's lab in the late 1970s and 1980s (reviewed in Schekman 2010).

At my arrival in June 2012, Prof. Schekman suggested that I join one of the lab's ongoing projects, to ease the process of adaption to the new working environment. I chose to team up with an experienced Postdoc in the lab, Zhiliang Cheng, who was working on post-Golgi trafficking in *Saccharomyces cerevisiae*. There is rather sparse knowledge about trafficking pathways from the trans-Golgi network (TGN), since proteins can enter many different routes to reach their destined compartment. The large diversity, in conjunction with low abundance, of vesicles emerging from the TGN, makes this field a challenging one to study. I reckoned that the methods I would learn studying post-Golgi trafficking pathways in *Saccharomyces* could also be applied to investigate trafficking of recombinant protein in *P. pastoris*. Using *Pichia* as host organism for these experiments proved to be difficult to start with, due to the lack of temperature-sensitive mutants in this organism, and the absence of appropriate cultivation conditions in the lab, i.e. baffled flasks, 28°C incubators.

Earlier work in the Schekman lab had focused on post-Golgi vesicular trafficking of two proteins: Chs3p and Fus1p, two membrane proteins localized to the plasma membrane (Ziman et al. 1996;

Valdivia et al. 2002; Valdivia et al. 2003; Sanchatjate and Schekman 2006; Barfield et al. 2009). Both proteins can also be found in intracellular compartments of the cell that colocalize with markers of the TGN and early endosomes (Chuang and Schekman 1996; Santos and Snyder 1997; Ziman et al. 1996). Chs3p and Fus1p cycle between these compartments and the plasma membrane via vesicular carriers. Transport to the plasma membrane had been shown to be dependent on exomer, a multi-protein complex (Wang et al. 2006; Barfield et al. 2009). If the exomer complex is disrupted, Chs3p and Fus1p do no longer reach the plasma membrane, but reside in the internal stores instead. The functional role of exomer in vesicular trafficking is still uncertain. It was first presumed to act as a coat-complex, but experimental results could not confirm these assumptions (Matsuoka et al. 1998; Wang et al. 2006). Exomer is not sufficient to form coated vesicles in budding assays, as would be expected for a complete coat complex. This poses the question what additional proteins are involved in coat formation.

Chitin synthase III (Chs3p) is responsible for the synthesis of chitin at the lateral cell wall and the bud neck. Its localization and activity are tightly regulated throughout the cell cycle. Chs3p is maintained within internal reservoirs at steady-state and rapidly shifted to the plasma membrane when needed, e.g. during bud emergence or cell stress. Redistribution to the plasma membrane is dependent on exomer. Thus, in cells deficient for exomer (*chs5Δ*) Chs3p accumulates in internal compartments. It cycles between these internal compartments, TGN and early endosomes, in clathrin-coated vesicles. Incorporation into these vesicles is dependent on clathrin adaptor complex 1 (AP-1). The beta-subunit of AP-1 is encoded by the gene *APL2*. In *chs5Δ apl2Δ* mutant cells, where both the exomer-dependent transport to the plasma membrane and AP-1-dependent recycling are disrupted, Chs3p reaches the plasma membrane again, supposedly via an alternative bypass pathway.

In my experiments, I worked with *S. cerevisiae* strains that combined the mutations in exomer and AP-1 with a temperature-sensitive mutation of *SEC6*. Sec6p is essential for the fusion of post-Golgi vesicles with the plasma membrane, and as a result temperature-sensitive *sec6-4* mutants accumulate a large amount of transport vesicles when shifted to the restrictive temperature (37°C). The work of Valdivia et al. (2002) suggested that Golgi and post-Golgi vesicles could be sufficiently separated from each other on a sucrose density gradient. This would allow me to isolate pure fractions of these Chs3p-containing compartments for subsequent pull-down

experiments and budding assays, with the ultimate goal to identify unknown proteins involved in vesicle formation.

To test the separation efficiency of density gradients, I compared three *S. cerevisiae* strain backgrounds with different localization patterns of Chs3p: *sec6-4*, *sec6-4 chs5Δ* and *sec6-4 chs5Δ apl2Δ*. The accumulation of post-Golgi vesicles was induced by shifting the cultures to the restrictive temperature for 40 min. In the next step, spheroplasts were formed, osmotically lysed and the large membranes of plasma membrane and ER removed by differential centrifugation. Then, slowly sedimenting membranes were loaded on a sucrose step gradient and centrifuged to equilibrium. Fractions were collected from the bottom and the distribution of proteins was determined by immunoblotting. I could observe a shift of the Chs3p peak to lower density fractions in *sec6-4 chs5Δ* cells, where Chs3p should reside in TGN and early endosomes, compared to *sec6-4* and *sec6-4 chs5Δ apl2Δ* cells that accumulate Chs3p-containing vesicles. In the *sec6-4* mutant, Chs3p colocalizes with Pma1p, a marker for plasma membrane-bound secretory vesicles. The triple mutant, *sec6-4 chs5Δ apl2Δ,* showed a distribution of Chs3p that was somewhat between the first two strains. We reasoned that this could be due to a lower efficiency of the alternative transport route.

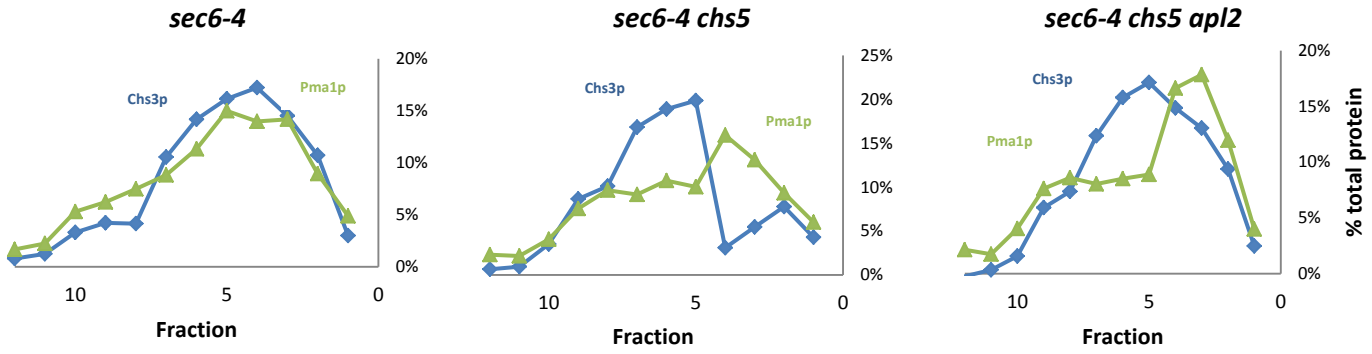


Figure 1: Chs3p distribution in different mutant strain backgrounds. Fractionation of post-Golgi secretory vesicles by density centrifugation. *sec6-4*, *sec6-4 chs5Δ* and *sec6-4 chs5Δ apl2Δ* cells were incubated at 37°C (40 min), converted to spheroplasts, osmotically lysed and fractionated by differential centrifugation (13,000 x g). S13 fractions were resolved on a sucrose step gradient and fractions collected from the bottom. Proteins present in the different fractions were identified by immunoblotting (Pma1p: secretory vesicle marker). The Chs3p peak shifts to lower

density fractions in the *sec6-4 chs5Δ* strain background, but cannot be fully separated from secretory vesicles (Pma1p).

Further modifications of the experimental set-up, e.g. using different incubation times at the restrictive temperature, modified sucrose step gradients or sequential sucrose gradients, did not lead to the desired clear separation of peaks for Chs3p and Pma1p in the *sec6-4 chs5Δ* strain background. The project was therefore abandoned for studying the traffic of another exomer-dependent cargo protein, Fus1p.

The membrane protein Fus1p is required for cell fusion in *S. cerevisiae*. Its expression is induced in the presence of mating pheromone of the opposite mating type. In about 85% of pheromone-induced cells Fus1p localizes to the mating projection, the so called "shmoo" (Trueheart et al. 1987). Besides its plasma membrane localization, Fus1p is also found in internal compartments that show the properties of TGN and early endosome, similar to Chs3p. In the absence of exomer, as in *chs5Δ* cells, the majority of Fus1p-GFP is detained in intracellular punctae (Santos and Snyder 2003). In *chs5Δ apl2Δ* double mutants, Fus1p can reach the plasma membrane again, probably via an alternate route of transport.

Previous work in the Schekman lab has identified the sorting signal in the cytosolic tail of Fus1p that is recognized by the exomer complex. Deletion or mutation of the signal leads to intracellular accumulation of Fus1p in wild-type cells. Clathrin-dependent transport between internal compartments is not affected by abolishing this signal, implying that AP-1 binds a different sequence stretch (Barfield et al. 2009). I screened a library of Fus1p-deletion constructs, generated by Robyn Barfield, for impaired AP-1 recognition. In this case, Fus1p-GFP should reappear at the plasma membrane in a *chs5Δ* strain. Indeed, I observed shmoo tip localization of Fus1p $^{\Delta 301\text{-}350}$-GFP and Fus1p $^{\Delta 350\text{-}400}$-GFP (Figure 2, c and d). The number of cells that showed this phenotype was not quantified, but seemed to be significantly lower than the 85% shmoo tip localization observed in the wild-type. It is probable that the exomer-independent transport route is not as efficient as the preferred, exomer-dependent route, and that this accounts for the reduced rate of Fus1p at the plasma membrane.

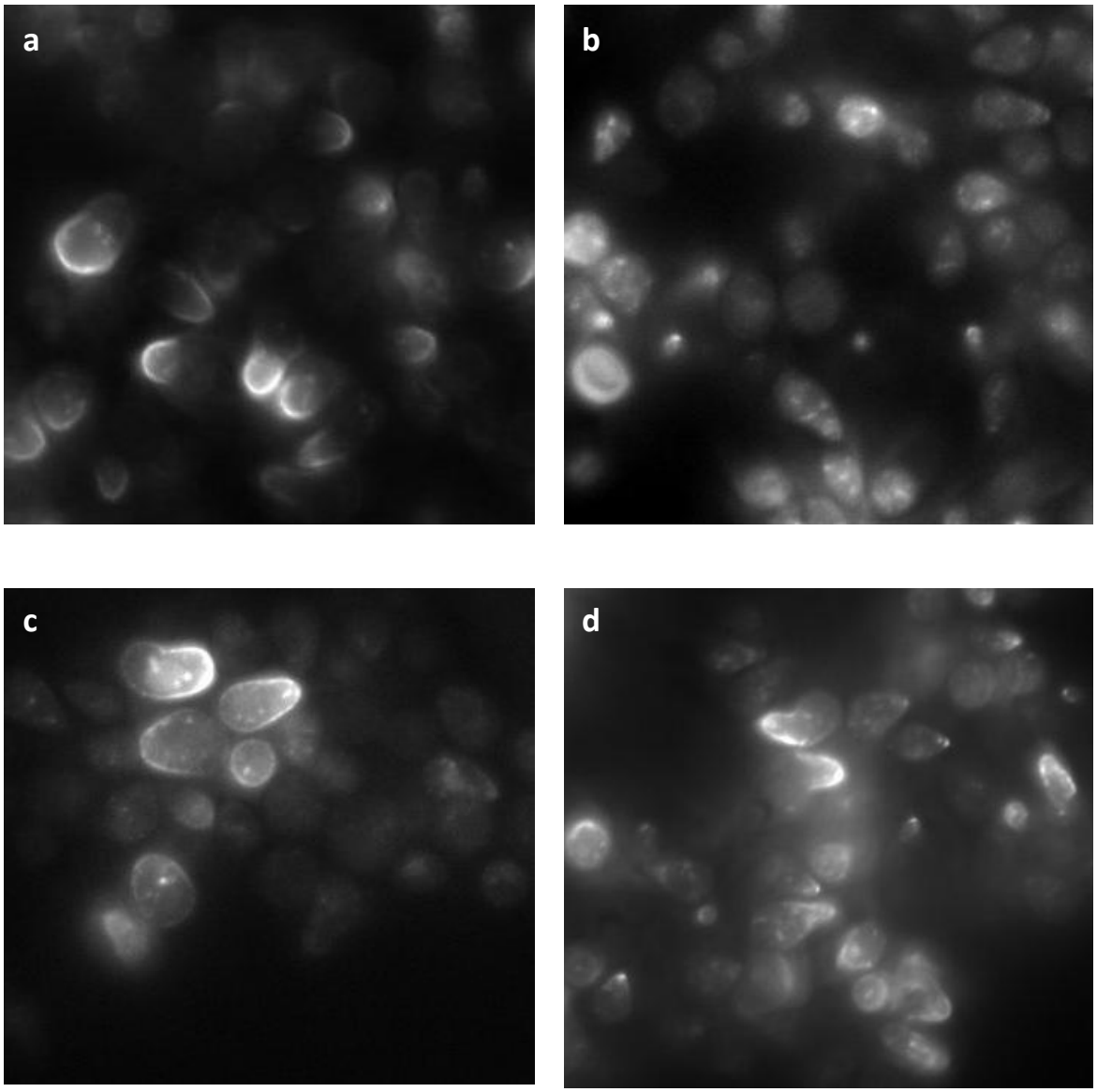


Figure 2: Fus1p-GFP localization in α-factor induced *S. cerevisiae* MATa *bar1Δ* cells. (a) Full-length Fus1p-GFP in wild-type cells localizes to the shmoo tip; (b) full-length Fus1p-GFP in *chs5Δ* cells is found in intracellular punctae; (c) Fus1p $^{\Delta 301\text{-}350}$-GFP in *chs5Δ* cells localizes to the shmoo tip in a certain percentage of cells; (d) Fus1p $^{\Delta 351\text{-}400}$-GFP in *chs5Δ* cells behave as in (c). Cells were treated with α-factor for 90 min, fixed with 4% para-formaldehyde and visualized by fluorescence microscopy.

Sorting signals recognized by AP-1 have already been described earlier. Known signals are the tyrosine-based YXXΦ and dileucine-based [DE]XXXL[LI] motives, reviewed in Bonifacino and Traub (2003). I found several variations of these motives in the 100 amino acid stretch, which when deleted, led to the expected phenotype in GFP-microscopy experiments. To further narrow down the sequence required for recognition by AP-1, I generated eight smaller deletions within amino acid 301 to 400 of Fus1p. Subsequent GFP-localization experiments did not lead to the expected unambiguous result. Instead, several of the smaller deletion constructs repeatedly showed both internal punctuate and plasma membrane localization of Fus1p, similar to the constructs with larger deletions.

The Yeast Two-Hybrid screen can be used to identify interactions on a molecular level. In this project, I used it to analyze interactions between the µ1-subunits of AP-1, which is the subunit of the complex that actually interacts with the substrate, and various deletion constructs of Fus1p. Upon interaction, growth of yeast cells is facilitated on selective medium. However, no growth was observed in my experiment, even though the importance of AP-1 on the trafficking of Fus1p had been confirmed through *in vivo* work before. These data indicate that the interaction between the two partners is either too weak to be recognized by the Yeast Two-Hybrid assay, or that additional proteins (or the complete adaptor complex) are necessary for the recognition process. These additional factors are absent in the synthetic Yeast Two-Hybrid system.

To overcome the obstacles of a completely artificial *in vitro* system, I tried to reproduce the binding of Fus1p to AP-1 in a pull-down experiment. Deletion constructs of Fus1p were expressed in *E. coli* as a fusion to the maltose binding domain (MBD). Upon binding of the fusion protein to amylose beads, the suspension was incubated with yeast cytoplasm. Immunoblotting of the bound protein fractions confirmed that AP-1 interacts with Fus1p, though from the result no reliable conclusions could be drawn regarding the specific interaction sequence.

Analyzing the distribution of Fus1p-GFP by fluorescence microscopy proved to be difficult. For each deletion construct of Fus1p, a certain percentage of cells showed accumulation of Fus1p in internal punctuate structures, others localized it solely to the plasma membrane, and a considerable number of cells had a mixed phenotype. I decided to quantify the distribution of Fus1p biochemically and conducted a protease protection assay as described in Barfield et at. (2009). With this assay it should be possible to quantify the amount of Fus1p localized to the plasma

membrane by determining the amount of Fus1p accessible to protease. After first attempts to reproduce Robyn Barfield's data, Prof. Schekman reasoned that I should try another method as the protocol had proven to be tricky and unreliable. We considered the possibility of doing a differential centrifugation to separate internal membranes and plasma membrane. To get a better separation, spheroplasts were treated with Concanavalin A, a lectin that specifically binds sugars and glycoproteins and increases the weight of the plasma membrane. Supernatant and pellet fraction were separated on a SDS-PAGE gel and the distribution of Fus1p-GFP was quantified by immunoblotting with a GFP-specific antibody. Surprisingly, only one band approximately of the size of GFP alone could be detected, which might be the product of degradation or unspecific binding. Further experiments with an alternative protein tag would be needed to validate this method for quantification of Fus1p distribution.

Barfield, R. M., Fromme, J. C., & Schekman, R. (2009). The Exomer Coat Complex Transports Fus1p to the Plasma Membrane via a Novel Plasma Membrane Sorting Signal in Yeast. *Molecular biology of the cell*, *20*, 4985–4996.

Bonifacino, J. S., & Traub, L. M. (2003). Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annual review of biochemistry*, *72*, 395–447.

Chuang, J. S., and Schekman, R. W. (1996). Differential trafficking and timed localization of two chitin synthase proteins, Chs2p and Chs3p. *J. Cell Biol. 135*, 597–610.

Matsuoka, K., Orci, L., Amherdt, M., Bednarek, S. Y., Hamamoto, S., Schekman, R., and Yeung, T. (1998). COPII-coated vesicle formation reconstituted with purified coat proteins and chemically defined liposomes. *Cell 93*, 263– 275.

Sanchatjate, S., & Schekman, R. (2006). Chs5 / 6 Complex: A Multiprotein Complex That Interacts with and Conveys Chitin Synthase III from the Trans-Golgi Network to the Cell Surface, *Molecular biology of the cell , 17*, 4157–4166.

Santos, B., and Snyder, M. (1997). Targeting of chitin synthase 3 to polarized growth sites in yeast requires Chs5p and Myo2p. *J. Cell Biol. 136*, 95–110.

Santos, B., & Snyder, M. (2003). Specific Protein Targeting during Cell Differentiation : Polarized Localization of Fus1p during Mating Depends on Chs5p in *Saccharomyces cerevisiae*, *2*(4), 821–825.

Schekman, R. (2010). Charting the secretory pathway in a simple eukaryote. *Molecular biology of the cell*, *21*(22), 3781–4.

Trueheart, J., Boeke, J. D., & Fink, G. R. (1987). Two genes required for cell fusion during yeast conjugation: evidence for a pheromone-induced surface protein. *Molecular and cellular biology*, *7*(7), 2316–28.

Valdivia, R. H., Baggott, D., Chuang, J. S., & Schekman, R. W. (2002). The yeast clathrin adaptor protein complex 1 is required for the efficient retention of a subset of late Golgi membrane proteins. *Developmental cell*, *2*(3), 283–94.

Valdivia, R. H., & Schekman, R. (2003). The yeasts Rho1p and Pkc1p regulate the transport of chitin synthase III (Chs3p) from internal stores to the plasma membrane. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(18), 10287–92.

Wang, C.-W., Hamamoto, S., Orci, L., & Schekman, R. (2006). Exomer: A coat complex for transport of select membrane proteins from the trans-Golgi network to the plasma membrane in yeast. *The Journal of cell biology*, *174*(7), 973–83.

Ziman, M., Chuang, J. S., Tsung, M., Hamamoto, S., & Schekman, R. (1998). Chs6p-dependent anterograde transport of Chs3p from the chitosome to the plasma membrane in *Saccharomyces cerevisiae*. *Molecular biology of the cell*, *9*(6), 1565–76.

CHAPTER 5

Supplementary data

Supplementary data

1 Transposon mutagenesis approach

Hermes transposon from *Musca domestica*

(Literature reference: Park JM, Evertts AG, Levin HL (2009) The Hermes transposon of *Musca domestica* and its use as a mutagen of *Schizosaccharomyces pombe*. Methods 49:243–7.)

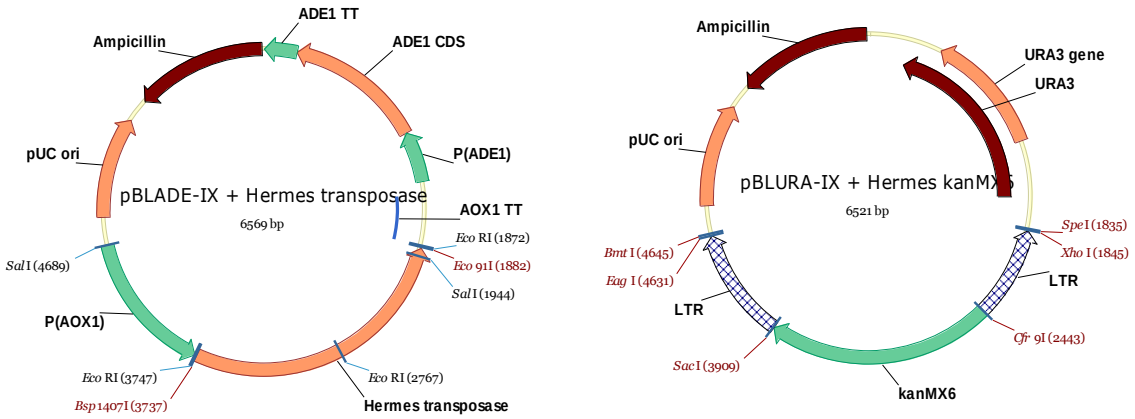


Figure 1: Hermes transposase expression plasmid (left) and Hermes transposon donor plasmid (right). Hermes transposase expression from P_{AOX1} is tightly regulated, as the enzyme should not be constantly expressed. The expression plasmid uses the *ADE1* gene as a marker. The *KanMX6* marker cassette on the donor plasmid is flanked by long terminal repeats (LTRs). Hermes transposase can bind to these LTRs and facilitate transposition of the cassette. The donor plasmid uses the *URA3* gene as a marker.

Zeta* transposon from *Yarrowia lipolytica

(Literature reference: Mauersberger S, Wang H, Gaillardin C, Barth G, Nicaud JM (2001) Insertional mutagenesis in the n-alkane-assimilating yeast *Yarrowia lipolytica*: generation of tagged mutations in genes involved in hydrophobic substrate utilization. J Bacteriol 183:5102–5109.***)***

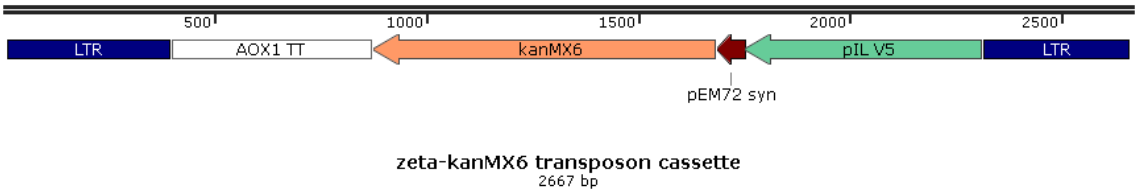


Figure 2: *Zeta* transposon cassette with *KanMX6* marker. The *KanMX6* marker cassette conferring resistance towards geneticin in *P. pastoris* was cloned between the long terminal repeats (LTRs), derived from the *zeta* transposon of *Y. lipolytica*.

2 Expression constructs for screening

Plasmids for expression of HRP

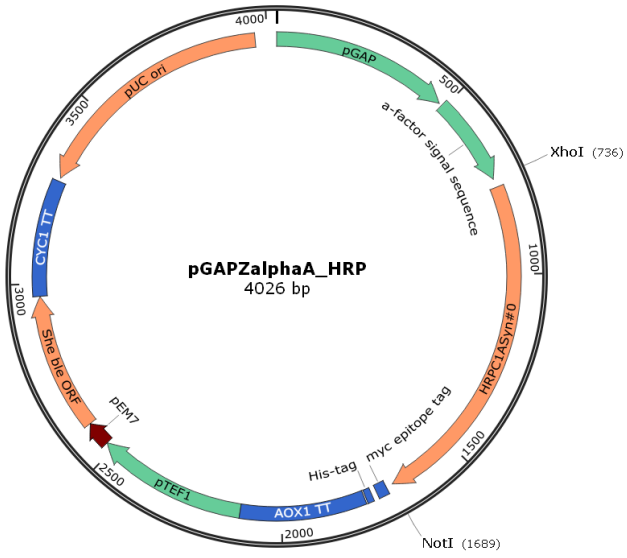


Figure 3: Plasmid pGAPZαA with HRP-C1A CDS integrated between *Xho*I and *Not*I sites.

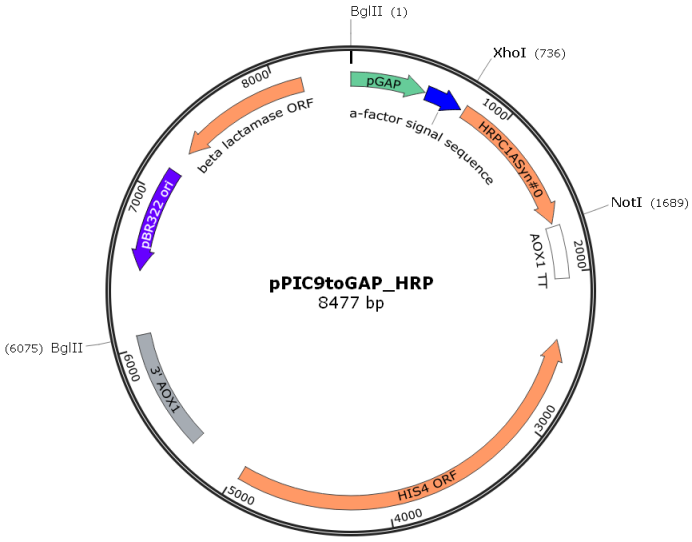


Figure 4: Plasmid pPIC9toGAP with HRP-C1A CDS integrated between *Xho*I and *Not*I sites. The plasmid was constructed by restricting pGAPZαA-HRP and pPIC9 with *Bgl*II and *Not*I. The resulting fragments containing P_{GAP}+α-signal sequence+HRP, *HIS4* marker and pBR322 ori+beta-lactamase CDS were then ligated to yield this plasmid.

Plasmids for expression of APLE

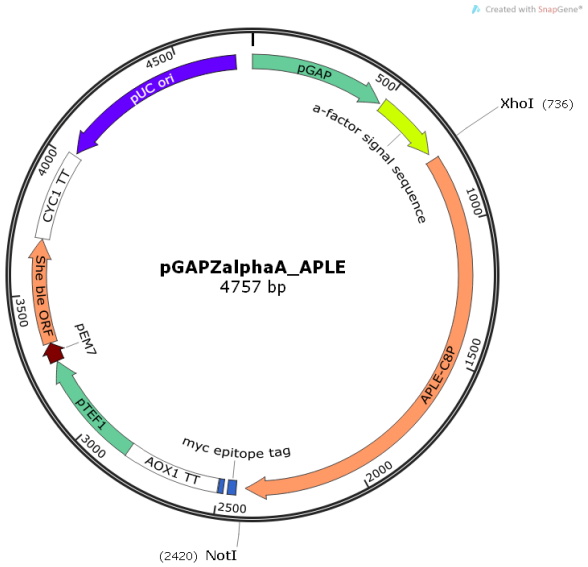


Figure 5: Plasmid pGAPZαA with codon-optimized APLE sequence integrated between *Xho*I and *Not*I sites.

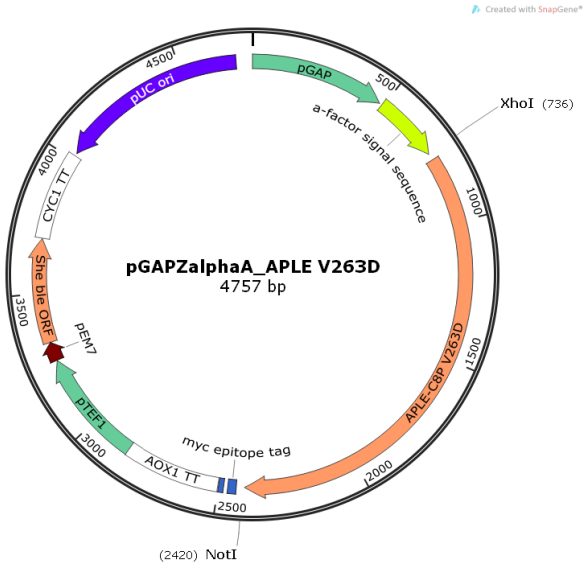


Figure 6: Plasmid pGAPZαA with codon-optimized APLE V263D sequence integrated between *Xho*I and *Not*I sites. The V263D amino acid exchange leads to disruption of the native trimeric structure of APLE, and triggers the formation of monomers.

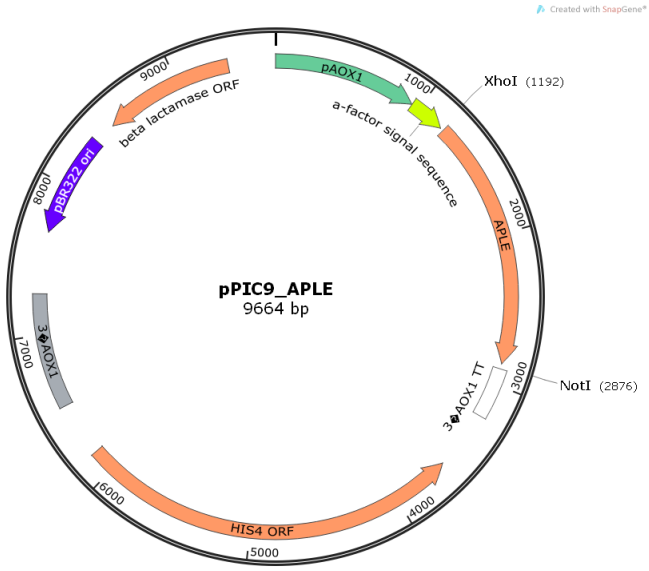


Figure 7: Plasmid pPIC9 with codon-optimized APLE sequence integrated between *Xho*I and *Not*I sites. A similar expression vector was constructed for the APLE V263D variant.

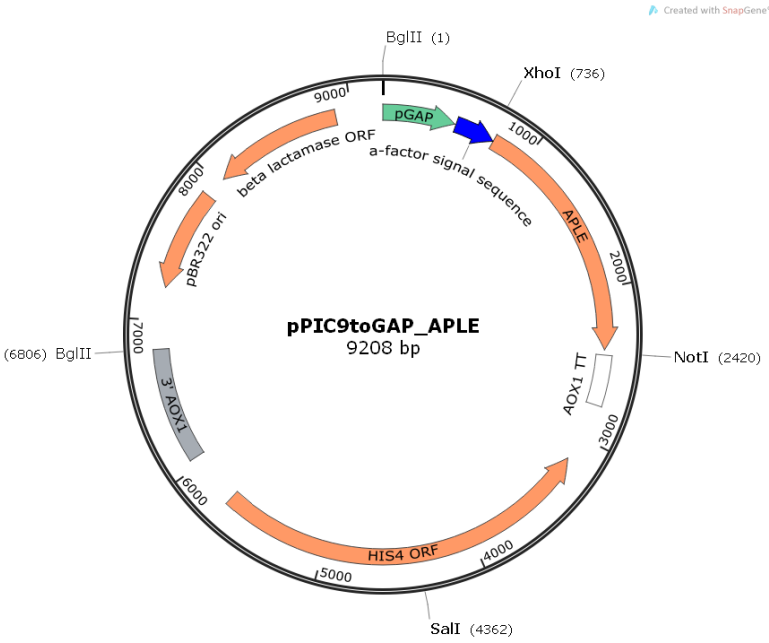


Figure 8: Plasmid pPIC9toGAP with the codon-optimized APLE sequence integrated between *Xho*I and *Not*I sites. The plasmid was constructed by restricting pGAPZαA-APLE and pPIC9 with *Bgl*II and *Not*I. The resulting fragments containing P_{GAP}+α-signal sequence+APLE, *HIS4* marker and pBR322 ori+beta-lactamase CDS were then ligated to yield this plasmid. A similar expression vector was constructed for the APLE V263D variant.

Plasmids for expression of MFα1

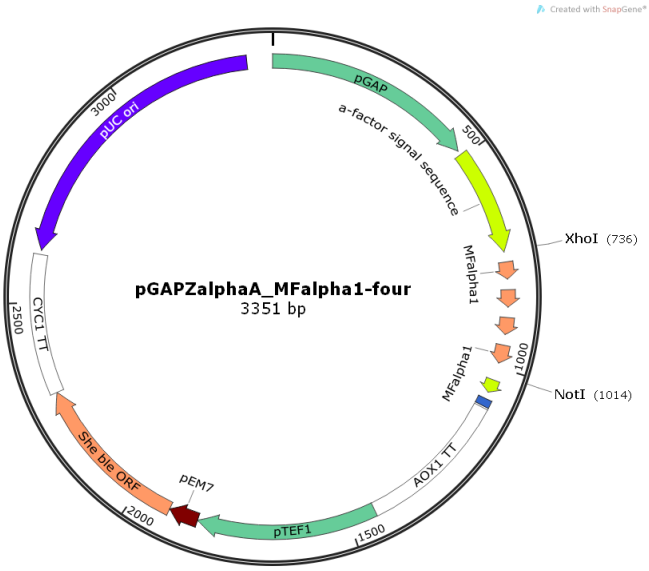


Figure 9: Plasmid pGAPZαA containing a complete MFα1 locus from *S. cerevisiae*. In *S. cerevisiae*, four copies of the mature MFα1 peptide are encoded at the genomic locus.

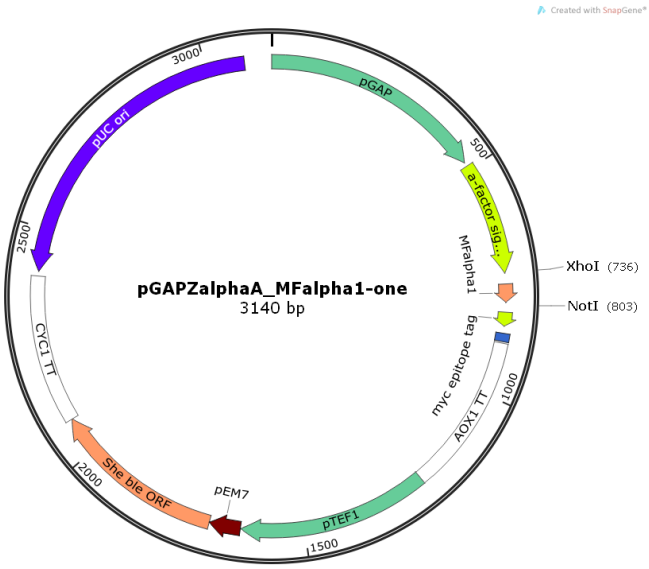


Figure 10: Plasmid pGAPZαA with a single copy of the MFα1 CDS from *S. cerevisiae* integrated between *Xho*I and *Not*I sites.

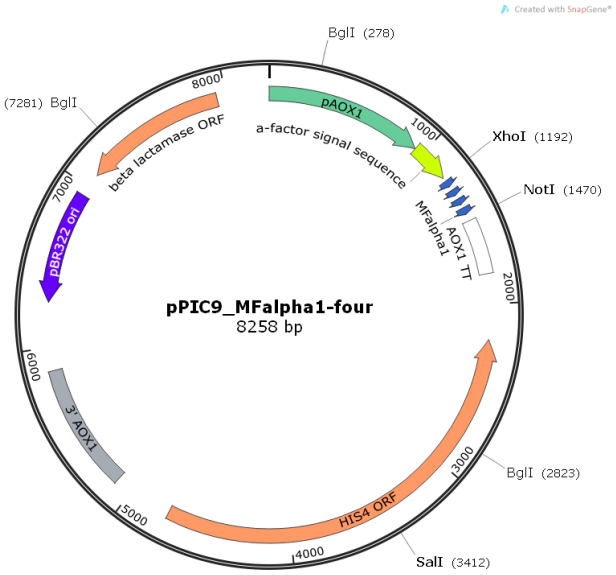


Figure 11: Plasmid pPIC9 with the complete MFα1 locus from *S. cerevisiae*. A similar plasmid was constructed containing only one copy of the MFα1 gene.

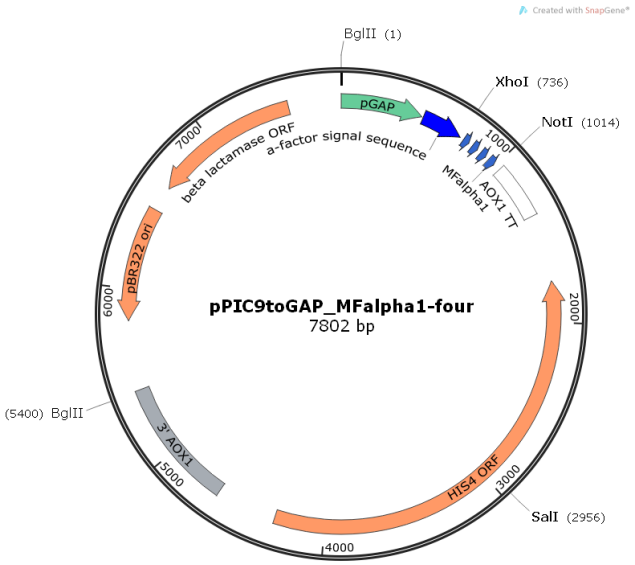


Figure 12: Plasmid pPIC9toGAP with four copies of the MFα1 sequence integrated between *Xho*I and *Not*I sites. The plasmid was constructed by restricting pGAPZαA-MFalpha1-four and pPIC9 with *Bgl*II and *Not*I. The resulting fragments containing P_{GAP}+α-signal sequence+MFα1, *HIS4* marker and pBR322 ori+beta-lactamase CDS were then ligated to yield this plasmid. A similar expression vector was constructed for the MFα1-one expression cassette.

3 Protocols for alternative screening assays

APLE-screening with para-nitrophenyl acetate (pNPA) assay

Esterase activity of APLE was quantified in 96-well format. For expression from P_{GAP}, cultures were grown in DWPs on 500 µL BYPD per well for 72 h. Cells were removed by centrifugation (10 min, 22°C, 1500 x g). Then, 10 µL of the culture supernatant were mixed with 290 µL of 2 mM pNPA in 100 mM Tris-HCl, pH 7.0. Increase in absorbance at 405 nm (ε=9.5946mM−1 cm−1) was determined at 25°C with a Biotek Gen5 spectrophotometer or by comparison to an APLE-secreting reference strain by eye.

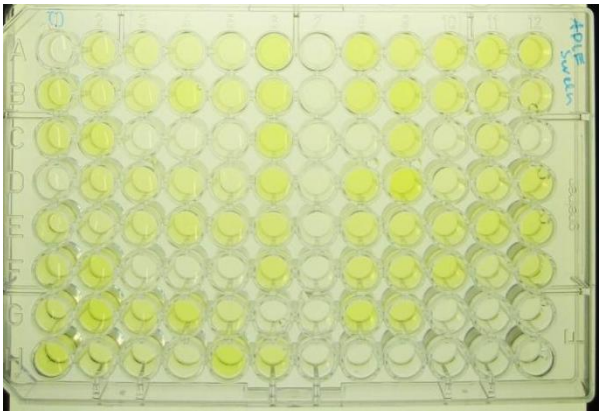


Figure 13: pNPA screening assay for APLE esterase activity. After cultures had been grown on BYPD for 72 h, 10 µL of culture supernatant were mixed with 290 µL of 2 mM pNPA in 100 mM Tris-HCl, pH 7.0.

MFα1-screening on Halo plates

We considered employing a variation of the halo assay developed by Manney (1983) to quantify recombinant *S. cerevisiae* α-mating factor secreted from *P. pastoris* cells. Therefore, we constructed *P. pastoris* strains expressing one or four copies of the *S. cerevisiae* α-mating factor gene from P_{GAP} or P_{AOX1}. Moreover, we aimed to modify the assay to allow higher screening throughput. Several rounds of further development led us to a protocol where YPD agar (2 % agar) was topped with a thin layer of "halo" agar (YPD and 0.8 % agar containing viable cells of *S. cerevisiae* MAT*a sst1 sst2 leu1 trp5 ade2 can1*). The two-layered agar was stable enough to allow

the pinning of *P. pastoris* culture supernatants from deep-well plates (DWPs) using a 96-pin metal pinning head. Previous tests, where the α-factor secreting yeast strains were spotted on the halo agar directly were complicated by growth to different colony size, which, in turn, made comparison of the formed halos impossible. We, therefore, decided to grow *P. pastoris* cells in 96-well DWPs on YPD for several days. To prevent transfer of viable cells, the cultures were autoclaved prior to pinning. Subsequently, the autoclaved DWPs were centrifuged and the culture supernatants, containing still intact α-factor peptide, were pinned onto the halo agar.

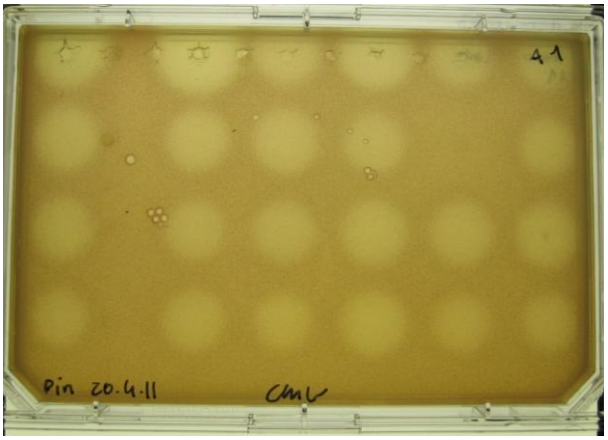


Figure 14: Halo assay involving pinning of autoclaved cultures. Every second well in a 96-well plate was inoculated with strains secreting MFα1 (4 copies of the gene, expressed from P_{GAP}). Column 3 was used for negative controls. Cultures were grown for 72 h on YPD, autoclaved and the supernatants were pinned onto Halo agar (YPD and 0.8 % agar containing viable cells of *S. cerevisiae* MAT*a sst1 sst2 leu1 trp5 ade2 can1*) layered on YPD agar containing 2 % agar. Agar plates were incubated at 30°C until the halos surrounding the pinning spots were clearly visible (3-4 days).

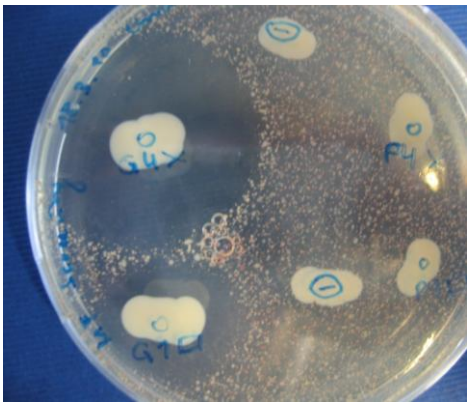


Figure 15: Original approach for Halo plate-screening. *P. pastoris* cells secreting MFα1 were streaked on Halo medium (YPD and 0.8 % agar containing viable cells of *S. cerevisiae* MAT*a sst1 sst2 leu1 trp5 ade2 can1*) and the plates were incubated for 4 days at 30°C. Growth to different colony size made comparison of the formed halos more difficult. Clearly visible is the different halo diameter of constructs with four copies of the MFα1 gene to constructs with only one copy of the gene. G = *GAP* promoter; P = *AOX1* promoter; 4 = four copies of MFα1; 1 = one copy of MFα1; Θ = negative control

Glucose-UV assay for quantification of levanase from *Bacillus subtilis*

Strains secreting *B. subtilis* levanase from P_{AOX1} were grown in DWPs and induced with methanol for 48 h. Cultures were centrifuged at 1,500 x g for 10 min to pellet cells. Fifty µL of the culture supernatant were mixed with 50 µL of 50 mg/mL sucrose in microtiter plates. Following incubation at 37°C for 10 min, 10 µL of the mix were added to 190 µL of Glucose-UV solution (DiproMed). The microtiter plates were again incubated at 37°C for 10 min. Absorbance at 340 nm was recorded with a Biotek Gen5 spectrophotometer.

Screen for intracellular EGFP expression

EGFP was expressed intracellularly from plasmid pPT4-GAP-EGFP- Arg4-Zeocin. Intracellular fluorescence of EGFP was quantified with a Biotek Gen5 spectrophotometer (Biotek) exciting at 395 nm and detecting emission at 507 nm. We also measured extracellular EGFP fluorescence to control possible leakage of cells. Here, the cultures were centrifuged 1,500 x g for 10 min and the supernatants were analyzed spectrophotometrically.

4 Additional results from random and targeted mutagenesis of HRP-secreting strains

Results from random mutagenesis:

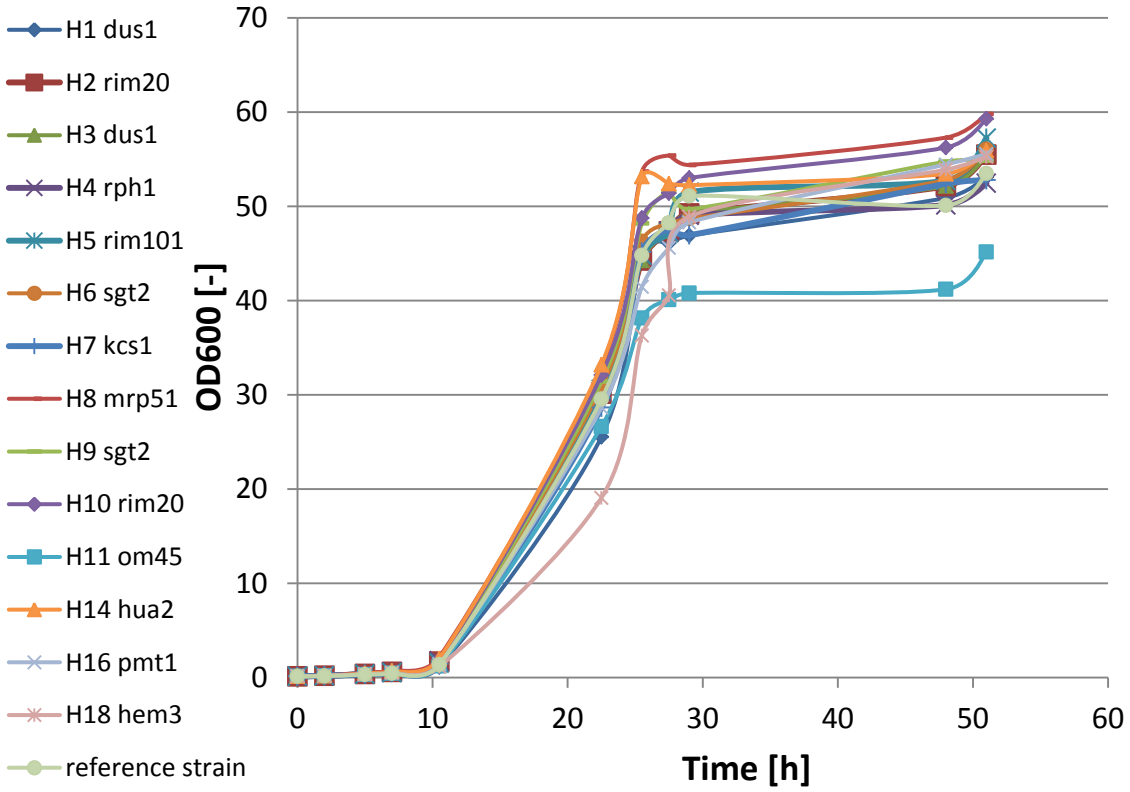


Figure 16: Growth curve analysis of control and mutant strains (from random mutagenesis) secreting HRP from *GAP* promoter. Cells were grown in BYPD (2% glucose) in baffled shake flasks at 28°C.

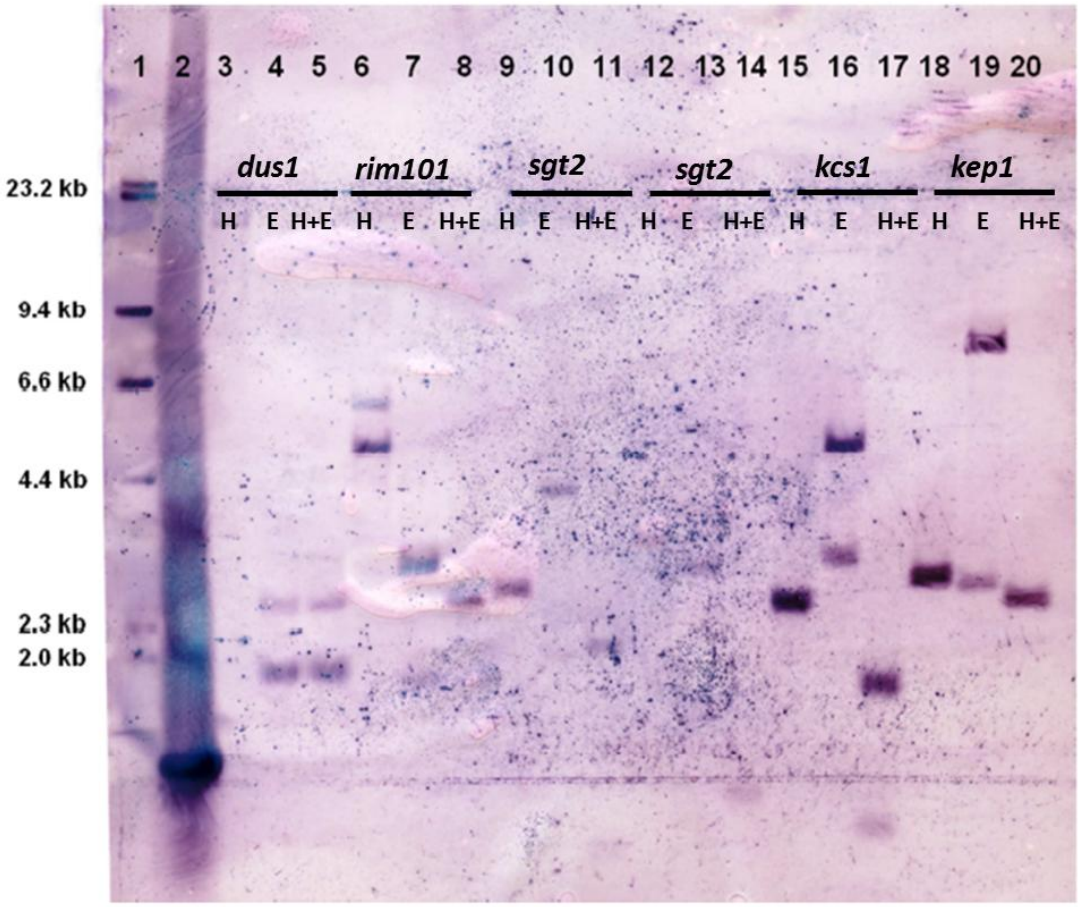


Figure 17: Southern blot analysis of HRP-secreting mutants selected in the screening. Probe: Zeocin^R cassette that was used for mutagenesis; E: *Eco*RV cuts in Zeocin^R cassette, therefore two fragments per integration event were expected; H: *Hin*dIII does not cut in cassette, therefore one fragment per integration event was expected; Ladder (lane 1): DNA Molecular weight marker II, DIG-labeled (Roche); positive control (lane 2): plasmid DNA, containing the Zeocin^R cassette. On the contrary to results obtained from genome walking, the *rim101* mutant strain seemed to harbor two mutagenesis cassettes, while apparently only one integration event had taken place in the *kep1* mutant strain.

Results from targeted knockouts:

Knockout maps

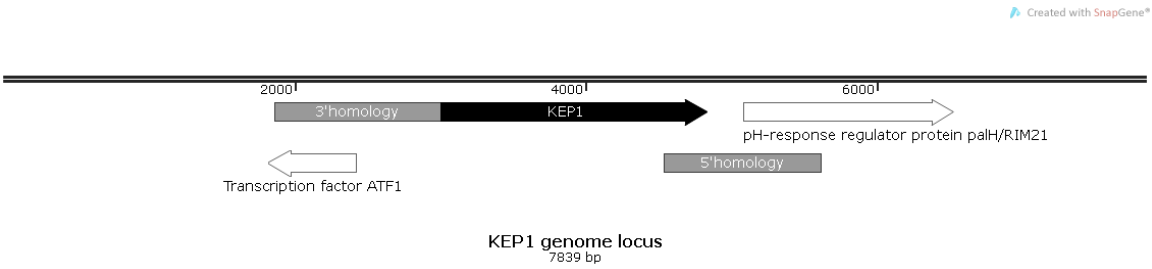


Figure 18: Targeted knockout of *KEP1* (CCA40244.1). Position of 5' and 3' homology regions integrated in knockout vector pPKC1 are indicated.

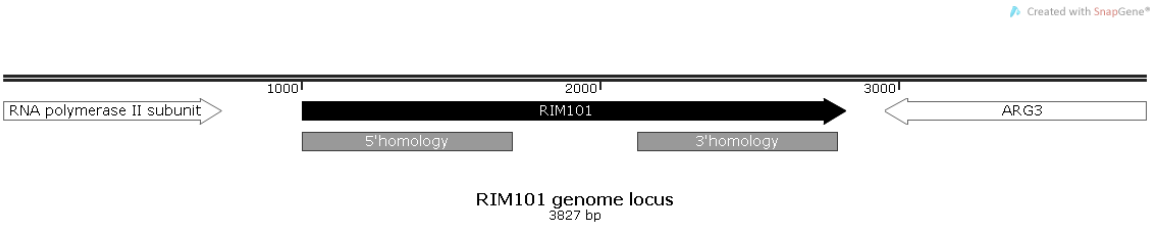


Figure 19: Targeted knockout of *RIM101* (CCA39536.1). Position of 5' and 3' homology regions integrated in knockout vector pPKC1 are indicated.

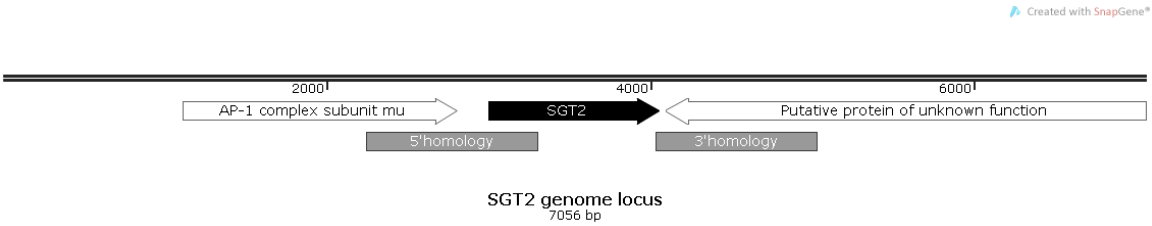


Figure 20: Targeted knockout of *SGT2* (CCA37018.1). Position of 5' and 3' homology regions integrated in knockout vector pPKC1 are indicated.

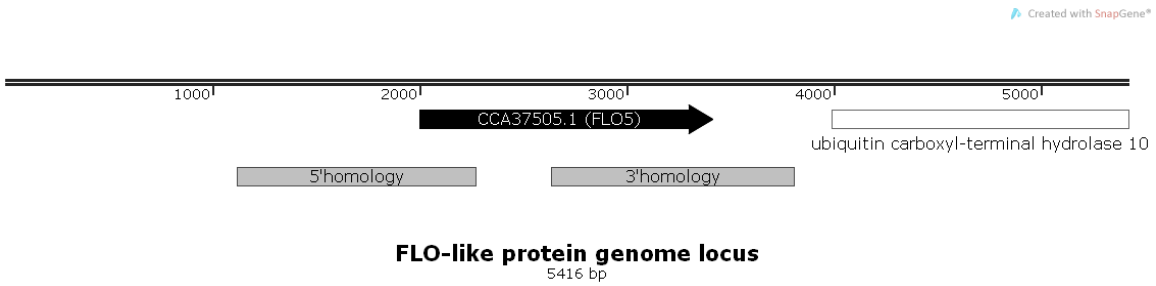


Figure 21: Strategy for targeted knockout of *FLO5* (CCA37505.1). Position of 5' and 3' homology regions integrated in knockout vector pPKC1 (Zeo^R) & pPKC3 (*HIS4*) are indicated. We were not successful in generating this knockout. The protein CCA37505.1, annotated as Flo9, is encoded by the ORF *FLO5*.

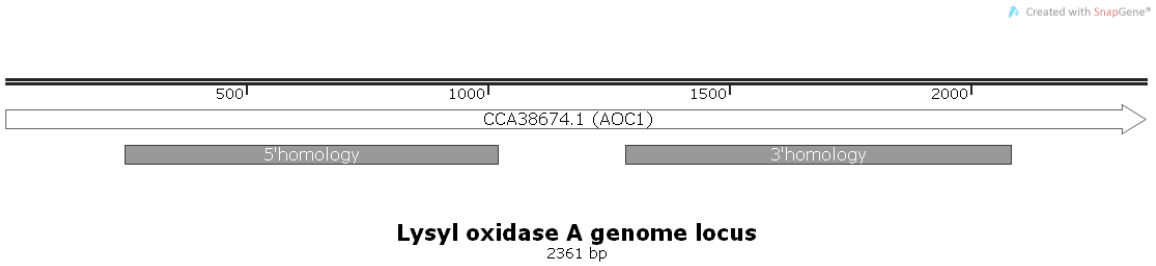


Figure 22: Targeted knockout of *Lysyl oxidase A* (CCA38674.1). Position of 5' and 3' homology regions integrated in knockout vector pPKC1 are indicated.

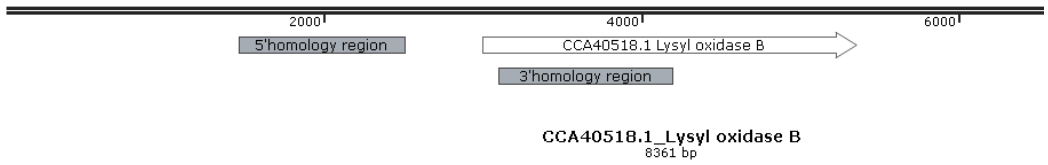


Figure 23: Targeted knockout of *Lysyl oxidase B* (CCA40518.1). Position of 5' and 3' homology regions integrated in knockout vector pPKC1 are indicated.

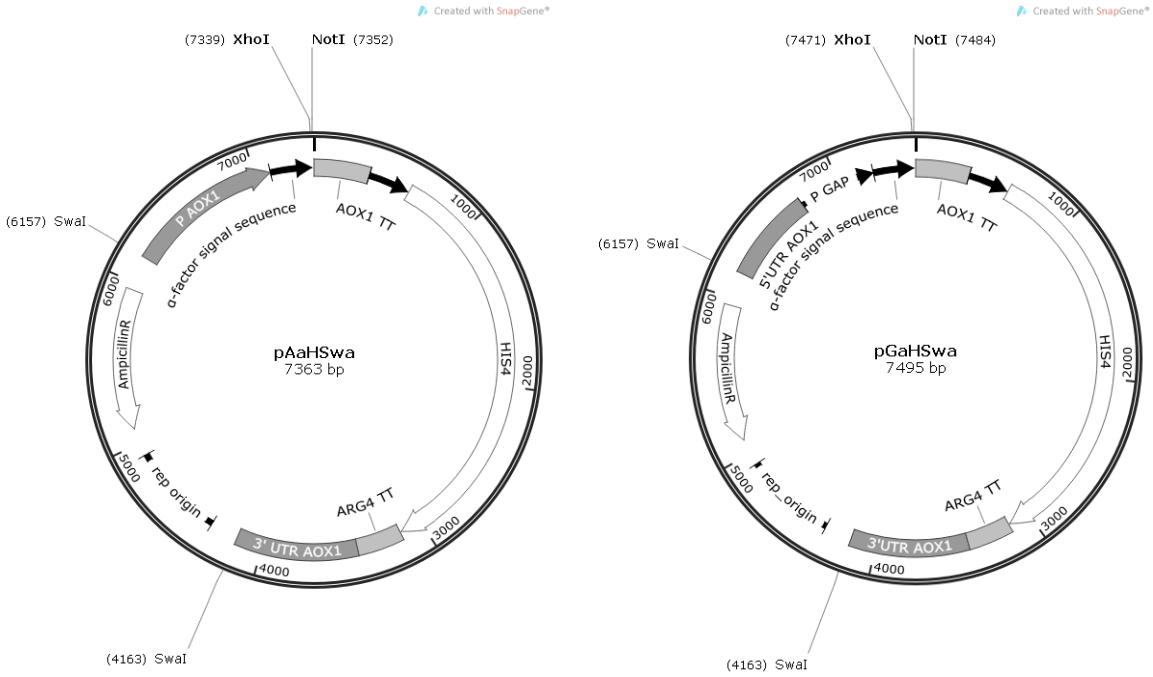
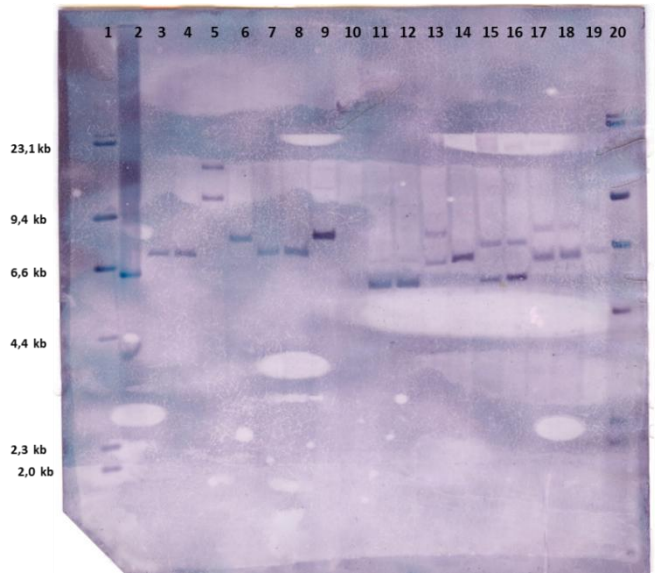
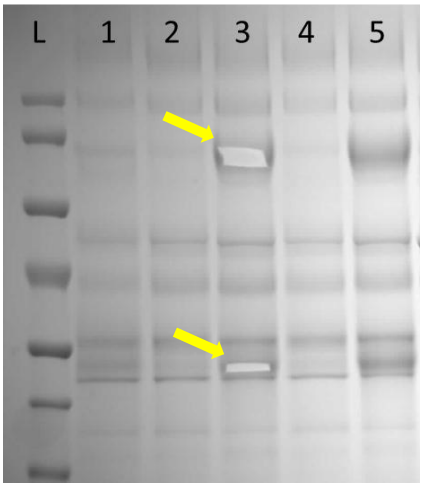


Figure 24: Expression plasmids pAaHSwa (left) and pGaHSwa (right). The plasmids can be targeted to the *AOX1* locus by restriction with *Swa*I. The target gene can be integrated downstream of P_{AOX1}/P_{GAP} and the α-mating factor signal sequence by cloning with *Xho*I and *Not*I. The *HIS4* gene acts as selectable marker. Kindly provided by Mudassar Ahmad.



Lane	Sample	Restr. Enzyme	Expected size
1	Ladder, 18 µL		
2	Plasmid fragment with *PAOX1*-HRP and *HIS4* (SwaI)		6330 bp
3	GAP- A1	HindIII	6900 bp
4	GAP- A2	HindIII	6900 bp
5	GAP- E1	HindIII	7600 bp
6	GAP- E2	HindIII	7600 bp
7	GAP- C1	HindIII	6900 bp
8	GAP- C2	HindIII	6900 bp
9	GAP- G1	HindIII	7600 bp
10	GAP- G2	HindIII	7600 bp
11	AOX1- A1	HindIII	5600 bp
12	AOX1- A2	HindIII	5600 bp
13	AOX1- E1	HindIII	6300 bp
14	AOX1- E2	HindIII	6300 bp
15	AOX1- C1	HindIII	5600 bp
16	AOX1- C2	HindIII	5600 bp
17	AOX1- G1	HindIII	6300 bp
18	AOX1- G2	HindIII	6300 bp
19	CBS7435 *his4* (control)	HindIII	No band
20	Ladder, 18 µL		

Figure 25: Southern blot to select reference strains for HRP and APLE screenings (representative result). Correct and single integration events of pAaHSwa and pGaHSwa expression cassettes at the *AOX1* locus were controlled with probes directed to the *HIS4* gene, the *AOX1* and *GAP* promoter, respectively. Strains that showed the expected bands where used as reference strains in the following HRP and APLE screening experiments.



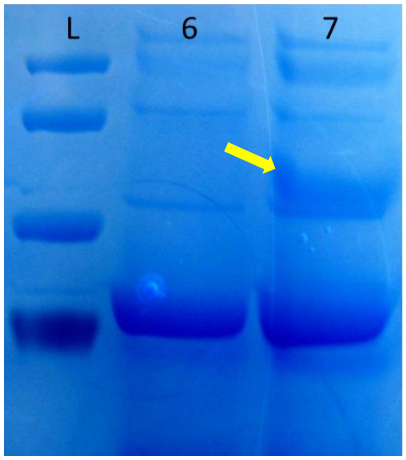


Figure 26: Coomassie blue-stained SDS-PAGE of native (left) and EndoH-treated (right) culture supernatants. The bands indicated by arrows were excised and analyzed by mass spectrometry. L: PageRuler^{TM} Prestained Protein Ladder (Thermo Scientific); 1: WT; 2: WT expressing HRP from P_{GAP}; 3: *kep1Δ* expressing HRP from P_{GAP}; 4: WT expressing HRP from P_{GAP}; 5: *kep1Δ* expressing HRP from P_{GAP}, 6: WT; 7: *kep1Δ*. Strains were grown on BYPD (2 % glucose) for 72 h. Proteins in 244 µL of culture supernatants were precipitated by adding TCA to a final concentration of 33%. In case of samples 6 and 7, the supernatant had been treated with $EndoH_f$ (New England Biolabs) before.

Secretory levanase expression in knockout strains

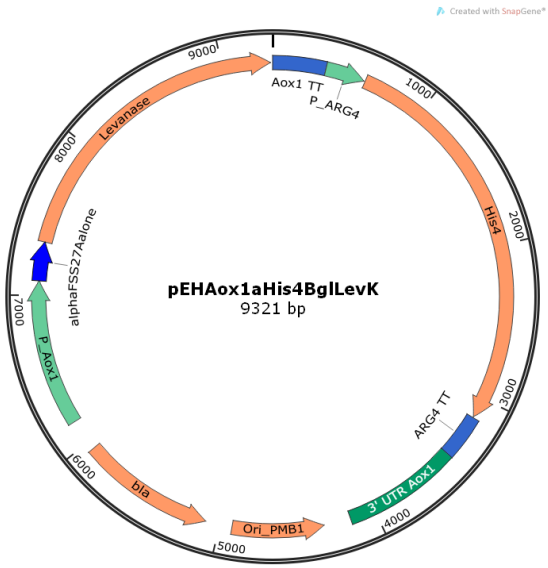


Figure 27: Plasmid pEHAox1aHis4BglLevK for secreted expression of levanase from *B. subtilis*. The gene encoding levanase was cloned downstream of P_{AOX1} and α-mating factor secretion signal sequence. Plasmid kindly provided by Mudassar Ahmad.

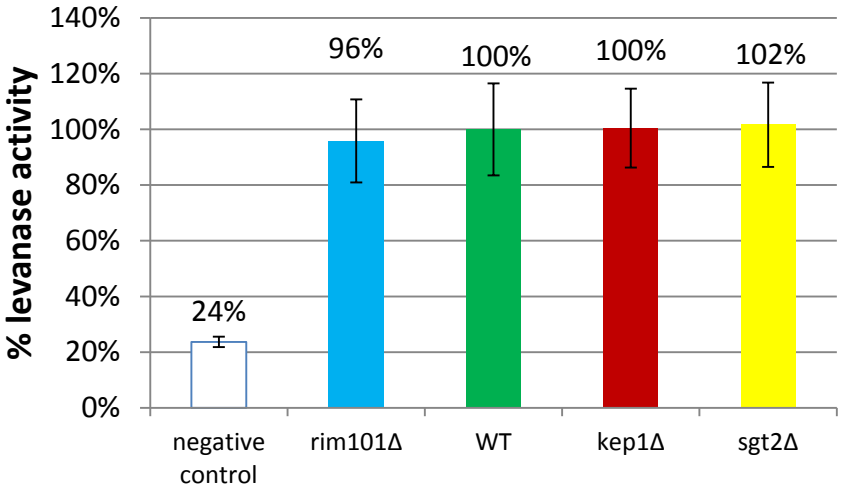


Figure 28: Relative activity of secreted levanase, expressed from P_{AOX1}, in deep-well plate culture supernatants of targeted knockout strains. WT control and mutant strains expressing levanase from P_{AOX1}, and control strain not expressing levanase, were grown in BMGY for 32 h and induced with methanol for 48 h. Levanase activity in supernatants was measured with Glucose-UV assay. Results represent the mean of one biologically independent experiment with 12 technical replicates per knockout strain.

Intracellular EGFP expression in knockout strains

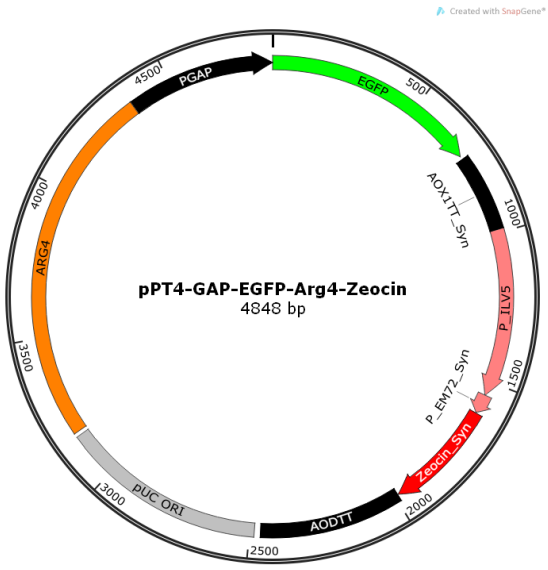


Figure 29: Plasmid pPT4-GAP-EGFP-Arg4-Zeocin for intracellular expression of EGFP. The gene encoding EGFP was inserted downstream of P_{GAP}. The plasmid was linearized in the *ARG4* homology region to target it to this locus in *P. pastoris*. Plasmid kindly provided by Lukas Sturmberger and Thomas Vogl.

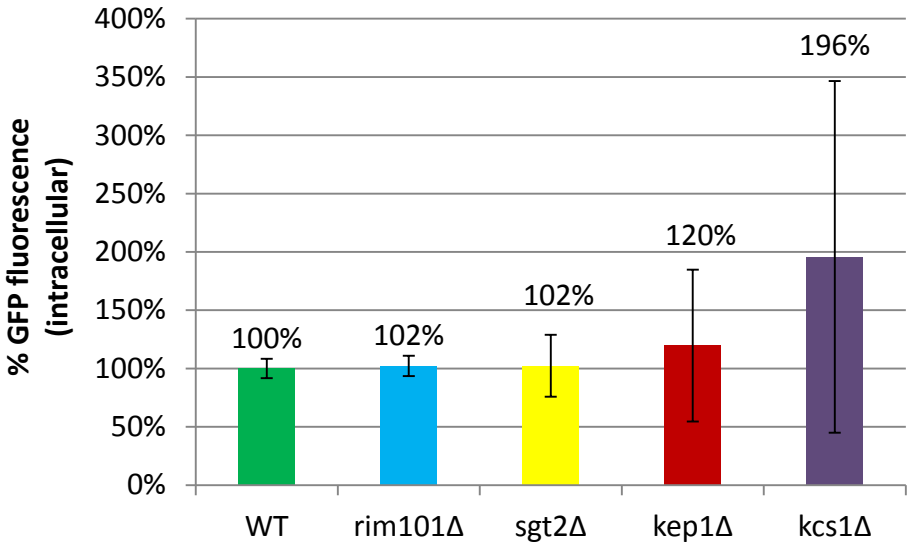


Figure 30: Relative fluorescence of EGFP, expressed from P_{GAP}, inside cells of targeted knockout strains. WT control and mutant strains expressing EGFP from P_{GAP} were grown in BYPD for 48 h. EGFP fluorescence was recorded spectrophotometrically. Results represent the mean of one biological experiment with 12 technical replicates per knockout strain.

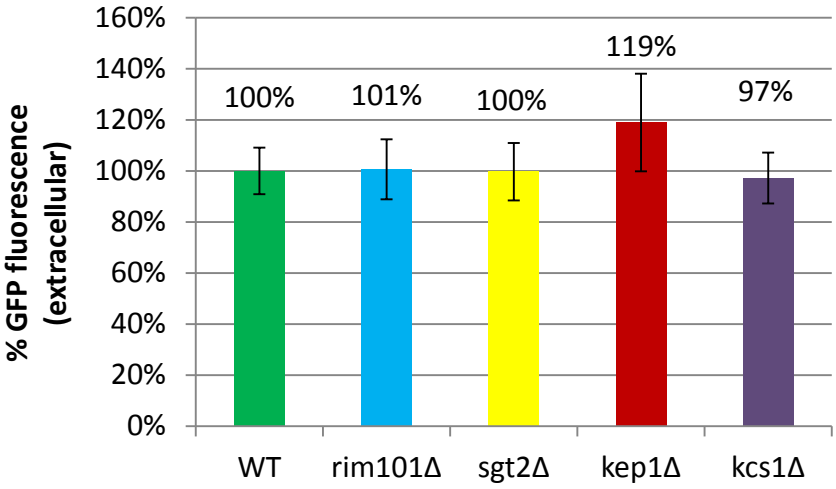


Figure 31: Relative fluorescence of EGFP, expressed from P_{GAP}, in the culture supernatant of targeted knockout strains. WT control and mutant strains expressing EGFP from P_{GAP} were grown in BYPD for 48 h and centrifuged for 10 min at 1,500 x g. EGFP fluorescence in the supernatants was recorded spectrophotometrically. Results represent the mean of one biological experiment with 12 technical replicates per knockout strain.

Restored gene expression with flag-tag

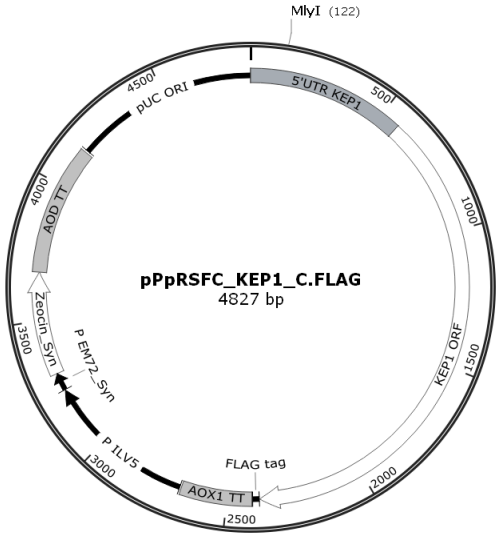


Figure 32: Plasmid pPpRSFC_KEP1_C.FLAG for restored expression of *KEP1* from its native promoter. The *KEP1* ORF and 5' untranslated region were cloned upstream of a flag-tag. A similar plasmid with a C-terminal EGFP-tag instead of the flag-tag, and variations with the *GAP* promoter instead of the gene's native promoter were also constructed.

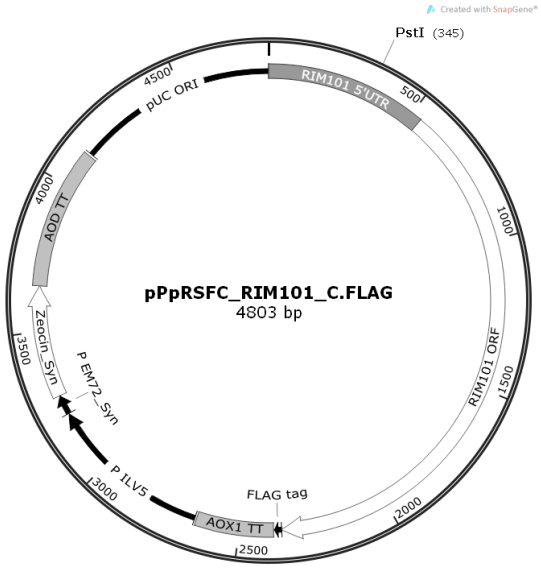


Figure 33: Plasmid pPpRSFC_RIM101_C.FLAG for restored expression of *RIM101* from its native promoter. The *RIM101* ORF and 5' untranslated region were cloned upstream of a flag-tag. A similar plasmid with a C-terminal EGFP-tag instead of the flag-tag, and variations with the *GAP* promoter instead of the gene's native promoter were also constructed.

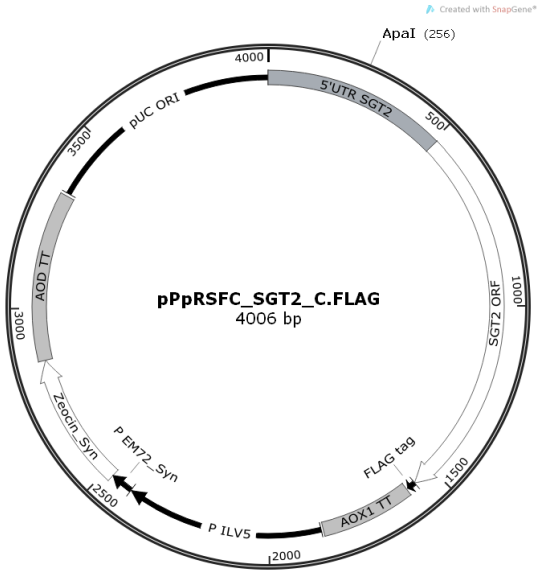


Figure 34: Plasmid pPpRSFC_SGT2_C.FLAG for restored expression of *SGT2* from its native promoter. The *SGT2* ORF and 5' untranslated region were inserted upstream of a flag-tag. A similar plasmid with a C-terminal EGFP-tag instead of the flag-tag, and variations with the *GAP* promoter instead of the gene's native promoter were also constructed.

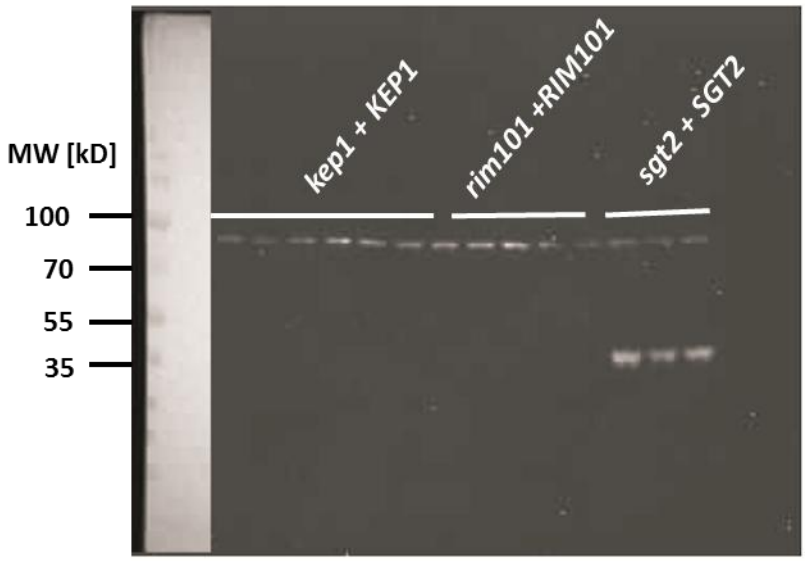


Figure 35: Western blot to detect expression of flag-tagged Kep1, Rim101 and Sgt2 in the respective knockout strains. The tagged proteins were expressed from P_{GAP}. Strains were grown on BYPD (2% glucose) for 72 h and cells were disrupted as devised by Horvath and Riezman, 1994. A primary anti-flag antibody was used for detection.

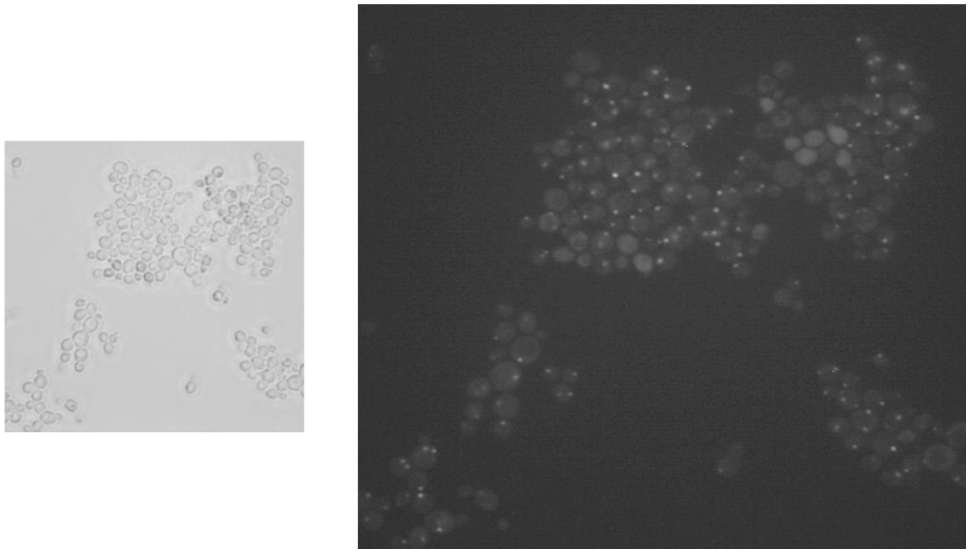


Figure 36: Fluorescence microscopy (right panel) and light microscopy (left panel) of CBS7435 *sgt2Δ* P_{GAP}-αHRP P_{GAP}-*SGT2*-EGFP cells. Cells were grown to OD_{600} = 1.0 on YPD. *SGT2* expressed from P_{GAP} seemed to accumulate in punctuate structures inside the cell.

5 Results from random and targeted mutagenesis of APLE-secreting strains

Table 1. Summary of identified mutants with positive effect on APLE activity in culture supernatant

Mut. N°	Accession N° of affected ORF in *P. pastoris* GS115	Accession N° of protein in *P. pastoris* CBS7435	Homologs [a]	Max. score	Query coverage	Max. identity	Comments
A1	XM_002490622.1	CCA37453.1	**Nam2p**	887	98%	50%	Mitochondrial leucyl-tRNA synthetase
A2	XM_002492489.1	CCA39851.1	Gtr1p	372	98%	59%	Cytoplasmic GTPase involved in stimulation of TORC1 in response to amino acids
A3	XM_002492067.1	CCA37798.1	Bem2p	676	90%	28%	Rho GTPase activating protein required for bud emergence
A3	XM_002490902.1	CCA38919.1	Mon2p	534	99%	28%	Protein with a role in endocytosis and vacuole integrity
A5	XM_002491728.1	CCA38127.1	Apl6p	281	82%	30%	Beta3-like subunit of the yeast AP-3 complex
A8	XM_002489328.1	CCA36203.1	Pam16p	92	74%	57%	Subunit of the PAM complex
A10	XM_002493302.1	no ORF annotated	Erd1p	92	88%	26%	Predicted membrane protein required for lumenal ER protein retention
A11	XM_002489731.1	CCA36593.1	**Cyc8p**	108	56%	24%	General transcriptional co-repressor; acts together with Tup1p
A13	XM_002491645.1	CCA38204.1	**Cat8p**	376	57%	35%	Zinc cluster transcriptional activator; binds carbon source responsive elements
A15	XM_002493014.1 (ATG 150 bp away)	CCA39327.1	Gpm1p	378	92%	80%	Tetrameric phosphoglycerate mutase; acts in glycolysis and gluconeogenesis
A16	XM_002489669.1	CCA36532.1	(Def1p)	58	7%	51%	No significant homology with *S. cerevisiae* proteome
A17	XM_002493877.1	CCA40655.1	**(Adr1p)**	133	10%	55%	Transcription factor Mxr1p (methanol expression regulator I) in *P. pastoris*
A17	XM_002492014.1	CCA37849.1	(Sul1p)	30	29%	27%	No significant homology with *S. cerevisiae* proteome
A18	XM_002490444.1	CCA37281.1	**Spt8p**	381	94%	42%	Subunit of the SAGA transcriptional regulatory complex
A20	XM_002493467.1	CCA41060.1	Rpl2bp	409	99%	83%	Ribosomal 60S subunit protein L2B

Table 2. Summary of identified mutants with negative effect on APLE activity in culture supernatant

Mut. N°	Accession N° of affected ORF in *P. pastoris* GS115	Accession N° of protein in *P. pastoris* CBS7435	Homologs [a]	Max. score	Query coverage	Max. identity	Comments
A23	XM_002493575.1	CCA40949.1	Bzz1p	421	99%	39%	SH3 domain protein implicated in regulating actin polymerization
A26	XM_002490318.1	CCA37157.1	Smm1p	344	80%	56%	Dihydrouridine synthase
A27	XM_002493376.1	CCA41153.1	(Flo19p)	34	9%	30%	No significant homology with *S. cerevisiae* proteome
A29	XM_002494206.1	CCA40323.1	Ydl206w	95	42%	34%	Putative protein of unknown function
A34	XM_002490042.1 (ATG 180 bp away)	CCA36889.1	Ena1p	1171	96%	57%	P-type ATPase sodium pump; involved in Na+ and Li+ efflux
A36	XM_002493242.1	CCA39094.1	**Pas1p (=Pex1p)**	468	64%	47%	AAA-peroxin; participates in the recycling of peroxisomal signal receptor Pex5p
A37	XM_002490640.1	CCA37471.1	Mph1p	644	80%	47%	3'-5' DNA helicase involved in error-free bypass of DNA lesions
A39	XM_002490902.1	CCA38919.1	Mon2p	534	99%	28%	Protein with a role in endocytosis and vacuole integrity
A41	XM_002493583.1	CCA40941.1	(Nst1p)	66	12%	30%	No significant homology with *S. cerevisiae* proteome
A42	XM_002492486.1	CCA39855.1	Vps36p	139	99%	26%	Component of the ESCRT-II complex

Entries in bold underline open reading frames that were selectively knocked out for further analysis

[a] Protein homolog with highest max. score in *S. cerevisiae* identified by Protein BLAST search

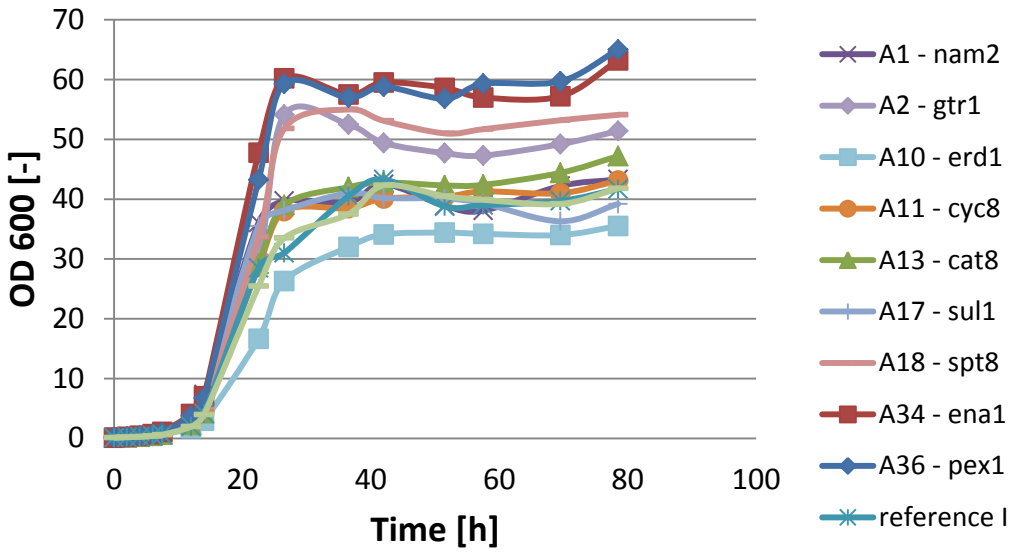


Figure 37: Growth curve analysis of control and mutant strains (from random mutagenesis) secreting APLE from *GAP* promoter. Cells were grown in BYPD (2% glucose) in baffled shake flasks at 28°C.

Table 3. Overview of APLE activity in mutant strains (from random mutagenesis) relative to wild type control in shake flask and DWP cultivation experiments (Mutant IDs A1 – A36 refer to Tables 1 and 2)

	Screen	Flask	Flask corr.	Gly.stock DWP	Gly.stock DWP corr.	Plate DWP	Plate DWP corr.
A1	+	~	~	~	-	~	-
A2	+	+	+	+	~	+	~
A3	+			+	-	+	~
A10	+	-	-	+	~	-	-
A11	+	+	+	~	-	+	~
A13	+	~	~	+	+	~	~
A17	+	+	+	+	+	+	~
A18	+	+	~	~	-	+	-
A34	-	+	-	-	-	~	-
A36	-	+	-	-	-	-	-

+ better than control - not as good as control as control corr.= corrected for growth

Targeted knockouts:

Knockout maps

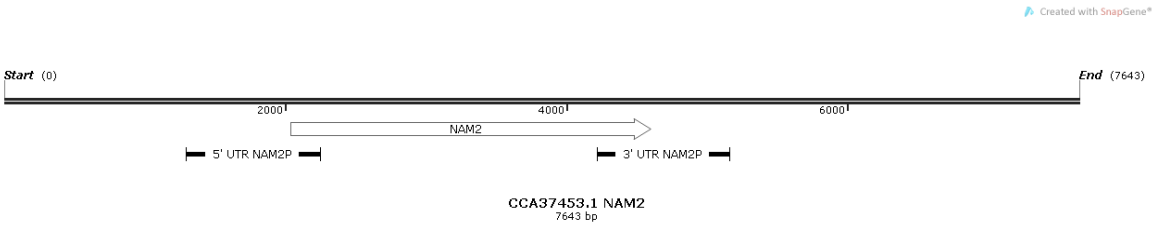


Figure 38: Targeted knockout of *NAM2*. Position of 5' and 3' homology regions integrated in knockout vector pPKC1 are indicated.

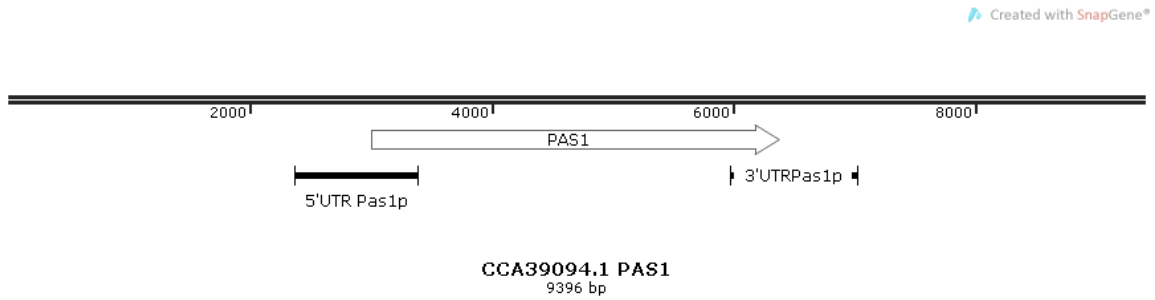


Figure 39: Targeted knockout of *PAS1 (PEX1)*. Position of 5' and 3' homology regions integrated in knockout vector pPKC1 are indicated.

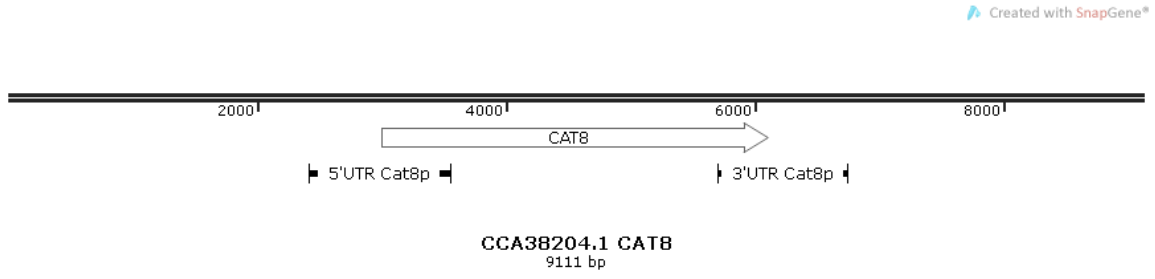


Figure 40: Targeted knockout of *CAT8*. Position of 5' and 3' homology regions integrated in knockout vector pPKC1 are indicated.

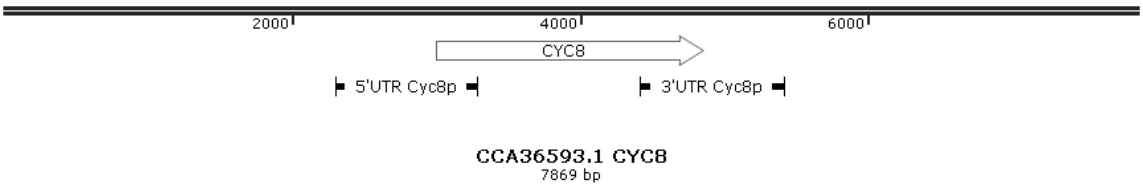


Figure 41: Targeted knockout of *CYC8*. Position of 5' and 3' homology regions integrated in knockout vector pPKC1 are indicated.

Figure 42: Targeted knockout of *SPT8*. Position of 5' and 3' homology regions integrated in knockout vector pPKC1 are indicated.

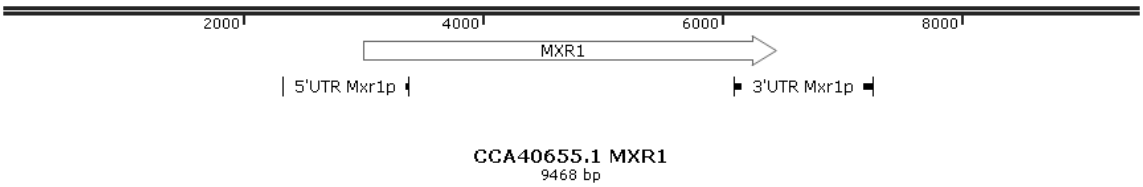


Figure 43: Targeted knockout of *MXR1*. Position of 5' and 3' homology regions integrated in knockout vector pPKC1 are indicated.

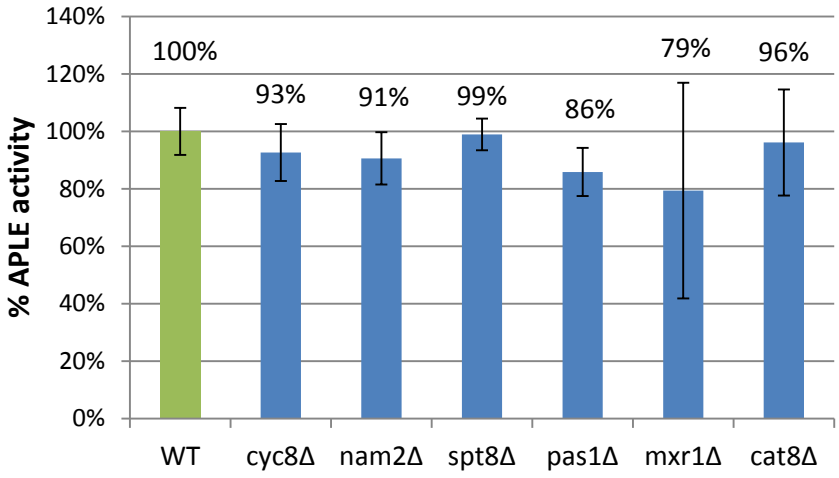


Figure 44: Quantification of secreted APLE, expressed from P_{GAP}, in deep-well plate culture supernatants of targeted knockout strains. Relative APLE activity upon expression from *GAP* promoter. WT control and mutant strains expressing APLE from P_{GAP} were grown on BYPD (2% glucose) in deep-well plates for 35 h before analysis. Esterase activity in supernatants was quantified with pNPA assay. Results represent the mean of one biological experiment with 12 technical replicates per knockout strain.

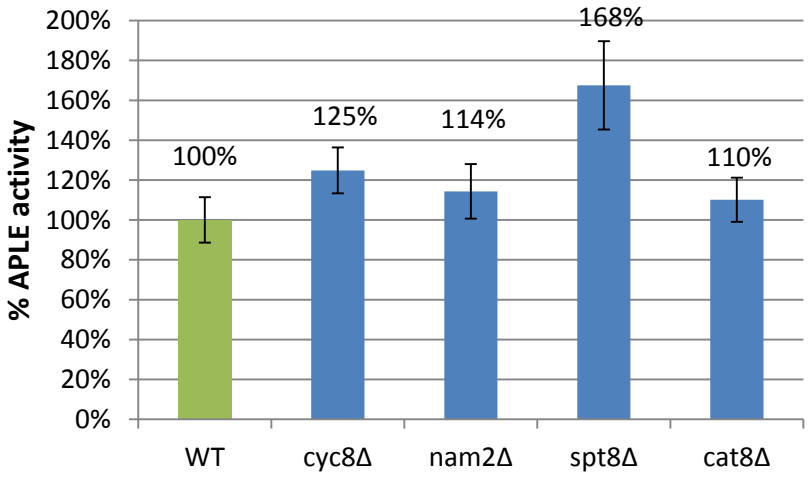


Figure 45: Quantification of secreted APLE, expressed from P_{AOX1}, in deep-well plate culture supernatants of targeted knockout strains. Relative APLE activity upon expression from P_{AOX1}. Strains expressing APLE from P_{AOX1} were grown in BMGY for 32 h and were induced with methanol for 48 h for activity assays as above. Results represent mean of one biological experiment with 12 technical replicate samples per knockout strain. The knockout strains *pas1Δ* and *mxr1Δ* cannot grow on methanol as sole carbon source, as the disrupted genes are involved in peroxisome biogenesis and cellular response to methanol, respectively.

6 List of *P. pastoris* strains transferred to the group's strain collection

Table 4. List of *P. pastoris* strains transferred to the group's strain collection

N°	Strains (+ transformed DNA)	Markers
Expression strains (*used as starting strain for random mutagenesis)		
Cmw01	GS115 + (pGAPZalphaA+MFalpha1-four)	*his4* Zeo^R
Cmw02	GS115 + (pGAPZalphaA+HRP)	*his4* Zeo^R
Cmw03	GS115 + (pGAPZalphaA+APLE WT)	*his4* Zeo^R
Cmw04	GS115 + (pGAPZalphaA+APLE V263D)	*his4* Zeo^R
Cmw05	GS115 + (pPIC9toGAP+MFalpha1-four)	*HIS4*
Cmw06*	GS115 + (pPIC9toGAP+HRP)	*HIS4*
Cmw07*	GS115 + (pPIC9toGAP+APLE WT)	*HIS4*
Cmw08	GS115 + (pPIC9toGAP+APLE V263D)	*HIS4*
Mutant strains from random mutagenesis		
Cmw09	Cmw06	*HIS4* Zeo^R
Cmw10	Cmw06	*HIS4* Zeo^R
Cmw11	Cmw06	*HIS4* Zeo^R
Cmw12	Cmw06	*HIS4* Zeo^R
Cmw13	Cmw06	*HIS4* Zeo^R
Cmw14	Cmw07 A1 *(nam2)*	*HIS4* Zeo^R
Cmw15	Cmw07 A2 (*gtr1*)	*HIS4* Zeo^R
Cmw16	Cmw07 A10 *(erd1)*	*HIS4* Zeo^R
Cmw17	Cmw07 A11 *(cyc8)*	*HIS4* Zeo^R
Cmw18	Cmw07 A13 *(cat8)*	*HIS4* Zeo^R
Cmw19	Cmw07 A17 *(sul1)*	*HIS4* Zeo^R
Cmw20	Cmw07 A18 *(spt8)*	*HIS4* Zeo^R
Targeted knockout strains		
Cmw21	CBS7435 *kep1*	*his4*
Cmw22	CBS7435 *rim101*	*his4*
Cmw23	CBS7435 *sgt2*	*his4*
Cmw24	CBS7435 *kcs1*	*his4*
Cmw25	CBS7435 *nam2*	*his4*
Cmw26	CBS7435 *cyc8*	*his4*
Cmw27	CBS7435 *cat8*	*his4*
Cmw28	CBS7435 *mxr1*	*his4* Zeo^R
Cmw29	CBS7435 *spt8*	*his4*

Cmw30	CBS7435 *pas1 (=pex1)*	*his4* Zeo^{R}
Cmw31	CBS7435 *kep1 lysoxA*	*his4*
Cmw32	CBS7435 *kep1 lysoxA lysoxB*	*HIS4*
	Reference expression strains (Mut^{S})	
Cmw33	CBS7435 + (pGaHSwa-HRP)	*HIS4*
Cmw34	CBS7435 + (pAaHSwa-HRP)	*HIS4*
Cmw35	CBS7435 + (pGaHSwa-APLE)	*HIS4*
Cmw36	CBS7435 + (pAaHSwa-APLE)	*HIS4*
	Expression strains (Mut^{S})	
Cmw37	Cmw21 + (pGaHSwa-HRP)	*HIS4*
Cmw38	Cmw22 + (pGaHSwa-HRP)	*HIS4*
Cmw39	Cmw23 + (pGaHSwa-HRP)	*HIS4*
Cmw40	Cmw24 + (pGaHSwa-HRP)	*HIS4*
Cmw41	Cmw21 + (pAaHSwa-HRP)	*HIS4*
Cmw42	Cmw22 + (pAaHSwa-HRP)	*HIS4*
Cmw43	Cmw23 + (pAaHSwa-HRP)	*HIS4*
Cmw44	Cmw24 + (pAaHSwa-HRP)	*HIS4*
Cmw45	Cmw21 + (pGaHSwa-APLE)	*HIS4*
Cmw46	Cmw22 + (pGaHSwa-APLE)	*HIS4*
Cmw47	Cmw23 + (pGaHSwa-APLE)	*HIS4*
Cmw48	Cmw24 + (pGaHSwa-APLE)	*HIS4*
Cmw49	Cmw21 + (pAaHSwa-APLE)	*HIS4*
Cmw50	Cmw22 + (pAaHSwa-APLE)	*HIS4*
Cmw51	Cmw23 + (pAaHSwa-APLE)	*HIS4*
Cmw52	Cmw24 + (pAaHSwa-APLE)	*HIS4*
Cmw53	CBS7435 + (pGaHSwa-hGH)	*HIS4*
Cmw54	Cmw21 + (pGaHSwa-hGH)	*HIS4*
Cmw55	Cmw22 + (pGaHSwa-hGH)	*HIS4*
Cmw56	Cmw23 + (pGaHSwa-hGH)	*HIS4*
Cmw57	Cmw21 + (pPT4-GAP-EGFP-Arg4-Zeocin)	Zeo^{R}
Cmw58	Cmw22 + (pPT4-GAP-EGFP-Arg4-Zeocin)	Zeo^{R}
Cmw59	Cmw23 + (pPT4-GAP-EGFP-Arg4-Zeocin)	Zeo^{R}
Cmw60	Cmw24 + (pPT4-GAP-EGFP-Arg4-Zeocin)	Zeo^{R}
Cmw61	Cmw25 + (pGaHSwa-APLE)	*HIS4*
Cmw62	Cmw26 + (pGaHSwa-APLE)	*HIS4*
Cmw63	Cmw27 + (pGaHSwa-APLE)	*HIS4*
Cmw64	Cmw28 + (pGaHSwa-APLE)	*HIS4*
Cmw65	Cmw29 + (pGaHSwa-APLE)	*HIS4*
Cmw66	Cmw30 + (pGaHSwa-APLE)	*HIS4*

Cmw67	Cmw25 + (pAaHSwa-APLE)	*HIS4*
Cmw68	Cmw26 + (pAaHSwa-APLE)	*HIS4*
Cmw69	Cmw27 + (pAaHSwa-APLE)	*HIS4*
Cmw70	Cmw29 + (pAaHSwa-APLE)	*HIS4*
	Strains expressing flag-tagged proteins	
Cmw71	Cmw37 + (pPpRSFC_KEP1_C.FLAG)	*HIS4* Zeo^R
Cmw72	Cmw38 + (pPpRSFC_RIM101_C.FLAG)	*HIS4* Zeo^R
Cmw73	Cmw39 + pPpRSFC_SGT2_C.FLAG	*HIS4* Zeo^R
Cmw74	Cmw37 + (pPpRSFC_PGAP_KEP1_C.FLAG)	*HIS4* Zeo^R
Cmw75	Cmw38 + (pPpRSFC_PGAP_RIM101_C.FLAG)	*HIS4* Zeo^R
Cmw76	Cmw39 + pPpRSFC_PGAP_SGT2_C.FLAG	*HIS4* Zeo^R

Further discussion, conclusion and outlook

In the first outline of the project, we planned to generate a library of *P. pastoris* mutants using transposon mutagenesis. Because no transposon native to *P. pastoris* is known, we had to adapt a transposon system from another organism for this yeast. The Hermes transposon from the housefly *Musca domestica* seemed to be an appropriate system, since it had been proven to facilitate mutagenesis in *Schizosaccharomyces pombe* (Evertts et al., 2007; Park et al., 2009). A serious obstacle for its use in *P. pastoris*, however, was that the donor plasmid (Figure 1 in Chapter 5, Supplementary data), carrying the transposon cassette, had to be removed from the cell after the transposition event had occurred. The transposon cassette encodes the antibiotic selection marker, which is used to screen for successful integrations of the cassette into the host genome. This screening is not feasible if the donor plasmid is constantly present in the genome. In *S. pombe*, the donor plasmid encoded a *URA3* marker along with the transposon cassette, and, hence, was removed from the cells by negative selection on 5-fluoroorotic acid (5-FOA). The drug 5-FOA is toxic to cells harboring a functional *URA3* gene, and therefore induces loss of the donor plasmid. In *P. pastoris*, on the other hand, the donor plasmid integrates into the genome and, thus, removal of the encoded antibiotic marker is more troublesome. Moreover, in our hands, selection on 5-FOA seemed to be less efficient in *P. pastoris* than in *Saccharomyes cerevisiae*. We hence tried mutagenesis with the *zeta*-transposon of the yeast *Yarrowia lipolytica*. This transposon, consisting of a selection marker flanked by two inverted *zeta* regions of 401 and 312 bp, was reported to efficiently integrate into the *Y. lipolytica* genome after transformation of the linear DNA fragment (Mauersberger et al., 2001). We soon discovered that the flanking *zeta* regions were not needed for efficient and random integration of the marker cassette into the *P. pastoris* genome. Our newly developed method for random mutagenesis of *P. pastoris*, presented and discussed in Chapters 1 and 2, uses simple linear DNA marker cassettes for mutagenesis. A favorable feature of this method is that the mutagenic cassettes can be readily amplified by PCR. Our method provides an efficient tool for the random mutagenesis of *P. pastoris* and can therefore assist the study of gene function in this yeast.

After a random mutant library has been successfully generated, the next step is the screening for the anticipated mutant phenotype. The design of this screening assay is especially critical, as it can significantly influence the outcome of the experiment. "You get what you screen for", is a phrase often heard in this connection. Indeed, the genes identified in our screening can in many cases be

ascribed to the cell cultivation and screening set-up. Disruption of the gene encoding Hem3, for example, was found to decrease HRP secretion. Depletion of Hem3, an enzyme involved in the synthesis of the heme cofactor, would, most probably, not influence the production of a protein that does not contain a heme group. Similarly, the highly interesting locus *KEP1* could have been missed if the cultivation time of 72 h growth on buffered YPD had been significantly shortened. Figure 2b of Chapter 1 shows that the beneficial effect of this gene knockout becomes more pronounced with longer cultivation times, while it is negligible in the early phase of cell growth. It can be assumed that, among other factors, the choice of promoter, secretion leader sequence and carbon source of the cultivation medium are similarly crucial for the outcome of the screening. Parameters of the screening assay should thus be selected carefully and resemble the actual requirements of the application. On the other hand, the screening should allow fairly high throughput of mutant colonies. The actual design of the screening will, therefore, often be a compromise between functionality and accuracy.

Inappropriate design of the screening strategy might have caused the problems we observed with APLE as a secreted model protein. While results from the screening of HRP-secreting mutant strains turned out to be highly reproducible, we obtained contradictory results for mutants secreting APLE. In contrast to the HRP experiment, where several genes turned up more than once in the screening, genes that came up in the APLE approach were usually single hits (Tables 1 and 2 in Chapter 5, Supplementary data). Only *MON2* was identified twice, but, to our surprise, the knockout of this gene seemed to benefit APLE secretion in one case and to decrease it in the other case. Knockout of the gene *MON2*, encoding an endosomal protein involved in endocytosis and vacuole integrity, had been indicated to increase secretion of recombinant luciferase in *S. cerevisiae* (Kanjou et al., 2007). Similarly confusing was that disruption of the gene *SSM1*, encoding a dihydrouridine synthase related to Dus1, negatively influenced APLE secretion, while, in contrast, *DUS1* disruption obviously benefited HRP production (Table 1 of Chapter 1). Moreover, the results of the initial screening were not evenly reproducible in shake flask and deep-well plate cultivations, as summarized in Table 3 of Chapter 5. This observation could possibly be explained by the nature of the genes identified in the APLE screening. Many of them are involved in regulatory pathways or act as transcription factors themselves. In addition, the mutations suspected to influence APLE secretion appeared to affect cell growth rates to a higher extent than the mutations selected in the HRP screening (compare growth curves of HRP- and APLE-secreting

mutant strains, Figures 16 and 37 of Chapter 5, respectively). Targeted knockouts of *NAM2*, *PAS1*, *CAT8*, *CYC8*, *SPT8* and *MXR1*did not significantly affect the secretion of APLE from P_{GAP}, as shown in Figure 44 of Chapter 5. Similarly, the secretion of HRP from P_{GAP} was unaltered in these knockout strains (data not shown). As integration of the mutagenesis cassette at a certain locus could eventually also effect the expression of neighboring genes, it is conceivable that the wrong genes were selected for targeted knockout. This hypothesis underlines the importance of creating "clean knockouts" of selected genes. Without confirmation by targeted disruption, the observed phenotype could be falsely attributed to a certain gene, while it is indeed caused by altered expression of another gene. Our novel strategy for the generation of targeted knockouts, presented in Chapter 3, greatly enhances the applicability of this approach.

By selectively targeting the genes *KCS1*, *RIM101*, *SGT2* and *KEP1*, we could show that at least knockouts of the latter three genes reproducibly benefited the secretion of HRP. Especially the *KEP1* locus caught our attention, since depletion of the encoded protein also enhanced secretion of the model proteins APLE and hGH (Figure 4 of Chapter 1). Strains engineered to universally increase recombinant protein secretion are of special interest for industrial production processes, and the *kep1Δ* knockout strain could have the potential to be such a universal production host. Surprisingly, the secretion of *B. subtilis* levanase did not seem to be affected in this knockout strain (Figure 28 of Chapter 5, Supplementary data). We aimed to test if increased leakage from the cell accounts for the enhanced secretion rates of certain model proteins in our knockout strains. Therefore, we expressed EGFP intracellularly, and observed fluorescence levels inside the cells and in the culture supernatant (Figures 30 and 31, Chapter 5). Interestingly, the knockout strains *sgt2Δ* and *rim101Δ* did neither affect intracellular nor extracellular levels of EGFP, while the results for *kep1Δ* and *kcs1Δ* were more controversial. The knockout strain *kep1Δ* seemed to increase EGFP both intra- and extracellularly, while *kcs1Δ* only affected intracellular fluorescence levels. However, the experiments would have to be repeated to confirm these initial results.

Possible alterations in the cell wall of *kep1Δ* knockout strains were indicated by the release of the potentially cell surface-attached proteins Flo9 and PPLO to the culture supernatant. We can only speculate about the role of PPLO in *P. pastoris*. Lysyl oxidases were proposed to be involved in the utilization of primary amines as the sole source of nitrogen or carbon in microorganisms, but were also shown to possess crosslinking activity (Duff et al. 2003). In any case, their function is not essential for cell survival, as knockout of both genes encoding PPLO was straightforward

using our novel gene targeting strategy. We also attempted to knockout the gene encoding Flo9, but were not successful in obtaining transformants that displayed the Zeocin resistance encoded on the knockout vector. As discussed in Chapter 3, using Zeocin in the selection process can pose a problem when the anticipated gene knockout decreases cell viability. As the Flo proteins were not reported to be essential in *S. cerevisiae*, it could be feasible to repeat the knockout experiment with a biosynthetic selection marker instead.

To conclude, our work has demonstrated that random mutagenesis of *P. pastoris* provides a potent tool to characterize molecular processes like secretion. Combining this method with our novel strategy for generating targeted knockouts in this yeast enabled us to identify a set of genes that had not been associated with recombinant protein secretion before. Future work will investigate how the most promising hits Rim101, Sgt2 and Kep1 influence recombinant protein secretion on a molecular basis. Furthermore, it would be interesting to test the generated knockout strains for the production of other recombinant proteins, and to combine multiple gene disruptions in one strain to check for combinatorial effects. Most importantly, a detailed characterization of the orphan gene *KEP1* could uncover unique features of the industrially important expression host *P. pastoris*.

References for introduction and conclusion

Balch WE, Dunphy WG, Braell WA, Rothman JE (1984) Reconstitution of the transport of protein between successive compartments of the Golgi measured by the coupled incorporation of N-acetylglucosamine. Cell 39:405-416

Barlowe C, Orci L, Yeung T, Hosobuchi M, Hamamoto S, Salama N, Rexach MF, Ravazzola M, Amherdt M, Schekman R (1994) COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. Cell 77:895–907.

Bernales S, Papa FR, Walter P (2006) Intracellular signaling by the unfolded protein response. Annu Rev Cell Dev Biol 22:487–508

Bevan A, Brenner C, Fuller RS (1998) Quantitative assessment of enzyme specificity in vivo: P2 recognition by Kex2 protease defined in a genetic system. Proc Natl Acad Sci U S A 95:10384–10389.

Block MR, Glick BS, Wilcox CA, Wieland FT, Rothman JE (1988) Purification of an N-ethylmaleimide-sensitive protein catalyzing vesicular transport. Proc Natl Acad Sci USA 85:7852-7856

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

Brake AJ, Julius DJ, Thorner J (1983) A functional prepro-alpha-factor gene in *Saccharomyces* yeasts can contain three, four, or five repeats of the mature pheromone sequence. Mol Cell Biol 3:1440–50.

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 81:4642–6.

Butz JA, Niebauer RT, Robinson AS (2003) Co-expression of molecular chaperones does not improve the heterologous expression of mammalian G-protein coupled receptor expression in yeast. Biotechnol Bioeng 84:292–304

Chan RK, Otte CA (1982) Physiological characterization of *Saccharomyces cerevisiae* mutants supersensitive to G1 arrest by a factor and alpha factor pheromones. Mol Cell Biol 2:21–9.

Childs BRE, Bardsley WG (1975) The Steady-State Kinetics of Peroxidase with 2,2'-Azino-di(3-ethylbenzthiazoline-6-sulphonic acid) as Chromogen. Biochem J 145:93–103.

Clary DO, Griff IC, Rothman JE (1990) SNAPs, a family of NSF attachment proteins involved in intracellular membrane fusion in animals and yeast. Cell 61:709-721

Connerly PL, Esaki M, Montegna EA, Strongin DE, Levi S, Soderholm J, Glick BS (2005) Sec16 is a determinant of transitional ER organization. Curr Biol 15:1439–47. doi: 10.1016/j.cub.2005.06.065

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

Dakin, HD (1903) The hydrolysis of optically inactive esters by means of enzymes: Part I. The action of lipase upon esters of mandelic acid. The resolution of inactive mandelic acid. J. Physiol. 30, 253–263.

Damasceno LM, Anderson KA, Ritter G, Cregg JM, Old LJ, Batt CA (2007) Cooverexpression of chaperones for enhanced secretion of a single-chain antibody fragment in *Pichia pastoris.* Appl Microbiol Biotechnol 74:381–389

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

De Maria PD, Garcia-Burgos CA, Bargeman G, van Gemert RW (2007) Pig liver esterase (ple) as biocatalyst in organic synthesis: from nature to cloning and to practical applications. Synthesis-Stuttgart 10:1439–1452. doi:10.1055/s-2007-966024

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 27:561–6. doi: 10.1038/nbt.1544

Dunn WA, Cregg JM, Kiel JAKW, van der Klei IJ, Oku M, Sakai Y, Sibirny AA, Stasyk OV, Veenhuis M (2005) Pexophagy: The Selective Autophagy of Peroxisomes. Autophagy 1:75–83.

Eakle KA, Bernstein M, Emr SD (1988) Characterization of a component of the yeast secretion machinery: identification of the *SEC18* gene product. Mol Cell Biol 8:4098-4109

Evertts AG, Plymire C, Craig NL, Levin HL (2007) The hermes transposon of *Musca domestica* is an efficient tool for the mutagenesis of *Schizosaccharomyces pombe*. Genetics 177:2519–23. doi: 10.1534/genetics.107.081075

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 73:6499–507. doi: 10.1128/AEM.01196-07

Glick BS, Nakano A (2009) Membrane traffic within the Golgi apparatus. Annu Rev Cell Dev Biol 25:113–32. doi: 10.1146/annurev.cellbio.24.110707.175421

Griff IC, Schekman R, Rothman JE, Kaiser CA (1992) The yeast *SEC17* gene product is functionally equivalent to mammalian alpha-SNAP protein. J Biol Chem 267:12106-12115

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 Fact 9:49

Hansen W, Garcia PD, Walter P (1986) In vitro protein translocation across the yeast endoplasmic reticulum: ATP-dependent posttranslational translocation of the prepro-alpha-factor, Cell 45:397–406

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

Hoyle MC (1977) High resolution of peroxidase-indoleacetic acid oxidase isoenzymes from horseradish by isoelectric focusing, Plant Physiol. 60:787–793

Hummel A, Brüsehaber E, Böttcher D, Trauthwein H, Doderer K, Bornscheuer UT (2007) Isoenzymes of pig-liver esterase reveal striking differences in enantioselectivities. Angew Chem Int Ed Engl 46:8492–4. doi: 10.1002/anie.200703256

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, Aryasomayajula D, Sinha J, Meagher MM (2006) Enhancement of protein secretion in *Pichia pastoris* by overexpression of protein disulfide isomerase. Biotechnol Bioeng 93:771–778

Jensen D, Schekman R (2011) COPII-mediated vesicle formation at a glance. J Cell Sci 124:1–4. doi: 10.1242/jcs.069773

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

Joshi S, Agrawal G, Subramani S (2012) Phosphorylation-dependent Pex11p and Fis1p interaction regulates peroxisome division. Mol Biol Cell 23:1307–15. doi: 10.1091/mbc.E11-09-0782

Kanjou N, Nagao A, Ohmiya Y, Ohgiya S (2007) Yeast mutant with efficient secretion identified by a novel secretory reporter, Cluc. Biochem Biophys Res Commun 358:429–434

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 95:104–12. doi: 10.1016/j.pep.2013.12.003

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 154:312–20. doi: 10.1016/j.jbiotec.2011.04.014

Kumar A, Seringhaus M, Biery MC, Sarnovsky RJ, Umansky L, Piccirillo S, Heidtman M, Cheung K-H, Dobry CJ, Gerstein MB, Craig NL, Snyder M (2004) Large-Scale Mutagenesis of the Yeast Genome Using a Tn7-Derived Multipurpose Transposon. Genome Res 1975–1986. doi: 10.1101/gr.2875304

Lange S, Musidlowska A, Schmidt-Dannert C, Schmitt J, Bornscheuer UT (2001) Cloning, functional expression and characterization of recombinant pig liver esterase. Chembiochem 2, 576–582

Levi SK, Bhattacharyya D, Strack RL, Austin JR, Glick BS (2010) The yeast GRASP Grh1 colocalizes with COPII and is dispensable for organizing the secretory pathway. Traffic 11:1168–79. doi: 10.1111/j.1600-0854.2010.01089.x

Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the *Pichia pastoris* expression system. Yeast 22:249–70. doi: 10.1002/yea.1208

Malhotra V, Orci L, Glick BS, Block MR, Rothman JE (1988) Role of an N-ethylmaleimide-sensitive transport component in promoting fusion of transport vesicles with cisternae of the Golgi stack. Cell 54:221-227

Manney TR (1983) Expression of the *BAR1* Gene in *Saccharomyces cerevisiae*: Induction by the alpha Mating Pheromone of an Activity Associated with a Secreted Protein. J Bacteriol 155:291–301.

Matlack KE, Misselwitz B, Plath K, Rapoport TA (1999) BiP acts as a molecular ratchet during posttranslational transport of prepro-alpha factor across the ER membrane. Cell 97:553–64.

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

Mauersberger S, Wang H, Gaillardin C, Barth G, Nicaud JM (2001) Insertional mutagenesis in the n-alkane-assimilating yeast *Yarrowia lipolytica*: generation of tagged mutations in genes involved in hydrophobic substrate utilization. J Bacteriol 183:5102–5109. doi: 10.1128/JB.183.17.5102

McClintock B (1950) The origin and behavior of mutable loci in Maize. Proc. Natl. Acad. Sci. USA 36, 344–345

Mogelsvang S, Gomez-Ospina N, Soderholm J, Glick BS, Staehelin LA (2003) Tomographic evidence for continuous turnover of Golgi cisternae in *Pichia pastoris*. Mol Biol Cell 14:2277.

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

Musidlowska A, Lange S, Bornscheuer UT (2001) By overexpression in the yeast *Pichia pastoris* to enhanced enantioselectivity: new aspects in the application of pig liver esterase. Angew. Chem. Int. Ed. 40, 2851–2853

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.

Novick P, Field C, Schekman R (1980) Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway. Cell 21:205–15.

Palade GE (1974) Intracellular aspects of the process of protein secretion. Nobel Lecture

Papanikou E, Glick BS (2009) The yeast Golgi apparatus: insights and mysteries. FEBS Lett 583:3746–51. doi: 10.1016/j.febslet.2009.10.072

Park JM, Evertts AG, Levin HL (2009) The Hermes transposon of *Musca domestica* and its use as a mutagen of *Schizosaccharomyces pombe*. Methods 49:243–7. doi: 10.1016/j.ymeth.2009.05.004

Planche LA (1810) Note sur la sophistication de la résine de jalap et sur les moyens de la reconnaître, Bull. Pharm. 2:578–580.

Richard M, Quijano RR, Bezzate S, Bordon-Pallier F, Gaillardin C (2001) Tagging Morphogenetic Genes by Insertional Mutagenesis in the Yeast *Yarrowia lipolytica*. J Bacteriol 183:3098–3107. doi: 10.1128/JB.183.10.3098

Rossanese OW, Soderholm J, Bevis BJ, Sears IB, Connor JO, Williamson EK, Glick BS (1999) Golgi Structure Correlates with Transitional Endoplasmic Reticulum Organization in *Pichia pastoris* and *Saccharomyces cerevisiae*. Cell 145:69–81.

Ross-Macdonald P, Coelho PS, Roemer T, Agarwal S, Kumar A, Jansen R, Cheung K, Sheehan A, Symoniatis D, Umansky L, Heidtman M, Nelson FK, Iwasaki H, Hager K, Gerstein M, Miller P, Roeder GS, Synder M (1999) Large-scale analysis of the yeast genome by transposon tagging and gene disruption. Nature 402:413–418.

Rothblatt JA, Webb JR, Ammerer G, Meyer DI (1987) Secretion in yeast: structural features influencing the post-translational translocation of prepro-alpha-factor in vitro. EMBO J 6:3455–63.

Ryan BJ, Carolan N, O'Fágáin C (2006) Horseradish and soybean peroxidases: comparable tools for alternative niches? Trends Biotechnol 24:355–63. doi: 10.1016/j.tibtech.2006.06.007

Schmidt FR (2004) Recombinant expression systems in the pharmaceutical industry. Appl Microbiol Biotechnol 65:363–72. doi: 10.1007/s00253-004-1656-9

Seifert HS, Chen EY, So M, Heffron F (1986) Shuttle mutagenesis : A method of transposon mutagenesis for *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 83:735–739.

Seringhaus M, Kumar A, Hartigan J, Snyder M, Gerstein M (2006) Genomic analysis of insertion behavior and target specificity of mini-Tn7 and Tn3 transposons in *Saccharomyces cerevisiae*. Nucleic Acids Res 34:e57. doi: 10.1093/nar/gkl184

Söllner T, Whiteheart SW, Brunner M, Erdjument-Bromage H, Geromanos S, Tempst P, Rothman JE (1993) SNAP receptors implicated in vesicle targeting and fusion. Nature 362:318-324

Suda Y, Nakano A (2012) The yeast Golgi apparatus. Traffic 13:505–10. doi: 10.1111/j.1600-0854.2011.01316.x

Veitch NC (2004) Horseradish peroxidase: a modern view of a classic enzyme. Phytochemistry 65:249–259. doi: 10.1016/j.phytochem.2003.10.022

Waterham HR, Digan ME, Koutz PJ, Lair S V, 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

Wilson DW, Wilcox CA, Flynn GC, Chen E, Kuang WJ, Henzel WJ, Block MR, Ullrich A, Rothman JE (1989) A fusion protein required for vesicle-mediated transport in both mammalian cells and yeast. Nature 339:355-359

Wu J, Kaufman RJ (2006) From acute ER stress to physiological roles of the unfolded protein response. Cell Death Differ 13:374–384

Wu M, Liu W, Yang G, Yu D, Lin D, Sun H, Chen S (2014) Engineering of a *Pichia pastoris* Expression System for High-Level Secretion of HSA/GH Fusion Protein. Appl Biochem Biotechnol. doi: 10.1007/s12010-013-0688-y

Zhang W, Zhao HL, Xue C, Xiong X, Yao XQ, Li X, Chen H, Liu Z (2006) Enhanced Secretion of Heterologous Proteins in *Pichia pastoris* Following Overexpression of *Saccharomyces cerevisiae* Chaperone Proteins. Biotechnol Prog 22:1090–1095.