DIPLOMARBEIT

Titel der Diplomarbeit

„Establishment and characterization of a novel secretion system in the methylotrophic yeast *Pichia pastoris*"

Verfasserin

Silvia Heiss

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag.rer.nat.)

Wien, 2011

For my Grandmother

Table of Contents

1 Abstract

The methylotrophic yeast *Pichia pastoris* is successfully used for recombinant protein production. The application of strong promoters (e.g. the glyceraldehyde-3-phosphate dehydrogenase promoter P_{GAP}, and the alcohol oxidase promoter, P_{AOX}) and efficient secretion signals (alpha mating factor secretion leader; MFα1pp) as well as proper folding and post-translational modifications are beneficial.

The extracellular protein X (Epx1) was detected as unknown, strongly secreted host protein in supernatants. As one of 20 endogenous secreted proteins in *P. pastoris*, little is known about Epx1 except C-terminal alignment with the SCP-like (secretory cysteine rich) extracellular protein domain. Therefore, the practicability as novel secretion system for enhancing recombinant protein expression was tested. The promoter sequence (1,000 bp upstream of ATG) was used to analyze intracellular eGFP expression. The Epx1 promoter, tested at three different carbon sources, was weaker in expressing eGFP compared to P_{GAP}. Furthermore, secretion and processing mediated by Epx1 secretion leader was studied on three structurally and functionally different recombinant proteins: pTRP (porcine trypsinogen), HSA (human serum albumin), and eGFP (enhanced green fluorescent protein).

The full length pre-pro Epx1 secretion leader partially showed processing difficulties. 21 amino acids at the pro-sequence remained N-terminally attached due to unexpected cleavage at the internal dibasic motif. By shortening the Epx1 leader, pTRP secretion was increased, and eGFP was secreted without N-terminal overhangs. Summarizing, we established a novel and more efficient secretion leader than the MFα1pp, by shortening the Epx1 pro-sequence.

2 Introduction

2.1 The methylotrophic yeast *Pichia pastoris*

The methylotrophic yeast *Pichia pastoris* is widely used as a host system for recombinant protein production. It is also a commonly used model organism for basic research of peroxisome and secretory organelles biosynthesis. Furthermore, it has come into focus for the production of glycol-proteins with human-like N-glycan structures (Hamilton et al., 2003), as well as for several metabolites and recombinant proteins.

Recently, *P. pastoris* has been reclassified into a new genus, namely *Komagataella* (Yamada et al., 1995), and divided into three species, *K. pastoris*, *K. phaffii*, and *K. pseudopastoris* (Kurtzman, 2005). The strains GS115 and X-33 have been made available by Invitrogen and belong to the species *K. phaffii*. Apart from that, other strains belonging to either *K. pastoris* or *K. phaffii* are freely used by researchers. The reference strain for all available *P. pastoris* strains (DSMZ 70382) belongs to the type species *K. pastoris* (Figure 1). Moreover, in accordance with published literature, all strains are further named *P. pastoris*, standing for the entire genus *Komagataella*.
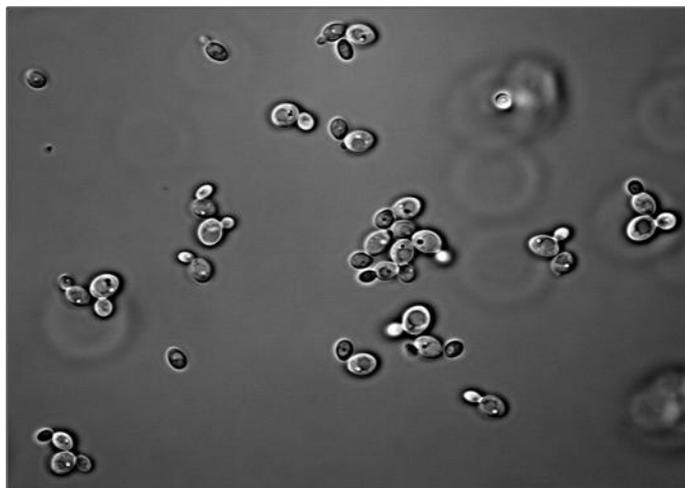
Figure 1: *Pichia pastoris* (by V. Puxbaum)

At present, the genomes of two *P. pastoris* strains (DSMZ 70382 and GS 115) have been fully sequenced (De Schutter et al., 2009; Mattanovich et al., 2009b). Therefrom, two genome browsers were set up (Mattanovich et al., 2009a). Until then, most data on genetic and physiological

background for strain- and process design relied on analogies to other, well studied yeasts like *Saccharomyces cerevisiae*. Accordingly, *P. pastoris* gene names follow mainly the format established for *S. cerevisiae*.

Generalizing, the number of functionally annotated genes (9.4 Mb; 5.450 ORFs) is comparable to other yeasts, most metabolic enzymes are present in single copies, and the number of actually secreted proteins is low, thus making secretory production of heterologous proteins attractive.

2.2 Recombinant protein production in *Pichia pastoris*

The use of *P. pastoris* as a cellular host for recombinant protein production steadily increases. It is genetically easy manipulated and cultured, and can reach high cell densities (> 130 g l^{-1} dry cell weight) on methanol and glucose (Cereghino & Cregg, 2000).

Equally important, as eukaryote, it provides the potential for producing soluble, correctly folded proteins, which have undergone post-translational modifications, such as glycosylation (O- and N-linked; less over-glycosylating than *S. cerevisiae*), disulfide bridge formation, as well as processing of signal sequences.

For intracellular expression, the amino-terminal methionine residue is cleaved off, unlike proteins expressed in *E. coli,* or the protein can also be acetylated and specific amino acid residues are likely to be phosphorylated, generating phospho-proteins (Daly & Hearn, 2005) without limitations and bottlenecks obtained by the secretory pathway. Summing up, recombinant proteins can either be secreted, thus generating a more desirable amount of protein and simultaneously mediating the first step of purification, or intracellular expressed; this choice will depend on how the protein is being expressed in its native system.

Transformation of the haploid, homothallic *P. pastoris* host with recombinant DNA is mediated either by integrative plasmids or by autonomously replicating plasmids. Directed integration or replacement requires homology of the introduced DNA with a chromosomal locus. Multiple integrations are often obtained on purpose (Porro et al., 2005).

Integration only requires restriction at a unique site, homologous to the *P. pastoris* genome. Transformation is mediated by electroporation and usually results in genetically stable transformants with high transcription rates (Daly & Hearn, 2005).

Most vectors are hybrids between bacterial and yeast sequences, possessing an origin of replication for *E. coli* and selection markers. The plasmids also contain a multiple cloning site (MCS). Both auxotrophic markers, such as the functional histidine dehydrogenase gene (HIS4; first described by Cregg and colleagues (Cregg et al., 1985)), which can be used in histidine dehydrogenase defective GS115, as well as dominant markers exist. The bacterial kanamycin resistance gene also confers resistance to the eukaryotic antibiotic G418 and the bacterial *Sh ble* gene applies resistance to the antibiotic Zeocin™, also appropriate for yeast (Cereghino & Cregg, 2000).

Further benefits of the *P. pastoris* system are strong inducible and constitutive promoter systems. One important issue of recombinant protein production is the transcription efficiency; therefore, the choice of the promoter is crucial. The number of available promoters is limited, however, and mainly comprises the methanol inducible alcohol oxidase 1 (AOX1) promoter and the constitutive glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter.

The *P. pastoris* alcohol oxidase and two other methanol regulated genes have first been isolated by Ellis and co-workers (Ellis et al., 1985). Alcohol oxidase is the key enzyme in the methanol utilization pathway; specific for methylotrophic yeasts. It is encoded by two genes, *AOX1* and *AOX2*, and functionally and structurally characterized (Koutz et al., 1989) as well as reviewed several times. The AOX promoters are tightly regulated by a carbon source-dependent repression/induction mechanism; showing full repression during growth on glucose or glycerol excess conditions, and maximal induction during growth on methanol.

In contrast, the glyceraldehyde-3-phosphate dehydrogenase promoter (P_{GAP}) is constitutively expressed, although its strength varies depending on

the carbon source used for cell growth. This offers an attractive alternative to P_{AOX1} on glucose, especially if induction by methanol may be inappropriate or inconvenient; simultaneously increasing cell viability. Furthermore, the activity of P_{GAP} in glucose-grown shake-flask cultures is stronger than P_{AOX1} in methanol-grown shake-flask cultures (Waterham et al., 1997) and slightly lower on methanol (Stadlmayr et al., 2010).

Identifying alternative promoters is a promising and important tool especially for co-expressing recombinant proteins. Additionally, through regulation of transcription, promoters with varying activities increase host engineering possibilities. Auxiliary constitutive promoters include P_{TEF1} (translation elongation factor 1-α) (Ahn et al., 2007), P_{PGK1} (glycolytic enzyme 3-phosphoglycerate kinase) (de Almeida, de Moraes & Torres, 2005), and P_{YPT1}, a GTPase involved in secretion (Sears et al., 1998). The formaldehyde dehydrogenase promoter (P_{FLD1}), on the other hand, is inducible by both methanol and methylamine (Shen et al., 1998), whereas the isocitrate lyase promoter (P_{ICL1}) is inducible by glucose or ethanol (Menendez, Valdes & Cabrera, 2003). Recently, the auspicious promoter of a protein involved in synthesis of the thiamine precursor hydroxymethylpyrimidine (P_{THI11}) has been identified by Stadlmayr and co-workers (Stadlmayr et al., 2010). It appears to be the first carbon and nitrogen source independent adjustable promoter. A major focus of several researchers is therefore the identification of new applicable promoter sequences.

2.3   The secretory pathway, protein folding and secretory capacity of *P. pastoris*

Heterologous proteins as well as endogenous proteins, dedicated to secretion, enter the endoplasmatic reticulum (ER). In *P. pastoris*, 88 proteins were predicted as secreted proteins and thereof, 55 had been functionally annotated out of 172 ORFs, which had been identified to encode proteins entering the secretory pathway. However, experimental determination (chemostat culture on glucose as limiting carbon source)

revealed only 20 actually secreted proteins as shown in Figure 2 (Mattanovich et al., 2009b).

The first step in targeting proteins co-translationally to the ER is mediated through an N-terminal signal sequence (discussed detailed in the next chapter), which is recognized by the signal recognition particle (SRP) after emerging from the nascent protein. The tightly bound SRP targets the translation machinery to the cytosolic side of the ER membrane (Walter & Johnson, 1994). There, it interacts with the SRP-receptor, a component of the hetero-trimeric Sec61/62/63 translocon complex with GTPase activity, thus mediating direct transport through the ER membrane into the lumen after a brief elongation arrest.

The ER lumenal signal peptidase, which is attached to the translocon on the lumenal side, recognizes and cleaves off the signal peptide. Secretory proteins, however, require a second sequence for targeting the polypeptide further, generally termed as pro-peptide. Nevertheless, folding starts as soon as the polypeptide emerges from the translocon.

ER resident folding factors assist in folding and quality control, such as

Figure 2: Categorization of *P. pastoris* secretome. A) predicted and B) detected secretome based on GO terms (intracellular proteins due to lysated cells). Proteins without functional annotation are marked as „unknown". (Mattanovich et al., 2009)

the heat shock chaperones (Hsp70s) BiP as well as GRP170 and their yeast equivalents Kar2p and Lhs1p (Gething & Sambrook, 1992). They sustain ER residency by a C-terminal retrieval sequence (HDEL in yeast), bind to hydrophobic sequences on nascent proteins and maintain vectorial protein translocation through an ATPase activity, supported by J-domain containing co-chaperones (Hsp40; Sec63p, Scj1p and Jem1p) and the nucleotide exchange factor Sil1p in yeast (van Anken & Braakman, 2005).

Perhaps the most distinctive feature of protein folding in the ER is the formation of disulfide bonds. Protein disulfide isomerase (PDI, a thiol oxidoreductase) catalyzes thiol-oxidation, reduction or isomerization. In yeasts, reduced Pdi1p is oxidized by Ero1p (ER oxidoreductase). Additionally, glutathione serves as reducing buffer that balances Ero1p activity (Tu & Weissman, 2004). Altogether, the list of PDI-like family members and other involved proteins has been growing steadily, hence enabling no universal folding strategy, but explaining the variety in speed and efficiency of folding (Kimura et al., 2005). Furthermore, it has been shown that BiP and Pdi1p act synergistically in the *in vitro* folding of a denatured and reduced Fab fragment, suggesting that BiP keeps the antibody chains accessible for Pdi1p (Mayer et al., 2000).

Apart from that, glycosylation can be added to target proteins; O-oligosaccharides (in the Golgi apparatus) or N-oligomannosyl-glycans (in the ER and Golgi apparatus). Despite the differences of glycosylation, there is evidence for a consensus recognition sequence for N-glycosylation in all eykaryotes, i.e. the sequon Asn-Xaa-Ser/Thr (Xaa any amino acid except proline) at the protein N-terminus (Macauley-Patrick et al., 2005). Glycans lead to more hydrophilic proteins and moreover, acting as a quality control mechanism even though they are trimmed and modified shortly after addition (van Anken & Braakman, 2005).

ER clients exit via COPII coated vesicles from the entire ER, if quality control standards are implemented. In yeasts, the membrane protein Erv29 determines that the soluble ER clients MFα1 pheromone and carboxypeptidase Y (CPY) are packaged and sent to the Golgi in accordance to hydrophobic sequences within the pro-sequence (Otte & Barlowe, 2004).

On the other hand, misfolded ER substrates are retro-translocated to the cytosol and become subject to ubiquitination and subsequent proteasomal degradation; a process called ER associated degradation (ERAD). If the folding load increases or productive folding is hindered, the unfolded protein response (UPR) pathway is switched on; sensed by the endonuclease Ire1p. After the Ire1p-mediated splicing of HAC1 mRNA,

Hac1p[i] acts as transcription factor leading to up-regulation of UPR target genes, such as chaperones (Patil & Walter, 2001).

In the Golgi apparatus, precursor processing and sorting mechanisms take place, concluding in targeting the proteins to the plasma membrane, endosomal compartments, the vacuole or via retrograde transport to the ER (Figure 3).

The fungal Kex2 protease (a subtilisin-like serine proteinase) is located in the late Golgi network as well as in the endocytic, pre-vacuolar compartment. It acts during pheromone maturation at lysine-arginine motifs, extensively studied in processing prepro α-factor of *S. cerevisiae* and killer toxin K28. Kex2p as well as the mammalian equivalent furin undergo autocatalytic activation at an analogous site having the sequence Lys-Arg (Leduc et al., 1992; Wilcox & Fuller, 1991).

Several studies have been carried out in *S. cerevisiae kex2* mutants, unraveling severe drawbacks, but also observing cellular backup mechanisms. Especially furin and aspartyl-like proteases, such as yapsins, a family of glycosylphosphatidylinositol (GPI) anchored aspartic proteases (discussed below), show occasional processing at sites consisting of single basic residues (Steiner et al., 1992) and partial processing coverage within *kex2* mutants (Bourbonnais et al., 1991).

Generally, the preferred processing site of subtilisin-like serine proteinases depends on the amino acids surrounding the processing site (Lys-Arg or less frequently Arg-Arg, Arg-X-X-Arg or Arg-X-Lys, succeeded by negatively charged or small residues in the P1´ [C-terminal right after the processing side], P2´ and P4´ position) as well as the three dimensional structure (Bader, Krauke & Hube, 2008). Glu-Ala respectively Asp-Ala repeats C-terminal of the cleavage site have been known already to enhance Kex2p activity. Additionally, they ensure that these precursor portions will be exposed on the outer surface (Brake, Julius & Thorner, 1983).

Those repeats improve correctly processed proteins by preventing steric hindrance of the Kex2p cleavage site and are removed by the

diaminopeptidase (Ste13), also localized in the late Golgi. For high yield recombinant protein production secreted by prepro α-factor, Ste13p might be unable to cope with processing, leading to a protein fraction with Glu-Ala amino terminal extensions. Noteworthy, those extensions on recombinant proteins may also create dominant immunogenic sites.

Two general pathways have been reported for protein secretion from the Golgi (Figure 3): a direct route and one via endosomes. Sorting to the endosome can follow targeting to the late endosome, which is also referred as multi-vesicular body (MVC) or pre-vacuolar compartment (PVC) via CPY pathway (soluble vacuolar hydrolase carboxypeptidase Y) or targeting the early endosomes. A third route directs the vacuole via alkaline phosphatase (ALP) pathway, named after its most studied cargo protein (Bowers & Stevens, 2005; Conibear & Stevens, 1995)
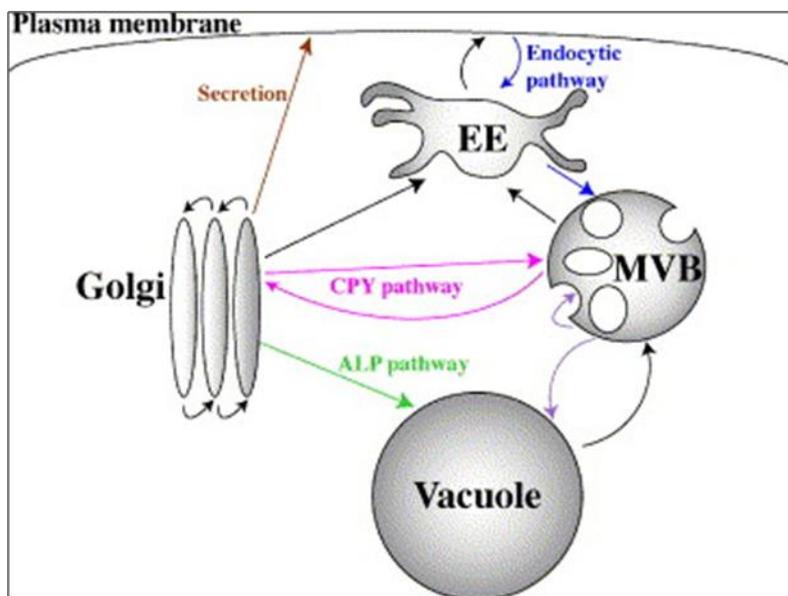
Figure 3: Protein trafficking pathways from the late Golgi to the vacuole in *S. cerevisiae*. Proteins are directly secreted, targeted to early endosomes (EE), to the multivesicular body (MVB) or to the vacuole (Bowers & Stevens, 2005).

In *S. cerevisiae*, Ste13p and Kex2p are localized within the same compartment (Bryant & Boyd, 1993) and have been reported to spend a proportion of their time in different retrieval or recycling pathways between the late Golgi network and the PVC. Golgi retention is guaranteed by a late Golgi localization sequence which is recognized by receptors (e.g. Vps10p and Grd19p) (Bowers & Stevens, 2005; Voos & Stevens, 1998).

It was also reported that Kex2p itself (soluble or Golgi persistent) takes part as a saturable limiting component in the secretory and sorting pathway. As a result, secretion of insulin was increased after substituting dibasic sites, as well as by *vps* and *kex2* deletions, suggesting endoproteolytic processing and sorting of Kex2p being the main cause for insulin retention and mislocalization to the vacuole (Zhang et al., 2001).

Vacuolar protein sorting has been studied extensively in *S. cerevisiae*, unraveling the role of more than 70 proteins involved in sorting, directing and secretion (61 currently known *VPS* genes). Additionally, the yeast vacuole is involved in degradation of plasma membrane-, cytosol- or organelle-derived proteins, ion storage, homeostasis maintenance and adaption to several stress factors (Li & Kane, 2009).

Protein secretion is mediated through post-Golgi secretory vesicles that dock and fuse with the plasma membrane through interaction with v-SNAREs and t-SNAREs. The cargo is then released outside after fusion of exosomes with the plasma membrane by help of the exocyst, the cytoskeleton and Rab GTPases (Burchfield et al., 2010).

Misfolded proteins are translocated from the ER to the cytosol. However, some proteins are reported to enter the Golgi before they retrieve ERAD, and have been shown to enter the vacuole, using autophagic pathways via the cytosol. Some misfolded proteins that evade ER quality control or lack a further targeting sequence are transported to the vacuole using the classical secretory pathway, pointing out the vacuole system as a prominent counterpart in UPR and stress response (Kawaguchi, Hsu & Ng, 2010; Li & Kane, 2009; Roberts, Nothwehr & Stevens, 1992). In plants, BiP was suggested to carry an additional vacuolar sorting signal, which enables an adaptor function between ill-defined hydrophobic regions of misfolded proteins and the sorting machinery, especially Vps10p, which recognizes hydrophobic regions in proteins subjected to vacuolar targeting (Pimpl et al., 2006).

Internal stress has been identified as ratelimiting for secretion of recombinant proteins leading to several consequences, such as protein misdirection. Hohenblum and colleagues (2004) reported that human trypsinogen intended for secretion accumulates in the cytosol when expressed under control of P_{GAP} on glucose, compared to methanol using P_{AOX1}, where accumulation was only shown in the ER. Intracellular expression of heterologous Hnl (hydroxynitrile lyase) resulted in 50 % of the total cytosolic proteins in *S. cerevisiae*, whereas secretion by addition of signal leaders resulted in accumulation of Hnl at the plasma membrane (Hasslacher et al., 1997). The stress in recombinant protein producing yeast has been reviewed in detail by Gasser and colleagues (2008), discussing several topics, but also noting the phenomenon of variability within populations and therefore the existence of sub-populations with quite distinct features influencing the outcomes (Mattanovich et al., 2004).

Additional factors that influence protein production include site and mode of integration of foreign sequences, mRNA 5´- and 3´ untranslated regions (UTR), translational start codon context, A+T composition of cDNA, nature of secretion signal, endogenous protease activity, host strain physiology, yeast codon bias (Sreekrishna et al., 1997) as well as media and growth conditions (Maurer et al., 2006). Codons that are common in other species, such as mammals, may be rare in *P. pastoris*. Altering the codon bias and therefore increasing the GC content within the DNA can be used to improve expression levels (Daly & Hearn, 2005). The recombinant proteins, as well as α-MF prepro leader sequence used in this study, therefore were codon optimized for *P. pastoris*.

2.4 Selection of secretion signals in the *Pichia pastoris* expression system

Signal sequences follow a common structural motif. The preregion typically consists of a positively charged N-terminus, followed by a hydrophobic middle sequence and a polar C-terminus. It targets the nascent

polypeptide to the ER via SRP dependent, SRP independent or both pathways, which is regulated by the hydrophobicity of the core sequence (Ng, Brown & Walter, 1996). In this context, increasing hydrophobicity was reported to augment protein secretion of human lysozyme in *S. cerevisiae* (Yamamoto et al., 1987), however, failed to improve secretion of chicken lysozyme in *P. pastoris* (Oka et al., 1999). The signal peptidase cleavage site is most often followed by three hydrophobic amino acids (Ala-Phe-Val), preceded by a basic residue, determining the pro-region of varying length and amino acid composition.

2.4.1 *Saccharomyces cerevisiae* mating factor alpha

In *S. cerevisiae*, the α-factor structural genes MFα1 and MFα2 and their role in mating have been studied extensively (Brake et al., 1983; Brake et al., 1984; Kurjan, 1985). The mating factor alpha gene 1 (MFα1) encodes a potent precursor (the prepro α-factor) of 165 amino acids that consists of a 19 amino acid signal peptide, a 64 amino acid pro region containing three sites for N-linked glycosylation, followed by four tandem repeats of the mature α-factor pheromone preceded by a spacer peptide. The nascent polypeptide is SRP-independently translocated into the ER, where the signal peptide is cleaved off and the three Asp sites within the pro-region are glycosylated, which however is not essential (Caplan et al., 1991; Chaudhuri, Latham & Stephan, 1992). Further modifications occur at the late Golgi, where Kex2p cleaves at dibasic sites, followed by Ste13p removal of hydrophobic spacer sequence (mostly two repeats of Glu-Ala) (reviewed in Porro et al., 2005).

MFα1pp is the most extensively used secretion leader for secretion of foreign proteins in yeast, such as human epidermal growth factor (Brake et al., 1984), insulin-like growth factor 1 (Chaudhuri et al., 1992), full-length antibodies (Rakestraw et al., 2009), cellulase endoglucanase (Zhu, Yao & Wang, 2010) and many more in *S. cerevisiae* as well as in various other

yeasts such as *P. pastoris* (Porro et al., 2005). Naming all would extend the scope of this prologue.

Generally, the MFα1pp obtains the most promising results and production rates and is in most cases even more efficient than native signal peptides (Zhu et al., 2010). Unfortunately, folding and processing sometimes hassles and the N-terminal (Glu-Ala)$_2$ overhangs influence quantity and quality of secreted proteins. However, the MFα1 pro-sequence can also be omitted, leading to correct N-termini (Zhao et al., 2009).

Figure 4: Analysis of secretion enhancing mutations in MFα1pp. a) secreted Fab of mutagenized MFα1pp compared to the wild type leader (error bars indicate one standard deviation of three trials), b) sequences of the most productive mutations indicate common mutations at position 22 (V22A) and 63-66 (more polar) (Rakestraw et al. 2009).

The MFα1pp spacer has ambivalent impact on heterologous protein production, such as preventing intracellular accumulation by decreasing biological activity (Brake et al., 1984), decreasing the enzymatic activity of lysozyme (Oka et al., 1999), and exposure of sequence to further processing enzymes (Brake, et al., 1983) as well as generation of N-terminal ragged ends and increased immunogenicity due to partial cleavage by saturated Ste13p, even though Kex2p processing is increased by the spacer peptide (Parekh, Forrester & Wittrup, 1995). To overcome processing difficulties leading to saturation of enzymatic capabilities and

mislocalization, many attempts have been made. A directed evolution of the alpha mating factor 1 leader peptide (MFα1pp) yielded a 16-fold secretory improvement over the wild type MFα1pp (Figure 4a), hence identifying two accountable mutations (Figure 4b; position 22 and 63-66 in the pro-sequence) (Rakestraw et al., 2009).

In designing secretion systems for heterologous proteins, one aims to maximize both the yield and fidelity of the product, and therefore the use of other secretion leaders might be more advantageous (Sleep, Belfield & Goodey, 1990).

2.4.2 Other used signal sequences

P. pastoris and *S. cerevisiae* have low specificity for recognition of signal sequences and therefore, the recombinant proteins' native signal may also be used successfully for protein expression. A ribosome inactivating protein from *Aspergillus giganteus* (α-Sarcin) has been successfully secreted in *P. pastoris* with its native signal sequence, but not in *S. cerevisiae*. A single mutation of the leader peptide, generating a more favored Kex2p recognition site, enhanced processing and secretion of α-Sarcin (Martinez-Ruiz et al., 1998). Human serum albumin (HSA) was efficiently secreted by its native leader (Kobayashi, 2006). Additionally, human lysozyme was directed correctly processed to the supernatant in *P. pastoris* by the HSA native signal leader (Xiong & Chen, 2008) and the chicken lysozyme native leader (Oka et al., 1999). The matrix metalloproteinases-1, 2, 3, 9 signal peptides are also functional in the *P. pastoris* system (Sreekrishna et al., 1997). Additionally, the plant lectin *Phaseolus vulgaris* agglutinin (PHA; both E- and L-forms) signal sequence was used to secrete correctly processed PHA and GFP in P*. pastoris* (Raemaekers et al., 1999).

Recently, a C-terminal truncated envelope protein from Dengue-2 virus has been targeted to the ER for further folding through the *S. cerevisiae* sucrose invertase 2 (Suc2) signal sequence in *P. pastoris* (Valdes et al., 2007). This signal sequence was also used for hepatitis C virus structural

viral protein (Martinez-Donato et al., 2006) and *A. niger* xylanase expression (Berrin et al., 2000) in *P. pastoris*, just to name a few. Protein secretion was also obtained by using a viral secretion signal derived from the K28 prepro toxin, the precursor of the yeast K28 virus toxin in *P. pastoris* (Eiden-Plach et al., 2004) as well as by using the *P. pastoris* Pho1 (acid phosphatase 1) secretion signal, containing a Kex2p cleavage site (Laroche et al., 1994).

To this end, usage of MFα1pp is mostly improving heterologous protein secretion compared to other or native secretion signals; however, fractions with N-terminal amino acids of the leader are often remaining. Different other prepro sequences were tested in *P. pastoris*, leading to various outcomes compared to the MFα1pp. Generally, the optimal secretion leader should be adjusted specifically for every recombinant protein as the leader influences folding and secretion. The establishment and characterization of novel powerful secretion signals is therefore a common and feasible goal for biotechnology applications and understanding yeast physiology.

2.5   Extracellular protein X; a member of the SCP-like (secretory cysteine-rich) protein super family

As previously reported, experimental determination revealed 20 actually secreted proteins (Mattanovich et al., 2009b) in *P. pastoris*. The most abundant one of them, PIPA00934, was predicted to be secreted by SignalP platform, but lacks an N-linked glycosylation site or transmembrane area. PIPA00934 is a member of the SCP–like (secretory cysteine-rich) protein super family with unknown function, it has a theoretical pI of 5.55 and a molecular weight of 31.72 kDa (Figure 5).
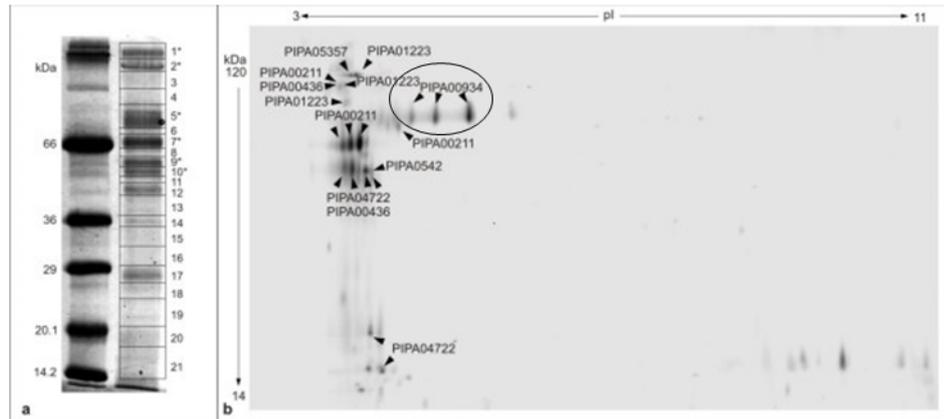
Figure 5: Secretome of *P. pastoris*. a) SDS-polyacrylamide gel. Left lane: molecular weight marker, right lane: supernatant of *P. pastoris* chemostat culture. Boxes indicate the gel slices used for LC-MS protein identification. Bands corresponding to glycoproteins are marked with an asterisk. b) 2D electrophoresis gel of *P. pastoris* culture supernatants. Proteins identified by LC-MS are indicated. Spots belonging to extracellular protein X are highlighted (Mattanovich et al., 2009b).

During several laboratory and large scale fermentations for heterologous protein production, a prominent unidentified host cell protein band at about 70 kDa was observed in *P. pastoris* supernatants. N-terminal sequencing of this protein revealed that it equals the PIPA00934-gene product; however it is present as dimer extracellularly and therefore migrates slower as predicted on SDS-PAGE. Little was known about this protein, named extracellular protein X (Epx1), thus it caught attention and interest because of its high secretion rates.

Subsequent NCBI-BLAST analysis of Epx1 revealed no clear statement concerning full alignment with a specific protein domain. Only the C-terminal half was dedicated to the SCP superfamily, which includes plant pathogen related (PRY) proteins, Golgi-associated PRY proteins and several hypothetical proteins in all kinds of species, such as its homologue Pry2p in *S. cerevisiae* (Figure 6).
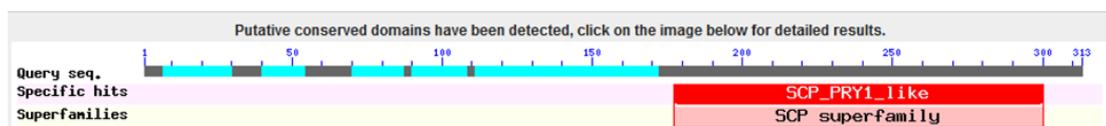
Figure 6: BLASTp search revealed Epx1 C-terminal alignment with SCP-superfamily.

23

This super family also includes mammalian cysteine-rich secretory proteins (CRISP), moreover having a signal peptide for secretion and a conserved cysteine-rich domain. In yeast, the Cys-rich domain is not conserved. PRY proteins in plants are divided into 17 major groups with 2 subclasses, whereof one comprises an acidic subclass that usually is secreted to the extracellular space and the other specifies deposition into the vacuole. Antifungal activity of PR-5 in plants is suggested according to their permeabilization of fungal plasma membrane, but not completely understood so far (Kitajima & Sato, 1999). Human Golgi-associated plant pathogenesis related protein (GAPR-1) mainly forms a membrane-bound dimer, generating a possible catalytic triad across the interface of the dimer structure (Serrano et al., 2004a).

Most prohormones and precursors in higher eukaryotes are processed in dense core secretory vesicles at dibasic sites by the prohormone convertase (PC)/furin family, which are structurally analogous to Kex2p, except that the Ser/Thr-rich domain in Kex2p was replaced by a cysteine-rich domain and some are likely to be soluble proteins within the late Golgi. For mammalian PC3, it was also shown that it undergoes C-terminal processing, presumably at one of several pairs of dibasic amino acids 100-200 residues before the C-terminus (Steiner et al., 1992).

Yapsins, members of aspartic proteases, were first explored as monobasic-specific cleavage enzymes and *kex2p* compensating endopeptidases (Steiner et al., 1992), and their deletion rather than their over expression has been useful in heterologous protein production (Cho et al., 2010). Notably, there are several examples of prohormones in which the processing site occurs at a single Arg site, e.g. the somatotropin releasing inhibitory factor (SRIF) which is endoproteolytically activated at a single Arg or internal dibasic cleavage site (Arg-Lys), generating two functionally different peptides (Bourbonnais et al., 1991). The fungal yapsin Sap9 has been indicated to enclose an autocatalytic processing ability (Albrecht et al., 2006).

The function of yapsins has yet not been fully elucidated, but presumed in cell surface integrity. They divert greatly among eukaryotes, but comprise a signal peptide, a pro-peptide of variable length boarded by dibasic residues (which might occupy the active site cleft in the inactive precursor and serves as sorting signal or chaperone), a conserved catalytic center, presence of disulfide bridges for dimerization, a loop with high Ser/Thr content and a GPI anchor, even though many aspartic proteases are soluble proteins (Gagnon-Arsenault, Tremblay & Bourbonnais, 2006). They are synthesized as inactive zymogens and activated through cleavage at di- or monobasic sites, although the precise course of activation is not yet completely explained, and they might themselves be subjected to late activation at the cell periphery (Komano, Rockwell, Wang, Krafft, & Fuller, 1999).

The pepsin-like aspartic proteinase A (PrA) is essential to the yeast vacuolar system and targeted to the organelle as zymogen. Activation occurs, amongst others, via a step-wise auto-activation product known as pseudo-proteinase A in a transport vesicle en route to the vacuole; eventually mediated through an additional signal either within the pro-segment, the mature molecule or both (Parr et al., 2007).

Summing up, the role and mechanisms of extracellular protein X (Epx1) in *P. pastoris* is not clear at all. Nonetheless, Epx1 has a signal peptide and a pro-peptide, which includes a dibasic motive in the middle and a monobasic terminal motive, whereafter the sequence of mature extracellular Epx1 starts. It is then followed by a Ser/Thr-rich stretch. The last 100 residues, including an additional dibasic C-terminal sequence, are homologous to SCP-like family, even though Epx1 contains only 4 cysteine residues (Figure 7). They might be involved in dimerization of Epx1 through disulfide bonding.

MKLSTNLILAIAAASAVVSAAPVAPAEEA
ANHLHKRAYYTDTTKTHTFTEVVTVYRT**L**
KPGESIPTDSPSHGGKSTKKGKGSTTHSG
APGATSGAPTDDTTSTSGSVGLPTSATSV
TSSTSSASTTSSGTSATSTGTGTSTSTGT
GTGTTGTGTTSSSTSSSATSTPTGSIDAISQ
TLLDTHNDKRALHGVPDLTWSTELADYA
QGYADSYTCGSSLEHTGGPYGENLASGY
SPAGSVEAWYNEISDYDFSNPGYSAGTG
HFTQVVWKSTTQLGCGYK**ECSTDRYYIIC**
EYAPRGNIVSAGYFEDNVLPPV

Figure 7: Extracellular protein X amino acid sequence with secretion leader. Secretion leader is underlined. Purified, secreted Epx1 was N-terminal sequenced revealing LKPGESI (Red) as N-terminal, extracellular start, while the C-terminal peptides were determined by LC-ESI-MS- analysis (Brown).

In contrast to other family members, Epx1 dimers were dominantly found in supernatant at recombinant protein production in *P. pastoris*. Epx1 neither has a transmembrane domain, nor a GPI-anchor. It is however unclear how i) Epx1 is up-regulated especially in recombinant protein production strains, ii) processing and function of propeptide works and iii) how and why Epx1 exists as dimers extracellularly.

3   Aim of the study

The aim of the study is the biotechnological application of Epx1 promoter and secretion leader to the *P. pastoris* host properties on recombinant protein production, and the investigation of underlying mechanisms of strong Epx1 secretion. Secretion leaders show diverse impact on particular recombinant proteins, therefore three structural and functional different proteins are extracellularly and intracellularly analyzed.

4 Material and Methods

4.1 Bacterial Strains

As bacterial hosts, the *Escherichia coli* strains Top10, JM107 or DH10B were used (Table 4.1.1). In order to transform *E. coli* with particular plasmids, either electrocompetent or chemical competent cells were prepared.

4.1. 1: Bacterial strains used for transformation

Strain	Reference/ Source	Genotype
Top10	Invitrogen	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*aral*eu) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G
JM107	Fermentas	F' *traD36 proA*$^+$*B*$^+$ *lacI*^{q} Δ*(lacZ)M15* / e14$^-$ (McrA$^-$) Δ *(lac-pro) endA1 gyrA96* (Nal^r) *thi-1 hsdR17* (r$_k^-$m$_k^+$) *glnV44 relA1*
DH10B	Invitrogen	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ- *rps*L *nup*G

4.1.1 Preparation of electrocompetent *Escherichia coli*

A single *E. coli* Top10 or JM107 colony was inoculated in 10 mL LB-medium (per liter: 10 g soy peptone (QUEST), 5 g yeast extract (MERCK) and 5 g NaCl at pH 7.4 – 7.6, set with 4 M or 8 M NaOH) and incubated over night at 37°C under vigorous shaking.

The main culture (500 mL LB-medium) was inoculated with preculture, approximately generating a start OD_{600} of 0.1. After growing at 37°C for several hours to a final OD_{600} of 0.5 – 1, the culture was transferred into a sterile, precooled centrifuge tube and put on ice for 10 min. Afterwards, the culture was centrifuged at 4,000 x *g* (BECKMAN Coulter, Avanti J-20 XP Centrifuge, rotor JLA-10.500) for 10 min at 4°C.

The pellet was resuspended and washed three times with 1 mM HEPES, followed by another centrifugation for 10 min at 5,000 x *g* at 4°C. Thereafter, the pellet was resuspended in 20 mL 10 % (v/v) glycerol,

followed by another centrifugation and removal of supernatant. The pellet was then resuspended in 500 μL 10 % (v/v) glycerol. Aliquots à 100 μL were quick-frozen in liquid nitrogen and stored at -80°C.

4.1.2 Transformation into electrocompetent *E. coli*

The first few transformations were carried out with Top10 strain, whereas afterwards we switched to JM107. Plasmids were prepared as described below, purified via *mi*-PP200 (metabion) and eluted in 20 μL sterile 1 mM Tris/HCl pH 8, or sterile ddH$_2$O.

For each transformation (ligated empty vector backbone as negative control), one 100 μL aliquot was put on ice 15 min prior to transformation and the eluted, ligated plasmid was added to the cells. Transformation was performed using a BTX Electroporator (BTX-Harvard Apparatus ECM 830 Square Wave Electroporator), a charging voltage of 2.5 kV, a resistance of 1,000 Ω and a capacitance of 25 μF in an electroporation cuvette with a 4 mm gap.

Thereafter, cells were transferred into 1 mL sterile SOC-medium (per liter: 20 g soy peptone (QUEST), 5 g yeast extract (MERCK), 0.58 g NaCl, 0.19 g KCl, 2.03 g MgCl$_2$ x 6H$_2$O, 3.96 g glucose monohydrate, 2.46 g MgSO$_4$ x 7H$_2$O) for a 30 min regeneration period. Then they were plated out in aliquots (50 μL, 200 μL and remains) on selective LB-agar plates (per liter: 10 g soy peptone (QUEST), 5 g yeast extract (MERCK), 5 g NaCl and 20 g agar-agar (MERCK) at pH 7.4 – 7.6 set with 4 N or 8 N NaOH) containing 25 μg mL^{-1} Zeocin™, and inoculated for 24 h at 37°C.

4.1.3 Preparation of chemical competent *E. coli*

Chemical competent *E. coli* cells were prepared by first inoculating a single colony *E. coli* DH10B in 5 mL LB-medium overnight at 37°C under vigorous shaking. The main culture (200 mL LB-medium) was inoculated with preculture, approximately generating a start OD$_{600}$ of 0.1. After growing at 37°C for several hours, to a final OD$_{600}$ of 0.4 – 1, the culture was transferred into a sterile, precooled centrifuge tube and put on ice for 10

min. Afterwards, the culture was centrifuged at 3,000 x g (BECKMAN Coulter, Avanti J-20 XP Centrifuge, rotor JLA-10.500) for 15 min at 4°C.

The pellet was resuspended in 50 mL FB-buffer (per liter: 7.4 g KCl, 7.5 g $CaCl_2$*$2H_2O$, 100 g glycerol, 10 mL 1 M K-acetate at pH 7.5; sterilized by filtration afterwards) and incubated on ice for 45 min, followed by centrifugation for 15 min at 2,200 x g at 4°C. Thereafter, the pellet was resuspended in 15 mL FB-buffer. Aliquots à 150 µL were quickly transferred into precooled 1.5 mL Eppendorf tubes and stored at -80°C.

4.1.4 Transformation into chemical competent *E. coli*

For each transformation (ligated empty vector backbone as negative control), one 100 µL aliquot of chemical competent *E. coli* was put on ice 15 min prior to transformation and the ligated plasmid was added to the cells. The reaction tube was put at 42°C for 30 s and immediately put on ice for 5 min afterwards.

Followed by a regeneration period in SOC-medium at 37°C, cells were plated out in aliquots (50 µL, 200 µL and remains) on selective LB-agar plates, containing 25 µg mL^{-1} Zeocin™, and incubated for 24 h at 37°C.

4.1.5 Plasmid preparation

Independent from the transformation protocol, transformation efficiency was recorded and several colonies were chosen for further analysis. Single cell colonies were picked and streaked out on selective LB-agar plates and further transferred to 2 mL selective LB-medium for 24 h at 37°C.

The master LB-agar plate was stored at 4°C, whereas the overnight culture was used for plasmid Mini-Prep (*mi*-PM200, metabion), following the manufacturer's instructions. Plasmids were eluted in 40 µL 1 mM Tris/HCl pH 8, or sterile ddH$_2$O.

Analytical restriction digestions with specific restriction endonucleases (supplied by Fermentas or New England BioLabs) were performed for verifying the expected plasmid by means of the DNA-fragments analyzed with gel electrophoresis (4.9).

After receiving the correct band pattern for the estimated plasmid size within one or more clones, the respective clone was used for inoculation of approximately 100 mL selective LB-medium and incubated overnight at 37°C under vigorous shaking.

The next day, cell pellet was harvested by centrifugation at 3,500 x *g* for 15 min at RT (C312 Jouan). The supernatant was removed and plasmids were collected via Midi-Prep (HiSpeed® Plasmid Midi Kit (25), QUIAGEN, #12643), following the manufacturer's instructions.

Plasmids were eluted in 1 mL 1 mM Tris/HCl pH 8, or sterile ddH$_2$O, and plasmid concentration was measured (NanoDrop 1000 Spectrophotometer, Thermo Scientific) or determined by comparing the DNA concentration of 1 µL linearized plasmid with 5 µL MassRuler™ DNA Ladder Mix (Fermentas, SM0403).

4.2 Preparation and transformation of electrocompetent *P. pastoris*

For the yeast experiments, an electrocompetent *Pichia pastoris* strain X-33 (Invitrogen, wild type strain; C180-00) was prepared.

4.2.1 Preparation of electrocompetent *P. pastoris* X-33

A single colony from a fresh *P. pastoris* X-33 agar-plate was inoculated in 5 mL sterile YPD-medium (per liter: 20 g yeast extract, 10 g soy peptone and 20 g glucose at pH 7.4 - 7.6) and incubated overnight at 28°C under vigorous shaking.

The overnight culture was used to inoculate 200 mL YPD-medium, receiving an OD$_{600}$ of 3 after an incubation time of 17 h. The therefore required inoculation volume of the preculture was calculated via x$_0$ according equation 1.

Equation 1: Calculating the inoculation volume via x_0

$$x = x_0 \times e^{\mu t}$$

x OD_{600} main culture after time t

x_0 OD_{600} main culture at time of inoculation

μ   0.347 h^{-1} for *P. pastoris* in YPD-medium at 28°C

t   incubation time

After the determined cultivation time and reaching calculated cell density, the culture was transferred into a sterile, precooled centrifuge tube and centrifuged at 1.500 x *g* (approximately 3.000 rpm; BECKMAN Coulter, Avanti J-20 XP Centrifuge, rotor JLA-10.500) for 5 min at 4°C. The pellet was resuspended in 100 mL prewarmed YPD-medium containing 2 mL of 1 M HEPES pH 8.0 and 2.5 mL of 1 M DTT. After incubation for 15 min at 28°C on the rotary shaker (180 rpm), 400 mL icecold ddH_2O were added and cells were harvested by centrifugation as described above. The pellet was washed with 250 mL sterile and icecold 1 mM HEPES. Cells were harvested by centrifugation and the pellet was resuspended in 20 mL sterile icecold 1 M D-sorbitol. Another centrifugation followed and the pellet was resuspended in 5 mL sterile, icecold 1 M D-sorbitol. Aliquots à 100 µL were stored at -80°C.

4.2.2 Transformation of *P. pastoris* by electroporation

For each transformation, one 100 µL aliquot of electrocompetent *P. pastoris* was put on ice 15 min prior to transformation. A small volume of linearized DNA (20 µL column purified plasmid; *meta*bion, PCR purification Kit, *mi*-PP200) was added prior to electroporation. Plasmids were linearized and integrated into the genome, either within the AOX transcription terminator (AOXTT) (via AscI, R0558), or within P_{GAP} (via AvrII, R0174) and P_{EPX1} (via NheII, R0131) if AOXTT was absent on the plasmid.

The cells were transferred to a pre-cooled electroporation cuvette with a 2 mm gap. Transformation was performed using a BioRad Micropulser, a charging voltage of 2 kV and a pulse length of 4 ms.

Thereafter, cells were transferred to sterile YPD-medium for 2 h regeneration period. Then, they were plated out in aliquots (50 µL, 100 µL and 200 µL) on selective YPD-agar plates (per liter: 20 g yeast extract, 10 g soy peptone, 20 g glucose and 20 g agar-agar) containing 25 µg mL^{-1} Zeocin™ and incubated for 48 h at 30°C.

4.2.3 Cultivation conditions for *P. pastoris*

Transformation efficiency was recorded and several colonies were chosen for further analysis. Subsequently, single cell colonies were streaked out again on selective YPD-agar plates. One single cell colony per each clone was used for shake-flask experiments and about 12 individual clones were inoculated for 24 h in 5 mL selective YPD-medium.

The optical density of preculture was determined, whereby 5-10 mL main culture was inoculated in 100 mL shake-flasks, generating a start OD_{600} of 0.1. Dependent on the experiment and examined recombinant heterologous protein, different media were used.

For screening the eGFP expression under the control of P_{GAP} and P_{EPX1}, and for screening the secretion of pTRP under the control of P_{GAP} and P_{EPX1}, autoclaved complex-screening-medium was used (per liter: 10 g pea peptone, 10 g yeast extract, 10.2 g $(NH_4)_2HPO_4$, 1.24 g KCl, 910 µL 1M $CaCl_2$; pH set to 5 with 25 % [v/v] HCl) and 2 mL 500 x biotin and 100 mL 10 x carbon-source (20 % [v/v] glucose, 10 % [v/v] glycerol or 10 % [v/v] methanol) were added before usage.

Synthetic-screening-medium was used for analysis of pTRP under the control of P_{GAP} with three different secretion leader sequences (per liter: 22 g glucose, 22 g citric acid, 3.15 g $(NH_4)_2HPO_4$, 0.027 g $CaCl_2*2H_2O$, 0.9 g KCl, 0.5 g $MgSO_4*7H_2O$, 2 ml 500 x biotin and 1.47 mL trace salts stock solution [per liter: 6 g $CuSO_4*5H_2O$, 0.08 g NaI, 3 g $MnSO_4*H_2O$, 0.2 g $Na_2MoO_4*2H_2O$, 0.02 g H_3BO_3, 0.5 g $CoCl_2$, 20 g $ZnCl_2$, 5 g $FeSO_4*7H_2O$

and 5 mL H_2SO_4]; pH set to 5 with 5 M KOH and sterilized by filtration [0.22 µm]).

The shake-flask cultivation of clones secreting HSA with three different secretion leader sequences under the control of P_{GAP} was performed with the synthetic-screening-medium too, but the pH was set to 6 for assuring increased stability of secreted HSA.

The screening of eGFP secretion was performed with synthetic-screening-medium at pH 5, but 12.6 g $(NH_4)_2HPO_4$ were used instead.

Shake-flask cultivation was carried out for 48 h at 28°C at 170 rpm. Every 12 hours cells were fed with 100 µL 50 x glucose; generating a total concentration of 3.5 % (v/v) glucose at screening end for yielding maximal 20 g/L biomass. Likewise, 400 µl 10 x glycerol or 50 µL 100 x methanol were fed at carbon sources comparing experiments.

Samples were taken either after 12 or 48 hours. Wet cell weight was determined per mL culture, and cell pellets as well as supernatant was directly used for further analysis or stored separately at -20°C in aliquots.

4.3 Plasmids and cloning procedure

The novel *P. pastoris* expression vector pPUZZLE was used, allowing easy exchange of vector components by restriction digestion and inserting different sequences (Stadlmayr et al., 2010). This *E. coli*/*P. pastoris* shuttle vector was used as vector backbone throughout this project. The artificial sequences were exchanged by using restriction sites (Figure 8), added via PCR to gene of interest (Table 4.4.1). Hence AOXTT had to be removed, because of its internal Pfl23II recognition sequence (which hindered pTRP insertion), and ZeoRlox site was exchanged by ZeoR. The Epx1 secretion leader sequences (native and shortened) were PCR amplified and integrated via PvuII, BglI or AccI.
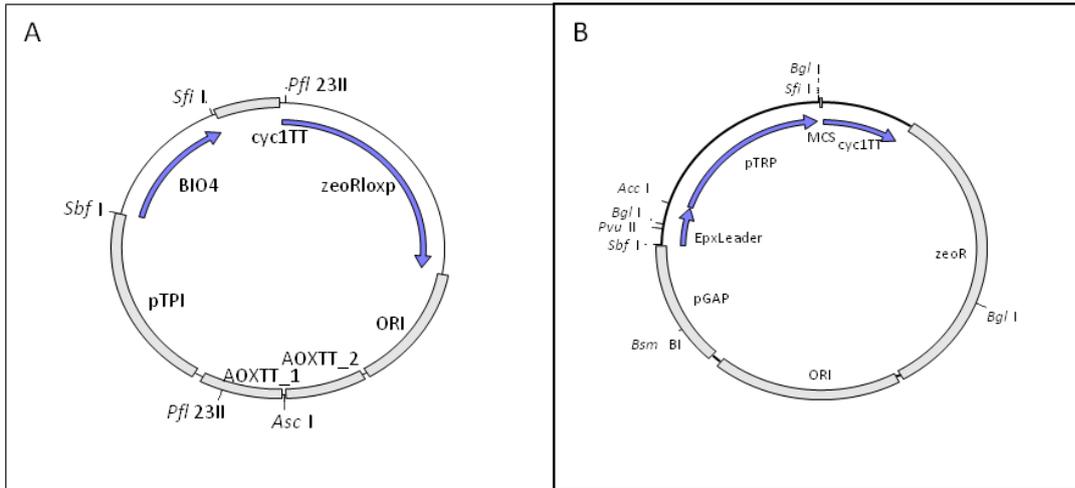
Figure 8: Vector maps. pPuzzle shuttle vector for *E. coli* and *P. pastoris* transformation. Main restriction sites are indicated as well as genes and coding regions. **A**: initial plasmid, described in Stadlmayr et al., 2010. **B**: working plasmid with leader and recombinant protein region and without AOXTT.

4.4 Primer sequences

Oligonucleotides were synthesized by *meta*bio*n* (100 pMol/µL) and are listed in Table 4.4.1. The calculated melting temperature (GC + AT Tm) was between 56 to 67°C, except primers used for amplifying the shortened Epx1 secretion leader.

For sequencing specific regions on plasmids, those primers were additionally used, as well as some more, which are listed here. Sequencing results obtained by MWG Eurofins were then analyzed via SeqMan™.

Table 4.4.1: Overview of forward and backward oligonucleotides with integration site for pPuzzle vector

Forward	Sequences 5´-3´	restriction site
Epx1pp	TATA<u>CCTGCAGG</u>ATGAAGTTCTCTACCAATTTGATC	SbfI
pEpx1	ATTA<u>CGCCGGCG</u>GCATCAAGGGTTCTGGATCTC	MreI (Sse232I)
Epx1pp-KR_pTRP	ATAC<u>CAGCTG</u>CTCCAGTTGCTCCAGCCGAAGAGGCAGCAAACCACTTGCACAAGCGTTTCACTGACGACGACGACAAG	PvuII
Epx1pp-RT_pTRP	ATAC<u>CGTACG</u>TTCACTGACGACGACGACAAG	Pfl23II (BsiWI)
αMF_eGFP	ATGA<u>GAATTC</u>GTGAGCAAGGGCGAGGAGC	EcoRI
Epx1pp-KR_eGFP	ATAC<u>CAGCTG</u>CTCCAGTTGCTCCAGCCGAAGAGGCAGCAAACCACTTGCACAAGCGTGTGAGCAAGGGCGAGGAGC	PvuII
Epx1pp-RT_eGFP	ATTT<u>GTCTAC</u>CGAACTGTGAGCAAGGGCGAGGAGC	AccI
Epx1pp-KR_HSA	ATTC<u>GCCGAAGAGGC</u>AGCAAACCACTTGCACAAGCGTGATGCACACAAGAGTGAGGTT	BglI
Epx1pp-RT_HSA	AGGC<u>GTCTAC</u>CGAACTGATGCACACAAGAGTGAGGTT	AccI

Backward	Sequences 5´-3´	
pEpx1	GATC<u>CCTGCAGG</u>TAGAATAGTGAACGTGTTTTAAAG	SbfI
Epx1pp	GAAG<u>ATGCAT</u>CGTACGGTAGACAGTGACAAC	NsiI
pTRP	TTTT<u>GGCCGAGGCGGCC</u>TTTCAGTTAGCAGCGATAGTTTG	SfiI
αMF_eGFP	TAGG<u>GCGGCCGC</u>TTACTTGTACAGCTCGTCCATG	NotI
eGFP	CGTT<u>GGCCGAGGCGGCC</u>TTACTTGTACAGCTCGTCCATG	SfiI
HSA	GAGT<u>GGCCGAGGCGGCC</u>TTATAAGCCTAAGGCAGCTTGA	SfiI

Sequencing	Sequences 5´-3´
Seq_GAP_forw	ACCAGAATCGAATATAAAAGGCG
pPUZZLE_bw_cycTT	AATTACATGATATCGACAAAGGA

4.5 PCR (polymerase chain reaction)

DNA sequences from plasmid or genomic DNA were amplified. Primers with accessory endonuclease cleavage sites were used (Table 4.4.1), generating a fragment which can be easily integrated into existent plasmids.

Generally we used a touch-down PCR method (Don et al., 1991) with BioRad DNA engine Peltier Thermal Cycler, starting 5°C above calculated annealing temperature and ending 5°C below, followed by 25 elongation cycles. One negative control (for every PCR setup) was prepared without template DNA.

4.6 DNA-polymerases

Polymerases with proof-reading function were used and PCR reaction mix varied according to polymerase, whereas dNTPs (10 mM, Fermentas, R0193) primers and sterile ddH$_2$O remained unchanged.

Most often *Pfu* DNA Polymerase (Fermentas, EP0502, 2.5U/µL) was used. The reaction setup in 50 µL ddH$_2$O was: 5 µL 10 X *Pfu* Buffer, 4 µL 25 mM MgSO$_4$, 1 µL 10 mM dNTPs, 1.5 µL Primer FW, 1.5 µL Primer BW, 1 µL polymerase and 1 µL template. Denaturing temperature was 95°C and elongation time was 2 min/kb at 72°C due to 3´-5´ exonuclease activity (proof-reading).

4.6.1 Generation of blunt ends

Additionally, the generation of blunt ends was performed with T4 DNA polymerase (NEB, M0203, 3U/µL) with 5´-3´ fill-in and 3´-5´ exonuclease activity (very low error rate). Then, 1 µL T4 DNA Polymerase and 1 µL 10mM dNTP were added, and incubated for 15 min at 12°C. Heat inactivation for 20 min at 75°C was beneficial. For ligation with another blunt end fragment, the larger fragment was treated with CIP (see below).

4.7 Digestion of DNA with restriction enzymes
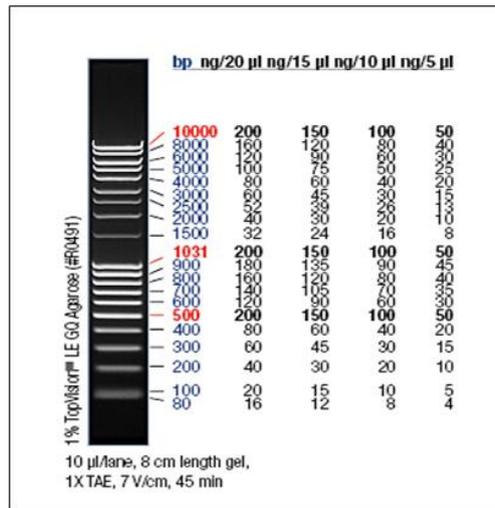
Figure 9: DNA ladder. MR, SM040 MassRuler™ DNA Ladder Mix ready-to-use.

PCR product and vector backbone were cut with respective restriction endonucleases. Therefore, the full PCR mix was digested with 2 µL per each enzyme (Fermentas or New England Biolabs), 5 µL buffer and 5 µL 10 x BSA (if required) for one hour. For preparative setups of plasmid DNA, 3 µL restriction endonuclease, buffer and BSA were added to 40- 80 µL plasmid DNA (High Speed® Plasmid Midi Kit, QUIAGEN, 12643, ca. 100 ng/µL) and incubated at the suggested temperature. After electrophoresis, desired bands were cut out and purified with a gel extraction kit (*m*i-GE100, metabion), following the manufacturer's instructions. DNA was eluted in 40 µL 1 mM Tris/HCl pH 8 or sterile ddH$_2$O. Digestion success and DNA amount (ng/µL) was checked by comparing 1 µL digestion reaction with 5 µL MassRuler™ DNA ladder (Figure 9) via agarose gel electrophoresis prior to ligation.

After ligation and successful transformation into *E. coli*, plasmid DNA (*m*i-PM200, metabion) of several clones was checked for expected band pattern of cleavage products of restriction digestion. Therefore, restriction mix (10 µL eluted plasmid, 5 µL buffer, 5 µL 10 x BSA (if required) and 1 µL endonuclease, ad 50 µL ddH$_2$O) were incubated at the required temperature.

4.8 Dephosphorylation and ligation of DNA fragments

For decreasing self-ligation, the vector backbone was treated with CIP (Calf Intestine Alkaline Phosphatase, M0290, NEB, 1 U/µL). Therefore, 1-3

µL CIP were added to preparative restriction digestion reaction of plasmid DNA, as CIP works with all buffers. After incubation for 30 min at 37°C and heat inactivation for 10 min at 65°C, the restriction digestion reaction was purified by using *mi*-PP200 kit and eluted in 40 µL 1 mM Tris/HCl pH 8, or sterile ddH$_2$O.

DNA concentration was either determined by fluorospectrometer (NanoDrop 3300, Thermo Scientific) or by estimating concentration via agarose gel electrophoresis. For that purpose, 1 µL sample (either insert or cut vector backbone) as well as 5 µL MassRuler™ (Figure 9) were applied to an agarose gel. The concentration was estimated by comparing band intensity of samples with those of known concentrations of the DNA standard. According to equation 2, a ratio of 3/1 of vector/insert was used and ratio of "free ends" was comprised via kilo-base pairs.

Equation 2: Calculating ligation setup

$$insert \; ^{ng}/_{\mu L} = \frac{vector \; kbp \; \times vector \; ^{ng}/_{\mu L}}{insert \; kbp} \times \frac{3}{1}$$

A ligation setup with total volume of 20 µL was prepared with 1 µL T4-DNA Ligase (NEB, M0202, 400,000 U/µL) and 10 x T4 Reaction Buffer (NEB, B0204S) for 2 h at RT. Subsequently, ligation setup was purified using *mi*-PP200 kit and eluted in 20 µL sterile ddH$_2$O, if transformation of electrocompetent *E. coli* followed. For transformation of chemical competent cells purification was not necessary.

4.9 DNA gel electrophoresis

For analysis of selective amplification of DNA fragments via PCR, for controlling DNA fragments after restriction digestion or preparation of fragments for ligation or transformation, DNA gel electrophoresis was used.

Due to cost and health reasons, we switched from ethidium bromide to Midori Green during the course of this work. While working with EtBr as fluorescent dye, per gel (150 mL) 1 % agarose (Biozym, LE Agarose, 840006), 1 x TAE and 6 µL EtBr were used. For Midori Green (NIPPON

Genetics Europe GmbH, MG02) we used 1.5 % agarose, 1 x TAE and 0.5 µl per gel (100 mL). Additionally, fluorescent dyes were added to the EtBr running buffer (1 X TAE with either 0.02 µL EtBr/mL) or none with Midori Green, whereas about 1 µL Midori Green was applied directly into the chamber prior to electrophoresis.

For analytical setups, 1 to 5 µL sample were mixed with ddH$_2$O and 6 x MassRuler® DNA Loading Dye (Fermentas, R0621) or 6 x self made loading dye (2.5 ng/mL Bromphenol blue, 33 % [v/v] glycerol) to a total loading volume of 6 µL. For preparative applications, after restriction digestion, the complete 50 µL mix was mixed with 13 µL 6 x Loading Dye and applied to the gel.

Electrophoresis was performed in running buffer at 130 V (analytical setup) or 90 V (preparative setup). The gel was illuminated for visualizing DNA bands (BioRad Gel Doc™) and DNA fragments for preparative usage were cut out and purified with gel extraction kit (*m*i-GE100, metabion), following the manufacturer's instructions. DNA was eluted in 40 µL 1mM Tris/HCl pH 8, or sterile ddH$_2$O.

4.10 Analysis of (recombinant) proteins

4.10.1 SDS-PAGE

For protein gel analysis the NuPAGE® Novex® Bis-Tris system was used. 15 µL sample, 4 x LDS loading dye (NP0007) and 2 µL Reducing Agent (NP0004) were cooked at 99°C and 15 µL were loaded per slot. Dependent on the subsequent visualization of proteins, several different standards were used. For silver stain, we used 5 µL of 1:10 diluted Bench Mark™, for Western blot we either used 5 µL or 2 µL (dependent on secondary antibody) MagicMark™ Standard (for chemiluminescence detection, Figure 10/B) or 5 µL PageRuler™ Prestained Protein Standard for colorimetric detection (Figure 10/C).

12 % Bis-Tris or 4-12 % Bis-Tris (especially for HSA analysis) gels with MOPS (NP0001) running buffer were used (Weber & Osborn, 1969). Inner and outer chamber (XCell SecureLock, Invitrogen) were filled with 1 x

MOPS and electrophoresis was performed at 180 V for approximately one hour.
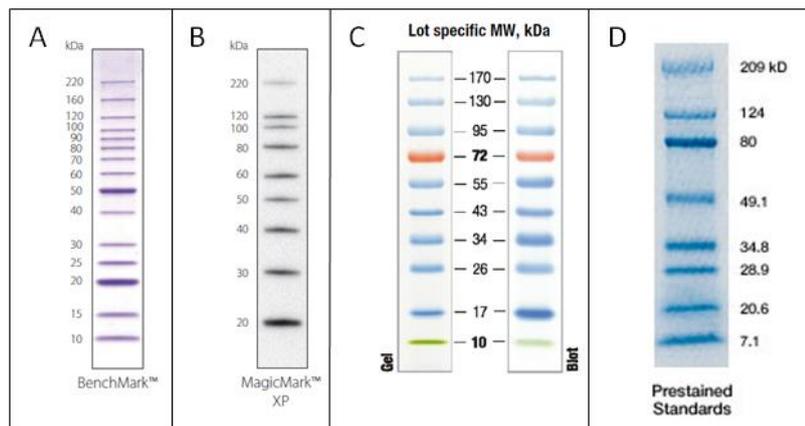
Figure 10: Overview of used protein ladders. A: BM, 10747-012 BenchMark™ **B**: MM, LC5602 MagicMark™ XP **C**: PS, SM0671 PageRuler™ Prestained Protein Ladder, **D**: PP, 161-0363EDU Precision Plus Protein Unstained Standard.

4.10.2 Silver Stain

Silver staining is a very fast, sensitive and unspecific visualization method for protein stains (Rabilloud, 1999). Therefore, it was mainly used to determine secreted proteins in *P. pastoris* culture supernatant.

After successful electrophoresis, the gels were incubated in fixing solution (per liter: 714 mL 70 % ethanol, 100 mL acetic acid) either for 1 h or overnight, followed by incubation for 1 h in incubation solution (per liter: 429 mL 70 % ethanol, 68.2 g sodium acetate, 3.2 g $Na_2S_2O_3*5H_2O$; per 50 mL ad 125 µL gluteraldehyde). The gels were washed three times with ddH_2O for 10 min and then 25 mL silver solution (per liter: 0.5 g $AgNO_3$) with 5 µL 37 % formaldehyde were used per gel. After incubation of 20 min the silver solution was removed, the gel washed with ddH_2O and 25 mL development solution (250 mM Na_2CO_3) plus 2.5 µL gluteraldehyde were added per gel until an appropriate staining was visible. Then, gels were washed with 25 mL stopping solution (0.5 M EDTA) and scanned for documentation (600 dpi, Epson Scan).

4.10.3 Western blot

For intracellular as well as extracellular visualization of specific proteins, Western blot analysis was performed. After gel electrophoresis, transfer buffer was prepared (per liter: 50 mL NuPAGE® 20 x Transfer Buffer, 100 mL MeOH, 1 mL Antioxidant for reduced samples (NP0005); 250 mL per blotting chamber), in which sponges and filter paper (Whatman® 3mm Chr) were soaked. The nitrocellulose membrane (BioRad Trans-Blot® pure nitrocellulose membrane 0.2 µm) was applied upon the gel, according to the manufacturer's instructions, and the blotting chamber was filled with Transfer Buffer. Blotting was mediated at 50 V for one hour in XCell SureLock® Mini-Cell Blot Module (Invitrogen).

Thereafter, the membrane was blocked with 25 mL blocking solution (1 x PBS, 0.1 % [v/v] Tween 20, 1 % [w/v] skim milk) for one hour, washed three times for 10 min with PBST. Incubation with primary antibody (dilution according to Table 4.10.5.1) in 7 to 10 mL blocking solution for one hour or overnight followed. After washing three times for 10 min with PBST, incubation with secondary antibody (dilution according to Table 4.10.5.2) and another washing step followed.

4.10.4 Dot blot

A dot blot was performed to distinguish the pTRP-binding ability of three different primary antibodies. Therefore, on four pieces of nitrocellulose membrane, 15 µL supernatant of *P. pastoris,* producing human trypsinogen and porcine trypsinogen as well as wild type supernatant, were spotted. The membranes were blocked with blocking solution for one hour, primary antibodies were diluted according to Table 4.10.5.1, Table 4.10.5.2 and Figure 19, washed again, incubated with secondary antibodies, washed again and incubated with AP Development Buffer until staining was visible.

4.10.5 Antibodies

Antibodies were used and diluted according to Table 4.10.5.1 and Table 4.10.5.2 in blocking solution.

4.10.5. 1: Primary antibodies and specificity

AB Target	Host	Monoclonal/ Polyclonal	Dilution	Product Name/Supplier	Additional notes
rpTRP	Mouse	whole serum	1: 75	SH22072010	Immunization by Silvia Heiss
rpTRP	Mouse	whole blood	1: 100	Polymun	Immunization by Polymun
rhTRP	Rabbit	polyclonal	1: 2,000 or 1: 100	Acris; AP00906PU-N	
rpTRP	Mouse	whole serum	1: 100 or 1: 10	Polymun	pTRP produced in *P. pastoris*
HSA	goat	polyclonal	1: 50,000 (WB) or 1: 30,000 (ELISA)	Bethyl; A80-129P	HRP conjugated
eGFP	Rabbit	whole serum	1: 5,000	Abcam; ab290	

4.10.5. 2: Secondary antibodies and conjugates

Antibody Target	Host	Dilution	Reactivity	Product Name/ Supplier	Conjugate
Rabbit IgG	Goat polyclonal; IgG fraction antiserum	1: 10,000	IgG whole molecule	Sigma; A8275-1ML	HRP
Rabbit IgG	Goat polyclonal; affinity isolated	1:16,000 to 1:160,000	IgG whole molecule	Sigma; A0545	HRP
Mouse IgG	Goat polyclonal; affinity isolated	1: 10,000	gamma-chain specific	Sigma; A3673	HRP
Mouse IgG	Goat polyclonal; affinity isolated	1: 80,000	Fab-specific	Sigma; A2179	AP
Mouse IgG	Goat polyclonal; affinity isolated	1: 30,000	gamma-chain specific	Sigma; A3438	AP

4.10.6 Conjugates and development of Western blot

For antibodies linked to Alkaline Phosphatase (AP), development buffer for colorimetric detection was prepared (per 50 mL: 2 mL 25 x AP Color Development Buffer (BioRad), 500 µL Solution A and 500 µL Solution B). Per membrane, 25 mL were used and incubated at RT until bands were visible. For antibodies linked to horse radish peroxidase (HRP),

development buffer for chemiluminescect detection was prepared (per membrane: 2 mL Solution A, 2 mL Solution B; Thermo Scientific, Super Signal West Chemiluminescent Substrate), and after 5 min incubation time, chemiluminescence was detected at the LumiImager™ (Böhringer Mannheim; software Lumina32).

4.10.7 Stripping

For testing a different set of antibodies, stripping was performed. Previously bound antibodies were removed by incubating a membrane with 25 mL stripping solution (per liter: 50 mM glycin, 1 mM EDTA at pH 2.3) for 1 h at RT, followed by washing three times for 10 min with PBST and further incubation with 25 mL blocking solution. Subsequent steps followed as described above.

4.11 Antibody production in mice

Due to the fact that none of the tested antibodies were useful to detect pTRP, we decided to produce an antibody for pTRP in mice ourselves.

Therefore, 4 mice (female, B6D2F1/Crl) were immunized with 25 µg porcine trypsinogen/mouse and injection. Right before immunization, approximately 400 µL blood per mouse (pre-bleed) was collected and merged as negative control for ELISA titer determination (Figure 11).

Trypsin from porcine pancreas (Sigma-Aldrich, T0303-1G) was solubilized in 10 ml sterile 1 x PBS and diluted to 1 mg/ 2 mL with sterile 1 x PBS prior to immunization, mixed with 2 mL Freund's Adjuvant Incomplete (Sigma-Aldrich, F5506) and shaken for at least 15 min at high-speed. After spinning down, 200 µL were injected per mouse, three times at an interval of three to four weeks. After the second boost, the antibody titer was determined by ELISA.

4.12 Antibody titer determination

For that reason, a Nunc-Immuno™ 96-well plate MaxiSorp™ (449824) was coated with 1 µg/mL (100 µL) porcine trypsinogen (T0303-1G; 100

mg/mL in 1 x PBS) in coating buffer (HSA carbonate buffer pH 9.6; 0.05 M carbonate-bicarbonate) over night. The next day, coating buffer was removed, the plate was washed three times with 1 x PBS with 0.1 %(v/v) Tween20 and 150 µL blocking buffer (1 x PBS, 1 % [w/v] BSA) were added to each well and incubated for 3 h at RT.

After another washing step, 50 µL serum samples (pre-bleed and after immunization) were applied at an initial dilution of 1:10 in blocking buffer, further diluted in 1:2 steps on the plate and incubated over night at 4°C. Initial dilution was set very low, because we were not sure if immunization with an antigen, such as porcine trypsinogen which differs only marginally from a host protein, would stimulate mouse immune response sufficiently.

After washing three times, 50 µL detection antibody (1: 500 in blocking solution; goat-α-mouse IgG, HRP linked, A3673) were added to each well and incubated for 3 h. Development of enzymatic color reaction was mediated with OPD (P9187). Therefore, 50 µL development buffer was added (200 µL OPD solution (one tablet in 20 mL ddH$_2$O), 16 µL 30 % H$_2$O$_2$ in 20 mL citrate buffer; pH 4.7) to each well and after approximately one minute, the reaction was stopped with 50 µL 3 M H$_2$SO$_4$.

The plate was read at 450 nm, the blank (solely blocking buffer) was subtracted, and it was taken account to the dilution factor within each well. The antibody concentration obviously could not be determined, but we could clearly distinguish between pre-bleed and serum samples after immunization (Figure 11).
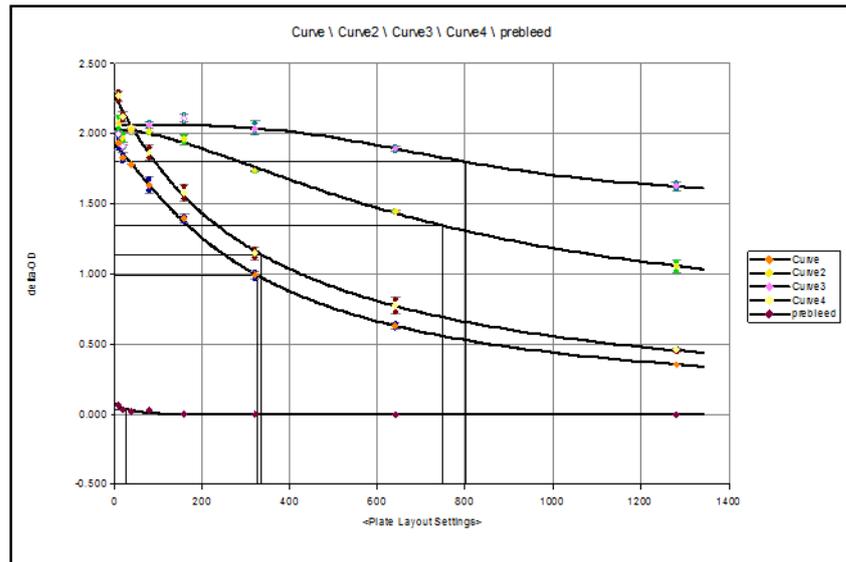
Figure 11: Anti-pTRP titer determination by ELISA. After the second boost with pTRP of four mice, anti-pTRP antibody generation was verified with pTRP-ELISA, with pTRP standard as catcher (1 mg/µl) and prebleed sample as negative control.

Subsequently, the immunization experiment with mice was stopped and the remaining blood was collected, merged and centrifuged for 5 min at 4 rpm after clotting. The serum (SH22072010) was then used for Western blot analysis of intracellular and secreted pTRP.

4.13 Trypsin activity assay

Secreted porcine trypsinogen was further determined and measured by means of TAME-assay (Walsh and Wilcox, 1970), where trypsin activity is based on its esterase activity to TAME (p-toluensulfonyl-L-arginin methyl ester HCL; Sigma, T-4626), and the increase of absorption at 247 nm was measured over time.

Photometric detection was mediated with HITACHI U-2910 double beam spectrophotometer in UV-plastic cuvettes (Brand 7592 00), UV-solutions software and Krüss PT-31 electronic thermostat with Peltier-element.

4.13.1 Buffer exchange

The salt amount and turbidity in the culture medium is debilitating for activity measurement. Therefore, supernatants were desalted by buffer change, using PD-10 small size-exclusion desalting columns (GE Healthcare). Together with LabMate Buffer Reservoir, a volume up to 25 mL could be loaded onto one column.

Prior to purification, PD-10 columns were equilibrated with 25 mL 1 mM HCl (5 x column volume). Thereafter, 2.5 mL well mixed supernatant was applied. Elution of proteins into fresh tubes with 3.5 mL 1 mM HCl followed and 70 µL 2 M $CaCl_2$ was added to each tube for stabilizing trypsinogen and retarding self digestion (autoproteolysis). Samples were either measured right after or stored at -20°C until analysis.

4.13.2 Activity measurement

Trypsinogen was activated to trypsin through enterokinase (lyophilized, Sigma E5510/E0632). By cleaving off the N-terminal hexa-peptide the thereafter active trypsin catalyzes the hydrolysis of p-toluensulfonyl-L-arginine methyl ester HCL by proteolytic cleaving on the C-terminal side of arginine.

Activation buffer (per liter: 6.1 g Tris base, 5.9 g $CaCl_2*H_2O$ at pH 8.6) and dilution buffer (per liter: 4.9 g Tris base, 5.9 g $CaCl_2*H_2O$ at pH 8.1) were prepared and stored at RT for up to one month. For activating trypsinogen at pH 8.6, 300 µL sample, as well as one negative control (300 µL 1 mM HCl), were incubated for 2 h at 37°C with 690 µL freshly prepared enterokinase solution (63.8 U/ml activation buffer).

As samples were measured at 30°C, the PT-31 was checked and the 6-cell changing chamber was attached to the thermostat. TAME solution was freshly prepared (1.18 mM; 22.3 mg TAME in 50 mL dilution buffer) and 1 mL were applied to each cuvette. For activity measurement of samples, 165 µL sample or 165 µL dilution buffer (blank) were added to 1 mL TAME solution, mixed and measurement started immediately for 5 min. With the 6-cell changing chamber, 6 samples could be measured at once, where the

extinction change is measured (15 s per cell) giving a linear slope (ΔE_{247}/min) which should not exceed 0.06. Values up to ΔE_{247}= 0.30 are directly proportional to the trypsinogen amount. Otherwise samples need to be diluted. Trypsin amount (µg/mL) was calculated according to equation 3.

As reported elsewhere (Vandermeers, 1972), the benefit of a high TAME concentration is that the assays are more sensitive and that the velocity of the reaction remains constant throughout most of the TAME hydrolysis. Maximal velocities for trypsin catalyzed hydrolysis are obtained at a pH of about 8.

Equation 3: Calculating µg trypsin per mL on the basis of extinction change

$$trypsin \left[{}^{µg}\!/_{mL} \right] = \frac{\Delta E_{247} \times DF_{sample} \times F}{\epsilon}$$

ΔE_{247} extinction at 247 nm per minute

DF   dilution factor in enterokinase activation solution (1:20)

F   dilution factor due to buffer change and $CaCl_2$ addition
           (3.57 mL/2.5 mL)

ϵ specific extinction coefficient $\epsilon = 0.0101$ E_{247} cm^{-1} min^{-1}

4.14 N-terminal protein sequencing

Edman-degradation was used to determine the N-terminal amino acids of an unknown secreted protein band (Figure 21). Therefore, 500 µL of respective supernatant was loaded onto a centrifugal filter for concentrating the protein (Amicon Ultra-0.5 mL 10 kDa centrifugal filter, Millipore, UFC5010), centrifuged for 5 min and 15 µL sample were recovered by reverse spin.

Thereafter, samples were prepared as described before, applied onto a 4-12 % Bis-Tris NuPAGE®. Western blot with borate buffer (per liter: 3.09 g Borate [50 mM], 100 mL MeOH at pH 9 set with 1 M NaOH) was performed for 2 h at 25 V but using a PVDF (polyvinylidene fluoride) membrane instead. Prior to blotting, the gel was incubated in borate buffer for 10 min,

whereas the membrane was dipped into methanol for 30 s followed by 3 min in borate buffer.

After blotting, the membrane was stained for 3 min with Coomassie (0.1 % [w/v] R250, MeOH [40 % v/v], acetic acid [10 % v/v]), followed by destaining (MeOH [40 % v/v], acetic acid [10 % v/v]). The membrane was rinsed with ddH$_2$O and the protein band of interest was cut out and sent to A. Univ.-Prof. Dr. Herbert Lindner (Medical University Innsbruck) for N-terminal sequencing.

4.15 HSA- ELISA (enzyme-linked immunosorbent assay)

For quantitative detection of human serum albumin, the enzyme-linked immunosorbent assay using the Human Albumin ELISA Quantitation Set (Bethyl, E80-129) were the method of choice.

Therefore, the coating antibody (A80-129A; Bethyl) was diluted to a concentration of 1 µg/mL (1: 1,000) in coating buffer (0.05 M carbonate-bicarbonate, pH 9.6). Therefrom, 100 µL were transfer to each well of a MaxiSorp™ 96-well plate (Nunc, 442404) and incubated for 1 h or over night at RT on a rotary shaker. Afterwards, each well was washed with 200 µL wash buffer (50 mM Tris, 0.14 M NaCl, 0.05 % [v/v] Tween20 at pH 8.0). Aspiration and washing was repeated four times with TECAN 96-well microplate wash station.

Subsequently, 200 µL blocking buffer (50 mM Tris, 0.14 M NaCl, 1 % [w/v] BSA at pH 8) were added to each well and incubated for 30 min at RT on a rotary shaker. Washing of each well was performed as described above.

Samples were diluted in dilution buffer (50 mM Tris, 0.14 M NaCl, 1 % [w/v] BSA, 0.05 % [v/v] Tween20 at pH 8.0) based on expected concentration of HSA in the samples. Necessarily, sample concentration has to be within the concentration range of the standard. To one row dilution buffer only was applied (blank). The standard (stock of 25 mg/mL; Bethyl, RS10-110) was diluted, to an initial concentration of 500 ng/mL was applied twice. 200 µL of standard and sample were transferred to the first assigned

wells and diluted 1:2 in dilution buffer in the following rows and incubated for 1 h at RT on a rotary shaker. Washing was performed as described above.

Detection antibody (anti-HSA, HRP-conjugated; Bethyl, A80-129P) was diluted 1:30,000 in dilution buffer, 100 μL were transferred to each well and incubated for 1 h at RT on a rotary shaker, followed by a washing step.

For HRP mediated enzymatic conversion of peroxidase substrate TMB (3,3',5,5'-tetramethylbenzidine, Bethyl, E102), 100 μL of 1:2 mixed substrate and Peroxidase Solution B were added to each well. For enzymatic reaction, plates were incubated for 15 min in the dark at RT. The incubation time of TMB substrate depends on the intensity of color development, which is direct proportional to the concentration of HSA in the samples. The highest standard concentration should have an OD_{450} of 2.0 – 2.5, whereas the lowest standard concentration should slightly exceed the blank.

The reaction was stopped by applying 50 μL of 2 M H_2SO_4 to each well. The color intensity was measured with a microplate reader (Infinite® 200 Pro, TECAN). Data analysis was performed using Magellan™ software.

The average blank value was subtracted from the average absorbance values. A standard curve was made using software that generates a four parameter logistic (4-PL) curve-fit. The calculated HSA concentration of samples and standard was calculated by using the standard curve. Obtained concentrations were multiplied by the dilution factor to determine the concentration of undiluted samples.

4.16 Cell disruption for intracellular protein analysis

Three different methods for receiving intracellular protein were used. They all had in common that the pellet of 1 mL culture was resuspended in the respective lysis buffer and transferred to ribolyzer tubes. 0.5 mL glass beads (Ø 0.3 mm) were added and cell disruption was performed by using MP-Fast Prep 24 at level 6 for 3x 20 s. Three different methods were used as result of simplifying the application.

The method used first implicated the addition of 50 mM Tris/HCl (pH 7.4), 300 mM NaCl, 5 mM EDTA and 0.02 % (w/v) sodium azide and Protease Inhibitor Cocktail (Sigma, P2714) to 1 x PBS and 1 % (v/v)

TritonX-100. The lysate was transferred by pipetting into a 1.5 mL Eppendorf tube and centrifuged for 15 min at full speed. The supernatant was applied to a new tube and used for analysis.

The second method implied the use of only 1 x PBS with 1 % (v/v) TritonX-100 and a Protease Inhibitor Tablet (Sigma, S8820-20Tab). After mechanical cell disruption, the tubes were put at 99°C for 10 min for denaturation. Further, a small hole was prepared at the bottom of the ribolyzer tube and then transferred on top of a 1.5 mL Eppendorf tube. Through centrifugation at 1.6 rpm for 2 min, the liquid phase without cell debris was collected and used for analysis.

The third method only differed in the last steps. Here, after mechanical cell disruption, the lysate was collected following the method described above. Therefrom, 150 µL were mixed with 50 µL LDS loading dye, put at 99°C for 10 min and centrifuged for 1 min at 13.2 rpm The addition of LDS stabilizes denatured proteins. The supernatant was collected and was directly applied to NuPAGE® gels.

4.17 Flow cytometry measurement (FCM) of eGFP fluorescence

Promoter strength of P_{GAP} and P_{EPX} was analyzed by the expression of eGFP which was determined by FCM. Subsequently, 5 µL culture was mixed with 1 mL sterile particle free 1 x PBS and cell clots were separated by ultrasonic exposure (3 s, 45 % pulser duty, output control 3.5; vibracell Sonic & Materials).

FACS Calibur (BD biosciences), equipped with a 488nm excitation 15 mW air-cooled argon-ion laser and a 630 nm diode laser, was prepared for analysis and FCS Express software was started. Gates were set to exclude cell debris and forward scatter (FSC) and side scatter (SSC) were measured. FSC geo mean and fluorescence emission signal (FL1) geo mean (of 10,000 cells per sample) measurement followed by using PBS as sheat fluid. The FL1 signal was gained in a logarithmic mode by using a 530/30 band-pass filter. Relative eGFP expression levels related to cell volume were calculated as described by Hohenblum (Hohenblum, Borth & Mattanovich, 2003). The untransformed wild type was used as blank.

4.18 Fluorescence photometry

Cell pellets of 1 mL culture were collected by centrifugation (2 min, 13.2 rpm), washed and resuspended in 1 x PBS. Intracellular fluorescence was measured with a spectrofluorimeter (excitation 490 nm; emission 510 nm, 25°C). Measured values were normalized for the intrinsic fluorescence of control *P. pastoris* cells.

5 Results

5.1 Extracellular protein X promoter (P$_{EPX1}$) analysis

To investigate the strength of P$_{EPX1}$, 1,000 bp upstream of the Epx1 start codon were amplified via PCR and used for experimental determination. Therefore, *P. pastoris* X-33 was transformed with plasmids expressing eGFP under the control of either P$_{GAP}$ or P$_{EPX1}$, and integrated into the same locus (AOXTT) in order to rule out position effects.

5.2 Expression analysis of eGFP under the control of P$_{GAP}$ and P$_{EPX1}$ at three different carbon sources

Approximately 12 clones of each construct were simultaneously screened in complex-screening-medium with glucose as carbon source and compared to the non-eGFP expressing wild type. Sampling was carried out after 24 h and 48 h for analyzing expression at different growth rates via FCM (flow cytometry measurement). Hence, promoter activity was measured indirectly by observing expressed gene product.

Figure 12 shows mean eGFP fluorescence per cell size, obtained by averaging over two independent screenings. The number of clones was varying, hence some clones were measured twice and some only once. Generally, eGFP expression was enhanced after 48 h screening period, hence P$_{GAP}$ throve eGFP expression about 10-fold (A: $5.65*10^{-3}$ mean fluorescence/cell size, SD: $2.68*10^{-3}$) compared to P$_{EPX1}$ (A: $3.79*10^{-4}$ mean fluorescence/cell size, SD: $2.77*10^{-4}$). Outliers may indicate multiple integrations (not verified).
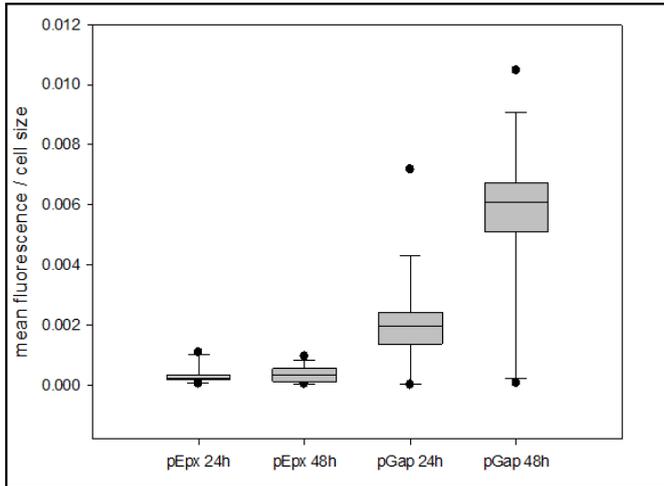
Figure 12: Expression analysis of eGFP grown on glucose and determined by FCM. Samples were measured after 24 h and 48 h under control of P_{EPX1} ($n_{24\,h}$= 12; $n_{48\,h}$= 15) and P_{GAP} ($n_{24\,h}$= 17; $n_{48\,h}$= 16). eGFP expression under the P_{GAP} control was significantly higher. The box represents 50 % of the calculated values, and the line indicates the median. The minima and maxima of each construct are given, indicating extreme values as dots.

As Epx1 has also been observed in methanol grown cells previously, promoter activity was also tested on other commonly used carbon sources, such as methanol and glycerol. In general, promoters are known to respond differently to varying carbon sources (Stadlmayr et al., 2010; Waterham et al., 1997).

Therefore, P_{EPX1} and P_{GAP} activity of approximately 12 clones was additionally analyzed in complex-screening-medium with glycerol (Figure 13/A) and methanol (Figure 13/B) as carbon source and compared to the non-eGFP expressing wild type.

After 48 h screening period, we observed that glycerol enhanced P_{GAP} activity (A: $1.43*10^{-2}$ mean fluorescence/cell size, SD: $3.76*10^{-3}$), but had little positive effect on P_{EPX1} activity (A: $1.87*10^{-3}$ mean fluorescence/cell size, SD: $1.28*10^{-3}$). Expression was up to 10-times higher compared to glucose, but P_{EPX1} was still significantly lower than P_{GAP} (Figure 13/A). There was rather no increase of P_{EPX1} (A: $1.73*10^{-4}$ mean fluorescence/cell size, SD: $8.99*10^{-5}$) and P_{GAP} activity (A: $1.61*10^{-3}$ mean fluorescence/cell size, SD: $1.11*10^{-3}$) by using methanol as carbon source (Figure 13/B).
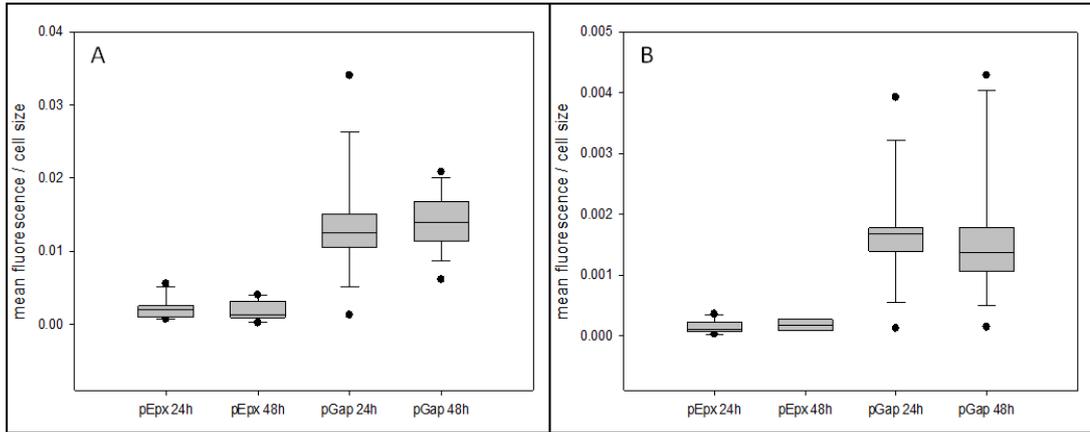
Figure 13: Expression analysis of eGFP grown on glycerol and methanol and determined by FCM. A: Growth with glycerol after 24 h and 48 h under control of P_{EPX1} ($n_{24 h}$= 14 ; $n_{48 h}$= 12) and P_{GAP} ($n_{24 h}$= 16; $n_{48 h}$= 15). eGFP expression under the control of P_{GAP} was significantly higher. **B**: Growth on methanol after 24 h and 48 h under control of P_{EPX1} ($n_{24 h}$= 15; $n_{48 h}$= 6) and P_{GAP} ($n_{24 h}$= 15; $n_{48 h}$= 16). eGFP expression under the control of P_{GAP} was significantly higher. The box represents 50 % of the calculated values, and the line indicates the median. The minima and maxima of each construct are given, indicating extreme values as dots.

5.3 Expression analysis of pTRP under the control of P_{GAP} and P_{EPX1}

Additionally, porcine trypsinogen secretion (pTRP) was analyzed, under the control of P_{GAP} and P_{EPX1}. As pTRP is secreted, we also evaluated the use of the extracellular protein X secretion leader prepro-sequence (Epx1pp) and compared it to the MFα1pp. The four therefore constructed plasmids were integrated into the *P. pastoris* P_{GAP} or the P_{EPX1} native promoter locus, respectively.

To monitor pTRP secretion, influenced by leader or promoter, 12 individual clones of each construct were incubated in YPD-Zeocin™ preculture plus the wild type as control in YPD without Zeocin™, and simultaneous shake-flask-cultivation for 48 h in complex-screening-medium was performed. Sampling occurred only once after 48 h as no growth rate dependency in expression level was observed during FCM analysis. Figure 14 shows measured trypsinogen in the cultivation supernatant, obtained by averaging over two independent screenings of identical clones.

Porcine trypsinogen (pTRP) secretion was quantified via TAME-assay (activity analysis) and protein quality was examined via SDS-PAGE (silver stain). Similarly to eGFP expression determined by FCM, promoter strength was indirectly measured through trypsin activity. In accordance to the

previously obtained results, P_{EPX1} did not lead to high levels of secreted pTRP (with Epx1pp, A: 2.341 µg trypsinogen/g wcw, rel. SD: 95 %; with MFα1pp, A: 2.553 µg trypsinogen/g wcw, rel. SD: 224 %).

In contrast, P_{GAP} with MFα1pp as secretion leader on average led to 76.2 µg/g trypsinogen per wet cell weight (rel. SD: 17 %), compared to 41.15 µg/g trypsinogen per wcw (rel. SD: 34 %) with Epx1 secretion leader (Figure 14). Apparently, Figure 16/A again shows differences in promoter strength of P_{GAP} and P_{EPX1}, analyzed via silver staining. Lanes 3 to 7 as well as 10 to 14 show extracellular pTRP, expressed under the control of P_{EPX1}. In comparison to lanes 8 and 15, pTRP secretion under the control of P_{GAP} was significantly stronger.
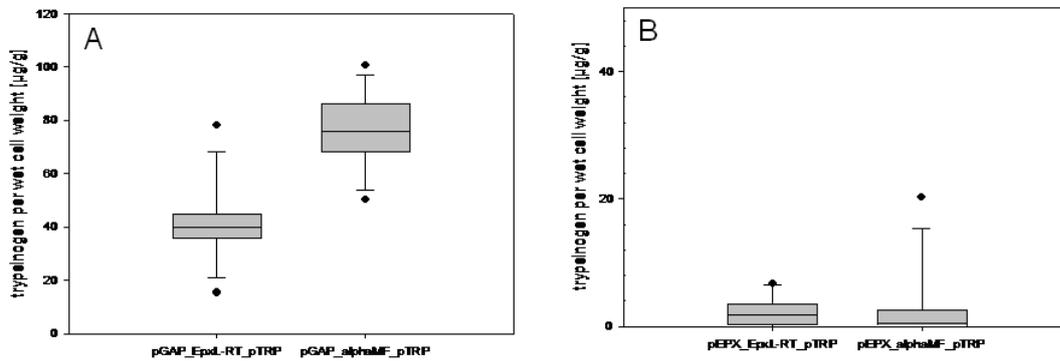
Figure 14: Expression analysis of extracellular pTRP determined by TAME-assay. A: pTRP secretion with two different leader sequences under the control of P_{GAP}. MFα1pp secreted significantly more pTRP compared to Epx1pp. **B:** pTRP secretion with two different leader sequences under the control of P_{EPX1}. TAME-assay revealed weak expression under the control of P_{EPX1} comparing both leader sequences. The box represents 50 % of the calculated values, and the line indicates the median. The minima and maxima of each construct are given, indicating extreme values as dots.

These results clearly indicated that high expression and secretion of extracellular protein X, might not be due to its promoter. However, pTRP secretion seems to be efficiently driven by using the Epx1 secretion leader. Therefore, we analyzed the secretion leader of Epx1 concerning its ability to enhance recombinant protein production.

5.4 Extracellular protein X secretion leader (Epx1pp) analysis

Secretion leaders do have significant impact on secretion efficiency in yeast (Sreekrishna et al., 1997). They share similar motifs (positively charged N-terminus, followed by a hydrophobic middle sequence and a polar C-terminus) which also Epx1pp shows. For example, we confirmed the hydrophobic Epx1 pre-sequence by hydropathicity analysis using Kyte and Doolittle algorithm (Kyte & Doolittle, 1982) with a window size of 9 (Figure 15) by using ExPASy Proteomics Server (http://expasy.org/tools/protscale.html). The dibasic motif within the pro-sequence is indicated, but discussed later.

The pre-sequence determines either co-translational or post-translational transport into the ER with or without SRP-dependency; embedded within the hydropathicity of the pre-sequence. According to Ng and co-workers (Ng et al., 1996), the Kyte and Doolittle blot indicates that most likely Epx1pp is SRP- independently targeted to the ER (Figure 15).
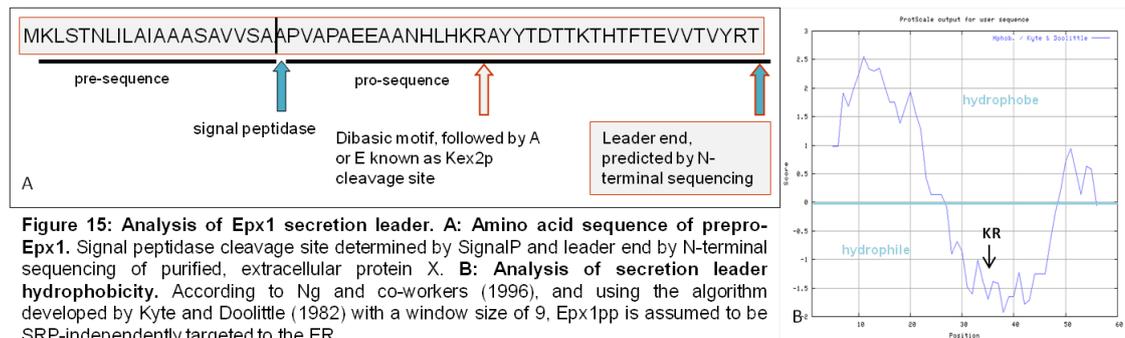
Figure 15: Analysis of Epx1 secretion leader. A: Amino acid sequence of prepro-Epx1. Signal peptidase cleavage site determined by SignalP and leader end by N-terminal sequencing of purified, extracellular protein X. **B: Analysis of secretion leader hydrophobicity.** According to Ng and co-workers (1996), and using the algorithm developed by Kyte and Doolittle (1982) with a window size of 9, Epx1pp is assumed to be SRP-independently targeted to the ER.

Epx1-pre sequence comprises 20 amino acids. Using SignalP platform, we determined signal peptidase cleavage site between position 20 and 21: VSA-AP (http://www.cbs.dtu.dk/services/SignalP/). The signal peptidase cleavage site is followed by three hydrophobic amino acids (Ala-Phe-Val), determining the start of the pro-region (Figure 15/A). The pro-region comprises 37 amino acids and Figure 15/B indicates its hydrophilic range. We determined the leader end by N-terminal sequencing of the purified, extracellular protein X, however an additional dibasic motif within the pro-sequence was noticed, which indicated a possible Kex2 favored cleavage site, but was not further implicated at this time.

5.5 Comparative analysis of pTRP with native Epx1pp and MFα1pp secretion leader under the control of P_{GAP} and P_{EPX1}

As already reported, the expression profile of pTRP was analyzed under the control of P_{GAP} and P_{EPX1} (integrated into the Gap- or Epx1 promoter locus) for investigating promoter strength on the one hand and Epx1 leader analysis on the other hand, in comparison to the well studied and efficient MFα1pp. As discussed above (Figure 14) and additionally shown below, expression under the control of Epx1 promoter showed diminished levels (Figure 16/A, lane 3-7 and 10-14; Figure 14/B) compared to the Gap promoter (Figure 16/A, lane 8 and 15; Figure 14/A) comparing both leader sequences. Therefore, P_{EPX1} analysis was discontinued and we performed further studies under the control of P_{GAP}.

5.5.1 Qualitative analysis of pTRP secretion

Unexpectedly, pTRP expression using the Epx1 secretion leader led to a protein smear of unidentified cause (Figure 16/A, and 16/B, black dot), whereas MFα1pp led to secreted pTRP of correct size (25 kDa; Figure B, arrow). Parts of secreted pTRP with Epx1pp, however, also led to correct size pTRP, albeit to a lesser extent. We hypothesized that the smeared band represents an Epx1pp-pTRP fusion protein, which appears due to incorrect processing, leading to partially unfolded pTRP.
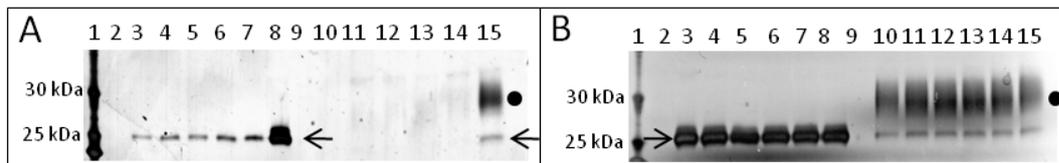
Figure 16: Silver stain detection of secreted pTRP. Comparison of pTRP secretion by reduced SDS-PAGE, after 48 h cultivation under the control of two different promoters and two signal leaders. Epx1pp secretion sequence leads to Epx1pp-pTRP fusion protein (dot). **A:** Overview of all four constructs. L1: BM, L3-7: pEpx1_MFα1pp_pTRP, L8: pGap_MFα1pp_pTRP, L10-14: pEpx1_Epx1pp_pTRP, L15: pGap_Epx1pp_pTRP. **B:** Effect of signal leader on pTRP secretion. L1: BM, L3-8: pGap_MFα1pp_pTRP, L10-15: pGap_Epx1pp_pTRP. Arrows indicate porcine trypsinogen (25 kDa).

We speculated whether this fusion protein (or just the small amount of correct size trypsinogen) could be activated by enterokinase or not. Diminished conversion of TAME then might not adequately represent the amount of secreted pTRP.

5.5.2 Activation test of Epx1pp-pTRP fusion protein with enterokinase

To test the capability of enterokinase in activating Epx1pp-trypsinogen fusion protein, two random clones per construct were analyzed. The supernatant before and after enterokinase incubation was visualized via reduced SDS-PAGE (Figure 17). Trypsinogen secreted by any leader, acquires zymogen activation to trypsin through enterokinase. A hexa-peptide at the trypsinogen N-terminus is thereby cleaved off by releasing active trypsin (Table 5.7.1).

Even though Epx1pp led to an aggregation of indeterminable cause, activation with enterokinase yielded correctly cleaved trypsin (Figure 17/lane 2 and 3) and as expected, trypsinogen secreted by MFα1pp was fully activated too (Figure 17/lane 6 and 7). These results indicate that the secreted amount of trypsin measured with TAME assay likely corresponds to the actual secreted amount of pTRP.
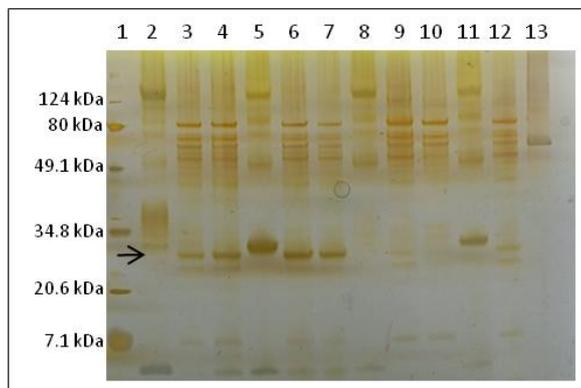
Figure 17: Trypsinogen activation by enterokinase. Reduced SDS-PAGE indicates that EK activates Epx1pp-pTRP-fusion protein, as well as trypsinogen secreted with MFα1pp, leading to trypsin. L1: PP, L2: #7 pGap_Epx1pp_pTRP, L3+L4: #7 and #9 after EK addition, L5: #6 pGap_MFα1pp_pTRP, L6+L7: #6 and #11 +EK, L8: #7 pEpx1_Epx1pp_pTRP, L9+L10: #7 and #12 + EK, L11: #2 pEpx1_MFα1pp_pTRP, L12+L13: #2 and #10 + EK. Arrow indicates trypsin (23.8 kDa).

5.5.3 Analysis of intracellular pTRP

As Epx1pp secreted trypsinogen, but shows significant amounts of Epx1pp-pTRP fusion-protein extracellularly, we speculated if we can also see effects intracellularly. Therefore, the cell pellet of several clones was

disrupted to obtain intracellular proteins and a Western blot was thereof performed.

Surprisingly, neither the anti-hTRP (Figure 18/A) nor the anti-pTRP (Figure 18/B) antibody bound at expected size (~25 kDa). In contrast, unspecific binding of each antibody to high molecular mass proteins, independent of the used secretion leader was observed.
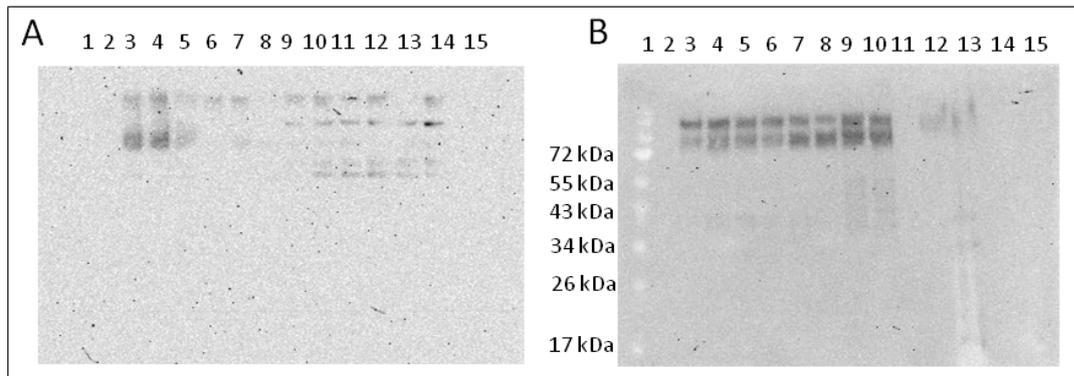
Figure 18: Western blot analysis to detect intracellular pTRP. Two different primary antibodies were compared concerning their ability to bind pTRP. **A**: Anti-hTRP antibody failed to bind pTRP. L1: pGAP_MFα1pp_pTRP supernatant, L2: purified pTRP, L3-5: pGAP_MFα1pp_pTRP, L6-8: pGAP_Exp1pp_pTRP, L9-11: pEPX1_MFα1pp_pTRP, L12-14: pEPX1_Exp1pp_pTRP, L15: PP ladder. **B**: Mouse-anti-pTRP antibody failed to bind pTRP. As *P. pastoris* derived pTRP was used as antigen, unspecific binding is likely. L1: PP, L3+4: pEPX1_Exp1pp_pTRP, L5+6: pEPX1_MFα1pp_pTRP, L7+8: pGAP_Exp1pp_pTRP, L9+10: pGAP_MFα1pp_pTRP, L12: pGAP_MFα1pp_pTRP supernatant, L13: intracellular; wild type.

The question arose whether no pTRP was within the cells, hence completely secreted or if the antibodies were not binding as expected. Therefore, a Dot blot was prepared (Figure 19) with samples of different supernatants, wherein trypsin was prior determined (not shown).

Figure 19/A shows that rabbit-α-hTRP also binds pTRP, whereas mouse-α-pTRP antibody detects pTRP even at a dilution of 1:100 (Figure 19/B and 19/C), and mouse whole blood also showed pTRP recognition (Figure 19/D). Unexpectedly, the same antibody, which detected secreted pTRP in this Dot blot (displayed in B and C) failed to bind intracellular pTRP specifically (Figure 18).
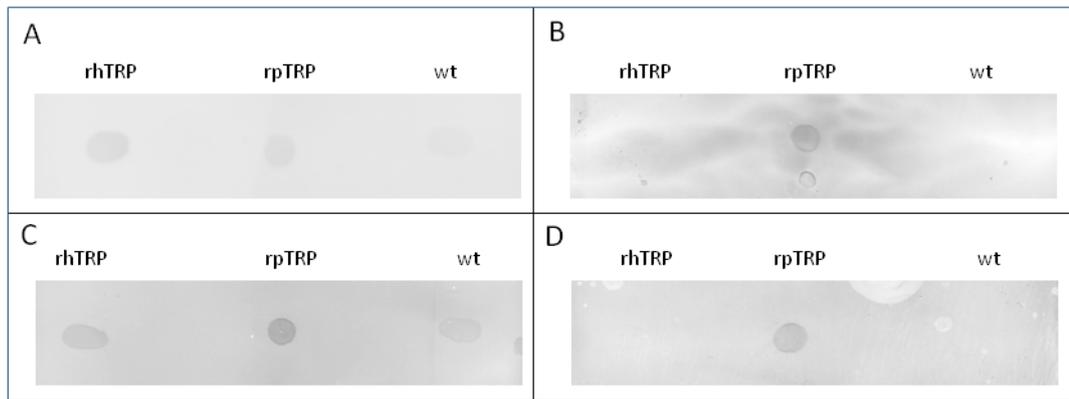
Figure 19: Dot blot concerning the rhTRP- and rpTRP-binding ability of available antibodies. Supernatants, comprising recombinant human trypsinogen (M149, G9, 080411MM), recombinant porcine trypsinogen (H138, APTRP_NIX G9, 070927) or none (H61, X33-wt G9, 050324MM), were spotted on nitrocellulose membrane and incubated with different self-made primary and purchased secondary antibodies. **A**: AB1, rabbit-α-hTRP, 1:100; AB2, A8275. **B**: AB1, mouse-α-rpTRP, 1:100; AB2, A2179. **C**: AB1, mouse-α-rpTRP, 1:10; AB2: A2179. **D**: AB1, mouse whole blood, 1:100; AB2: A2179.

This antibody (α-pTRP, Figure 19/B and C) was created in mice but the antigen (pTRP) was produced in *P. pastoris*. We therefore assumed that on the one hand, it binds abundantly available pTRP in the supernatant, but on the other hand it binds unspecific yeast protein rather than small amounts of pTRP.

Nonetheless, for subsequent visualization of intracellular pTRP an antibody was produced in mice. Therefore, mice were immunized with trypsin from porcine pancreas (Sigma-Aldrich, T0303-1G). After the second boost, the titer determination by ELISA (Figure 11) revealed sufficient anti-pTRP antibody production (compared to pre-bleed). The sera were used for extra- and intracellular pTRP detection by Western blot. Unfortunately, the new antibody again detected secreted pTRP, but failed to detect small amounts of intracellular pTRP (data not shown).

Moreover, once more unspecific band patterns were observed intracellelularly. Due to similar patterns received by other blots and mouse derived antisera within our group in the meantime, we reasoned that this pattern was due to unspecific binding of secondary antibodies.

5.5.4 Analysis of Epx1pp-pTRP fusion protein in accordance to altered pH-values and incubation time

As described within the introduction, proteins related to Epx1 were identified to undergo autocatalytic activation and cleavage of pro-peptide; most likely mediated through an additional cleavage site within the pro-sequence (Albrecht et al., 2006; Leduc R, 1992; Parr et al., 2007; Wilcox & Fuller, 1991). Furthermore, other proteins such as yapsins and aspartic proteases assist at the cell surface in processing at mono- or dibasic sites, however, most of them are cell wall anchored (Gagnon-Arsenault et al., 2006). Notably, the secretome of *P. pastoris* was analyzed and no proteolytic enzymes were found outside of glucose grown cells (Mattanovich et al., 2009b). Cell free supernatant was analyzed, therefore membrane-bound processing enzymes can still be assumed to cleave at specific motives shortly after secretion.

Epx1 also has an auxiliary dibasic site within the pro-sequence (Figure 15) and therefore, we tested whether the Epx1pp-pTRP fusion protein appearance changed after time in supernatants, also containing natively secreted Epx1. We reckoned if extracellular dimerized Epx1 is present, it may cleave the Epx1 pro-sequence attached to pTRP over time and altered pH values.

Trypsin is a serine protease, cleaving peptides C-terminal of Lys or Arg. We also supposed the possibility of active trypsin itself cleaving at the Epx1pro Lys-Arg sequence of unprocessed Epx1pp. Another characteristic of trypsin is its optimal operating pH of about 8. Screening medium in contrast has a buffered pH of about 4 at the end of the screening. The pH was set from 4 to 6 to observe differences, but not to enhance its autocatalytic properties. The supernatant per clone was incubated at three different pH values for 76 h at 30°C with sampling every 24 hours (Figure 20).
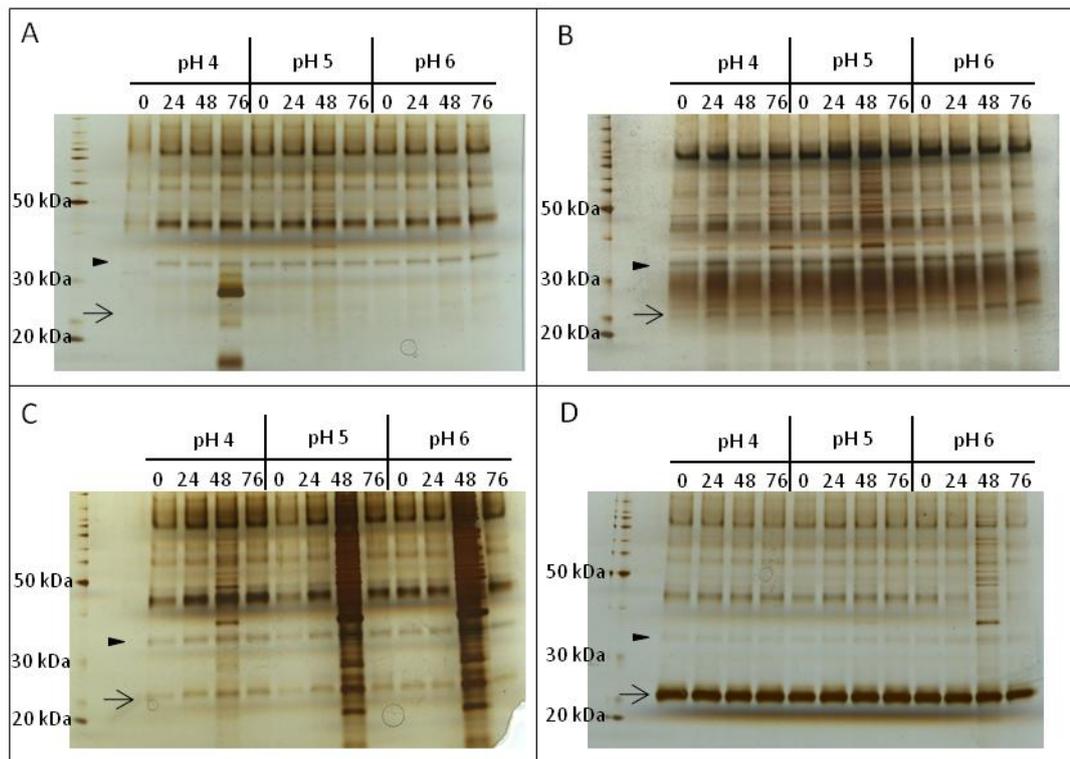
Figure 20: Autocatalytic cleavage test of Epx1pp-pTRP fusion protein in relation to pH and incubation time. No distinctive difference or autocatalytic activation of trypsinogen in accordance to pH value of media or incubation period at 30°C was observed via reduced SDS-PAGE. **A**: pEpx1_Epx1pp_pTRP, clone #8, **B**: pGap_Epx1pp_pTRP, clone #7, **C**: pEpx1_MFα1pp_pTRP, clone #2, **D**: pGap_MFα1pp_pTRP, clone #11. Arrow indicates trypsinogen (25 kDa). Arrowhead indicates potential Epx1 (32 kDa).

No distinctive difference or degradation of Epx1pp-pTRP fusion-protein through native secreted Epx1, in accordance to pH value of complex-screening-medium or incubation time was observed for any construct. The suggested impact of proteolytically active trypsin to the protein smear as well as the possibility of extracellular protein x cleaving the Epx1pp pro-sequence therefore was rejected.

5.6 Qualitative analysis of HSA, secreted with native Epx1 and native HSA secretion leader

Human serum albumin (HSA) is a globular monomeric protein which possesses no enzymatic activity and was often successfully secreted by its native signal leader in the *P. pastoris* expression system (Kobayashi, 2006). Therefore, HSA in combination with its leader sequence was used for

comparison; the expression vector was obtained by G. Stadlmayr (Stadlmayr et al., 2010). Additionally, Epx1pp was studied as alternative secretion leader for HSA, and was amplified via PCR with respective C-terminal Epx1pp overhangs and replaced pTRP in the previously used vector.
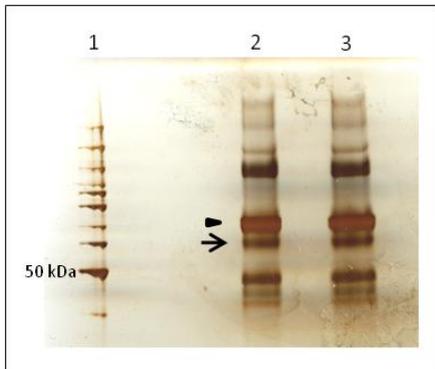
Figure 21: Silver stain detection of secreted HSA. Secreted HSA with EpxLpp secretion leader after 48 h cultivation indicates a double band. Lower band indicates HSA (67 kDa; arrow) on reduced SDS-PAGE. Upper band (arrowhead) and the lower band were sent to N-terminal sequencing. L1: BM, L2 + 3: pGap_Epx1pp-RT_HSA.

12 individual clones of each construct were analyzed concerning secretion behavior after 48 h shake-flask cultivation in synthetic screening medium. Qualitatively, secreted HSA was examined by reduced SDS-PAGE and visualization by using silver stain. Apparently, the expression pattern with the Epx1pp variants was different than observed with pTRP. Once again, an Epx1pp-fusion protein by expression with the native Epx1 leader was observed, hence no smear, but a distinct double band appeared (Figure 21, lane 2 and 3). In comparison with HSA secreted by its native leader and purified HSA (positive control, Figure 25/B) we concluded that the lower band was correctly processed HSA. The upper band was unexpected, but due to its slightly bigger molecular weight, we presumed that the Epx1 pro-sequence was partially cleaved at the dibasic site (Lys-Arg).

Therefore, the supernatant of one random clone was used to prepare a Western blot. The upper and lower bands were cut out and sent for N-terminal sequencing. Our presumption concerning cleavage after the dibasic Lys-Arg motive, instead of the monobasic Arg-Thr predicted end, was verified by N-terminal sequencing. For the upper band, AYYT was determined N-terminally, and for the lower band, D was determined N-terminally as the first amino acid, conclusively with the first amino acid of HSA (Table 5.7.1). Hence, we obtained HSA with 21 amino acid overhang

by remaining Epx1 pro-sequence (Figure 21, upper band) in addition to the correctly cleaved form (lower band).

5.7 Generating a shortened Epx1 secretion leader sequence by omitting amino acids of Epx1 pro-sequence

The presumption arose that the Lys-Arg motif within the pro-sequence somehow disturbs correct processing. Numerous proteins which might be Kex2p target proteins have been annotated and analyzed by Bader and colleagues. In short, Kex2p substrate recognition and specificity varies between yeasts. These authors also depicted that Kex2p processing additionally depends on three-dimensional structure and species/protein related specifications (Bader et al., 2008).

We reckoned if omission of 21 amino acids at the C-terminal pro-sequence and the therefrom resulting dibasic leader end (Figure 22) will favor and enhance correct processing, in particular by Kex2p. The shortened Epx1 leader (Epx1pp-KR) was therefore amplified via PCR and its impact on secretion of pTRP, HSA and eGFP was analyzed.
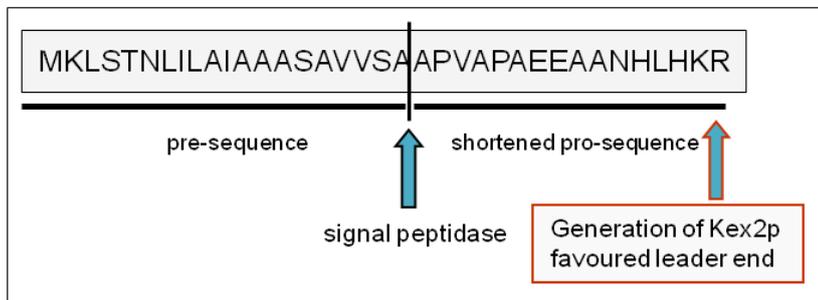
Figure 22: Amino acid sequence of shortened Epx1 secretion leader. Due to processing difficulties of pTRP and HSA using Epx1pp, the pro-sequence was shortened, creating an alternative end namely the dibasic site Lys-Arg, which is a known Kex2p target site.

Table 5.7.1 gives an overview concerning recombinant proteins used in this study. It lists the first 15 N-terminal amino acid sequences of porcine trypsinogen (with hexapeptide; underlined; chipped off for receiving active trypsin), human serum albumin (with native HSA secretion leader; positive control; underlined), and enhanced green fluorescent protein (with alpha mating factor leader and spacer peptide; positive control; underlined).

5.7. 1: Amino acid sequences of recombinant proteins used in this study

Accessory sequence and recombinant protein		15 N-terminal amino acids
hexapeptide + porcine trypsin		<u>TDDDDK</u>IVGGYTCAA
native HSA leader + HSA	<u>MKWVTFISLLFLFSSAYSRGVFRR</u>	DAHKSEVAHRFKDLG
MFα1pp (incl. spacer peptide) + eGFP	<u>MRFPSIFTAVLFAASSALAAPVNT TTEDETAQIPAEAVIGYSDLEGDF DVAVLPFSNSTNNGLLFINTTIASI AAKEEGVSLEKREAEAEF</u>	VSKGEELFTGVVPIL

5.7.1 Shortened Epx1pp (Epx1pp-KR) and the impact on pTRP secretion

Simultaneously, a population of approximately 12 individual clones for each construct was screened. Therefore, the same previously screened clones were compared with 12 clones employing the shortened Epx1 leader (Epx1pp-KR).

First, supernatants were analyzed after 48 h shake-flask cultivation in synthetic-screening-medium via SDS-PAGE and silver staining. We surprisingly observed that processing was indeed enhanced by using the shortened Epx1 secretion leader (Figure 23, lane 6 to 9). Moreover, no protein smear or the like was detected.

Figure 23: Silver stain detection of secreted pTRP. Comparison of pTRP, after 48 h cultivation with three different secretion leader, via reduced SDS-PAGE. In contrast to native Epx1pp (Exp1pp-RT), the shortened Epx1pp secretion signal (Epx1pp-KR) leads to increased pTRP secretion without fusion protein. L1: BM, L2-5: pGap_Epx1pp-RT_pTRP, L6-9: pGap_Epx1pp-KR_pTRP, L10-13: pGap_MFα1pp_pTRP, L14: porcine trypsin (1 mg /ml), L15: wt. Arrow indicates porcine trypsinogen (25 kDa).

Secondly, the expression mediated by the shortened leader in comparison to native Epx1pp and MFα1pp was quantitatively analyzed. The same clones for native Epx1pp-pTRP and MFα1pp-pTRP were used as in Figure 14/A. Expression was generally augmented due to using synthetic-screening-medium in contrast to complex-screening medium, hence the same clones showed marginal expression differences between the two screenings. However, the distribution remained similar.

Clearly, the shortened Epx1 secretion leader enhanced protein expression (Figure 24), yielding approximately 100 µg correctly processed trypsinogen per gram wet cell weight.
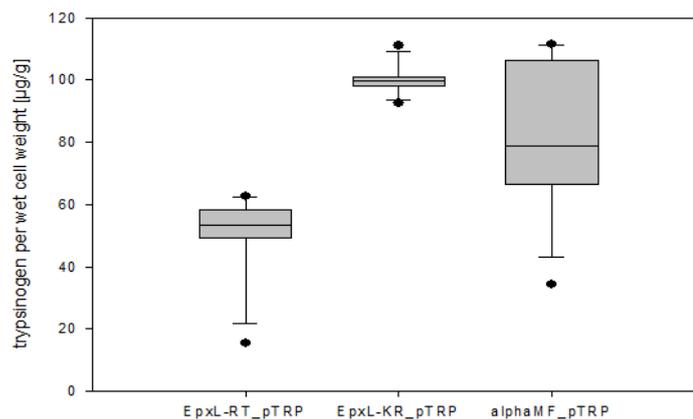
Figure 24: Expression analysis of pTRP secretion with three different leader sequences. pTRP secretion determined by TAME-assay after 48 h cultivation (n=12). The box represents 50 % of the calculated values, and the line indicates the median. The minima and maxima of each construct are given, indicating extreme values as dots.

5.7.2 Shortened Epx1pp (Epx1pp-KR) and the impact on HSA secretion

A population of 12 individual clones from each construct was simultaneously screened in shake-flask cultivation with screening medium for 48 h and the supernatant was analyzed via Western blot (Figure 25/A) and silver stain (Figure 25/B).

As reported previously, native Epx1 secretion leader led to a double band when secreting HSA. We approved that the Epx1 pro-sequence was partially cleaved at the intrinsic Lys-Arg sequence. Therefore it was assumed, that omitting the subsequent 21 amino acids would facilitate

secretion leader processing, thus leading to improved secretion of HSA. In contrast, the shortened leader failed to augment secretion of HSA, even though HSA was processed correctly (Figure 25/A and 25/B, lane 6 to 9).
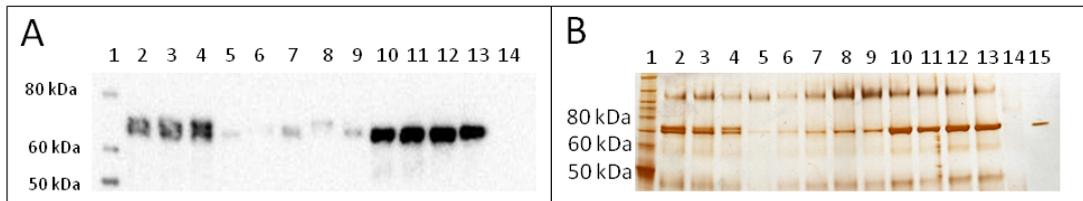
Figure 25: HSA secretion with three different leader sequences. The native Epx1 secretion leader partially cleaved at the dibasic Lys-Arg or monobasic Arg-Thr motif, leading to a HSA double-band. **A**: Western blot, reduced 4-12 % Bis/Tris NuPage®. **B**: Silver stain, reduced 12 % Bis/Tris NuPage®. L1: MM (A), BM (B), L2-5: Epx1pp-RT_HSA, L6-9: Exp1pp-KR_HSA, L10-13: native HSA signal peptide_HSA, L14: wt, L15: HSA Standard (2.5mg/L; 67 kDa).

Additionally, secreted HSA was quantitatively analyzed via HSA-ELISA. At a dilution of 1:50, supernatants of 9 (respectively 6 for HSA) clones were analyzed (Figure 26). In accordance to reduced SDS-PAGE analysis, the shortened Epx1pp averaged 1.3 mg/L HSA, whereas the native Exp1pp yielded 6.3 mg/L HSA which was significantly lower ($p < 0.05$) than native HSA leader (13.1 mg/L).
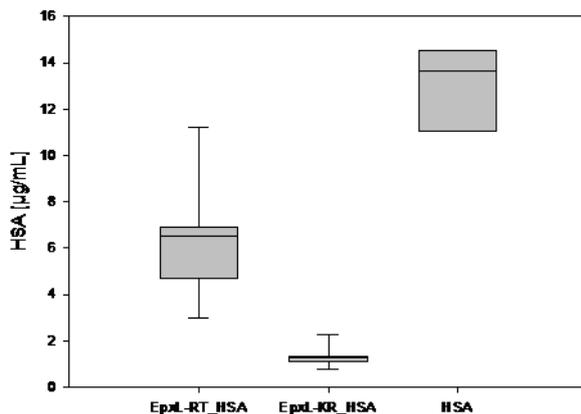
Figure 26: Expression analysis of HSA secretion with three different leader sequences. Determined by ELISA, Epx1pp-RT and Epx1pp-KR signal leader showed no increase of secreted HSA compared to native HSA leader. The box represents 50 % of 9 (respectively 6 for HSA) individual clones, and the line indicates the median. The minima and maxima of each construct are given.

5.7.3 Comparative analysis of intracellular HSA with native and shortened Epx1pp, and native HSA leader

Intracellular determination of HSA by Western blot revealed a different band pattern as expected regarding to extracellular HSA, even though performed several times, comparing different screenings and using different

methods for intracellular protein analysis (Figure 27). However, intracellular and extracellular expression level matched. Interestingly, no intracellular double band pattern was seen for native Epx1pp. HSA levels with the shortened Epx1pp secretion leader were very weak intracellularly (Figure 27, lane 6 to 9); indicating a bottleneck in leader processing.
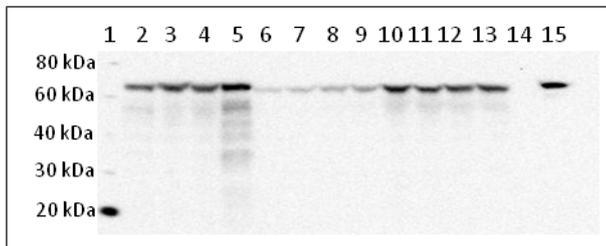
Figure 27: Western blot analysis to detect intracellular HSA. L1: MM (A), BM (B), L2-5: Epx1pp-RT_HSA, L6-9: Exp1pp-KR_HSA, L10-13: HSApp _HSA, L14: wild type, L15: HSA Standard (2.5 mg/L; 67 kDa).

Immunofluorescence analysis of intracellular HSA performed by Verena Puxbaum, however, achieved visualization of clear differences within the cellular distribution (Figure 28). The native Epx1pp led to distribution of HSA at the plasma membrane and the bud (Figure 28/A.1). Plasma membrane staining was already reported for hydroxynitrile lyase in *S. cerevisiae* (Hasslacher et al., 1997), where it indicated periplasmatic retention, while bud staining usually is a sign for secretion. The shortened leader in contrast showed disposition into the cytosol (Figure 28/B.1), and the native HSA leader illustrated a weak staining, probably of the ER/nuclear envelope (Figure 28/C.1).
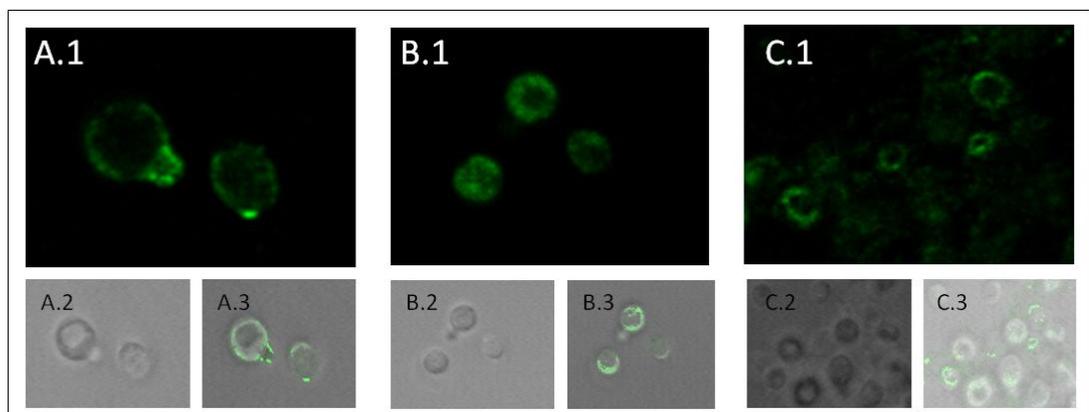
Figure 28: Immunofluorescence staining of HSA, secreted by three different secretion sequences. Immunofluorescence staining was performed with FITC labeled anti-HSA antibody (1) and LCM images (2) are merged with IF (3). The original leader (Epx1pp-RT) targets HSA to the plasma membrane mainly at budding sites (**A**), the shortened leader (Epx1pp-KR) shows strong accumulation in the cytosol (**B**) and the native HSA leader shows little intracellular HSA mainly at ER (**C**) (by V. Puxbaum).

It seemed that the expressed HSA is transported more or less efficiently out of the cells by native Epx1pp and HSA leader. In contrast, the shortened leader showed weak expression at extracellular and intracellular analysis. Therefore, the cytosolic localization of HSA determined by IF rather displayed folding and processing difficulties, which may lead to ERAD, and furthermore ubiquitin mediated degradation of HSA within the cytosol.

5.7.4 Comparative analysis of secreted eGFP with native and shortened Epx1pp, and MFα1pp secretion leader

Additionally enhanced green fluorescent protein (eGFP) was tested concerning its secretory behavior, triggered by native Epx1pp, shortened Epx1pp and MFα1pp as control (expression vector described in Stadlmayr et al., 2010). For achieving the two different Epx1pp constructs, eGFP with respective C-terminal Epx1pp overhangs was amplified by PCR and replaced HSA in the previously used vector.

Green fluorescent protein has been studied and exploited extensively in biochemistry and cell biology since its development in the 1960s. The advantage of GFP's *in vivo* detection makes it a powerful tool for protein localization analysis. Several studies in yeasts concerning secretion of GFP and GFP-fusion proteins were however deflating, as many secretion signals failed to secrete eGFP in *S. cerevisiae* (Li et al., 2002). In *P. pastoris*, the K28pptox was reported to allow eGFP secretion (under control of P_{AOX1}) (Eiden-Plach et al., 2004) as well as the PHA-E signal peptide (Raemaekers et al., 1999); showing no N-terminal extension compared to MFα1pp (Campbell & Choy, 2002). Additionally, eGFP-lipase fusion protein was secreted efficiently under the MFα1pp signal leader, approximately similar to wild type lipase (Passolunghi et al., 2003).

A population of 12 individual clones of each of the three constructs was simultaneously screened in shake-flasks with synthetic-screening-medium for 48 h. The reduced supernatants were analyzed via Western blot (Figure 29/A) and silver stain (Figure 29/B). Apparently, native Epx1pp failed to

secrete eGFP completely (Figure 29, lane 2 to 5), in contrast to Epx1pp-KR and MFα1pp.

Interestingly, only the Epx1pp-KR leader led to correct size eGFP (Figure 29, lane 6 to 9), whereas leftovers of MFα1pp leader led to higher molecular weight eGFP (Figure 29, lane 10 to 13). This was a well-known effect due to remaining $(EA)_2$ overhang of spacer peptide.
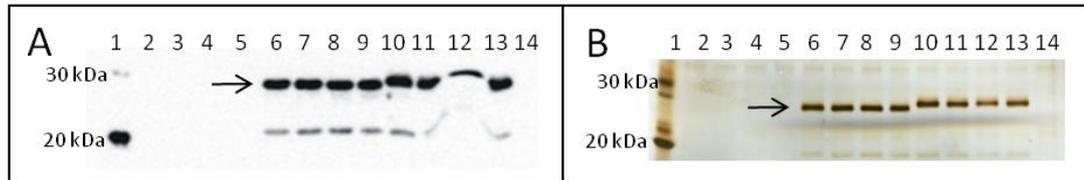
Figure 29: eGFP secretion with three different leader sequences. Shortened Epx1pp secretion leader efficiently secretes eGFP. A: Western blot analysis. B: Silver stain analysis, via reduced SDS-PAGE. L1: MM (A), BM (B), L2-5: pGap_Exp1pp-RT_eGFP, L6-9: pGap_Epx1pp-KR_eGFP, L10-13: pGap_MFα1pp_eGFP, L14: wild type. Arrow indicates eGFP (27 kDa). MFα1pp leader peptide leads to 27,5 kDa eGFP due to $(EA)_2$ overhangs.

5.7.5 Comparative analysis of intracellular eGFP with native and shortened Epx1pp, and MFα1pp secretion leader

Intracellular analysis of eGFP via Western blot (Figure 30) revealed that the native Epx1pp was completely attached to eGFP (Figure 30, lane 2 to 5) even though only low amounts were detected. In contrast, by using the shortened Epx1pp, not all protein was fully processed, but fractions of correctly cleaved eGFP were observable (Figure 30, lane 6 to 9). The MFα1pp was completely cleaved off, except leaving N-terminal amino acids of spacer sequence (Figure 30, lane 10 to 13).
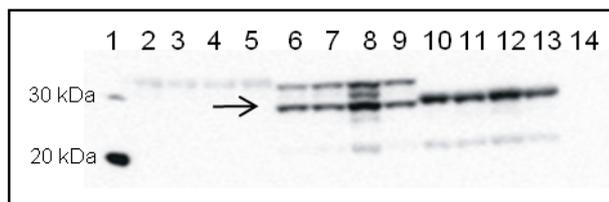
Figure 30: Western blot analysis to detect intracellular eGFP. eGFP retention by three different secretion leaders indicates that Epx1pp-RT and Epx1pp-KR secretion leaders are partially attached to eGFP. L1: MM, L2-5: pGap_Exp1pp-RT_eGFP, L6-9: pGap_Epx1pp-KR_eGFP, L10-13: pGap_MFα1pp_eGFP, L14: wild type. Arrow indicates eGFP (27 kDa). MFα1pp leader peptide leads to 27,5 kDa eGFP due to $(EA)_2$ overhangs.

We performed another approach to study intracellular eGFP and its fluorimetric capability. Therefore, the washed cell pellets of 12 individual clones per construct were

analyzed and measured regarding intracellular fluorescence units per mg wet cell weight (Figure 31). The obtained results are conclusive with intracellular eGFP detected via Western blot, indicating higher intracellular expression levels with the shortened Epx1pp (A: 79.84 FU/mg wcw) compared to native Epx1pp (A: 20.69 FU/mg wcw), The calculated FU/mg wet cell weight of intracellular eGFP with MFα1pp revealed very low levels of intracellular eGFP (A: 5.56 FU/mg wcw).
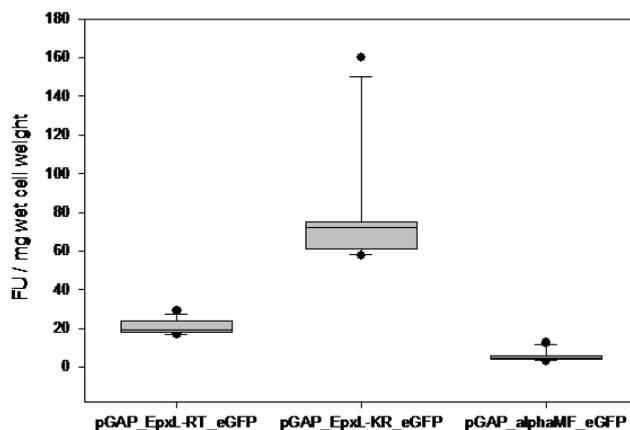
Figure 31: Quantification of intracellular eGFP. Fluorescence units per mg wet cell weight were measured to quantify intracellular eGFP with three different secretion sequences. The shortened Epx1pp secretion leader improves expression compared to the MFα1pp. The box represents 50 % of the calculated values (n=12), and the line indicates the median. The minima and maxima of each construct are given, indicating extreme values as dots.

Immunofluorescence performed by Verena Puxbaum revealed a vacuolar eGFP signal for both the native Epx1pp (Figure 32/A.1) and the MFα1pp (Figure 32/C.1) mediated secretion. This was also reported for *S. cerevisiae*, suggesting an unknown encoded vacuolar signal within eGFP itself (Kunze et al., 1999). Apart from that, eGFP sufficiently entered the secretory pathway, but could not be secreted by native Epx1pp. An intense intracellular fluorescent signal by using the shortened Epx1pp-eGFP construct was observed (Figure B.1), indicating Golgi (often appearing as small dots) or pre-vacuolar compartment stain as well as cytosolic background (Figure 32/B.1). The same localization pattern was reported for *S. cerevisiae*, expressing high levels of intracellular eGFP with acid phosphatase secretion leader, yet without secretion (Li et al., 2002).
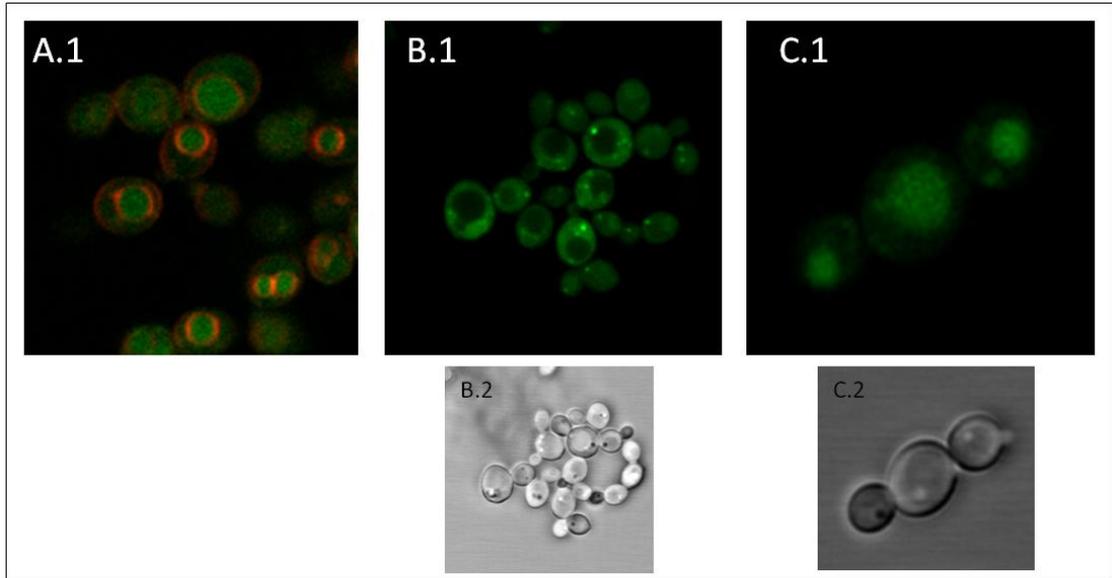
Figure 32: Immunofluorescence staining of eGFP, secreted by three different secretion sequences. Intracellular localization (1) and LCM images (2) of eGFP, expressed by three different secretion leaders. The original Epx1pp leader (Epx1pp-RT) targets eGFP to the vacuole, verified with vacuolar (FM4-64) co- staining (**A**), the shortened Epx1pp leader (Epx1pp-KR) shows weak accumulation in the cytosol (**B**) and the MFα1pp leader targets eGFP also to the vacuole (**C**) (by V. Puxbaum).

6 Discussion

6.1 Transcriptional regulation of Epx1 expression

Epx1 was identified as one of the major native extracellular proteins in *P. pastoris* under various different conditions. Microarray data from different cultivations of *P. pastoris* X-33 revealed that *EPX1* (RPPA09240) is up-regulated at 30°C compared to 20°C, but shows similar expression profiles at 25°C and 30°C. No increase was shown at overexpression of the oxidative stress response transcription factor *YAP1* (Delic et al., in preparation), but the overexpression of the unfolded protein response transcription factor *HAC1* (which is also involved in regulation of cell wall synthesis; Graf et al., 2008) led to two-fold up-regulation (fold change values) and high osmolarity also induced up-regulation (Dragosits et al., 2010).

Searching for transcription factor (TF) binding sites in the putative promoter region (within 1,000 bp upstream of ATG) revealed no conclusive results, but indicated potential binding sites for Stre5 (stress induced TF), Gcn4 (amino acid stress TF), Hstf (heat shock TF) and Mcm1 (TF involved in extracellular matrix remodeling), just to name a few.

Gcn4p was identified to be required for induction of UPR target genes during ER stress, providing further regulation in UPR response and additionally mediating protein degradation under starvation. It was suggested that Hac1p and Gcn4p act together in fine-tuning UPR response (Patil, Li & Walter, 2004). Presumably, Hac1p stabilizes Gcn4p as they form heterodimeric, ternary complexes at a conserved palindromic sequence. Therefore, two-fold up-regulation of *EPX1* observed at *HAC1* overexpression (Graf et al., 2008) is conclusive. Furthermore, these results imply that Epx1 may be involved in stress response.

Some similarities in pro-sequence processing were assumed between Epx1 and yapsins, a class of endopeptidases. Another connection might be their involvement in cell wall remodeling. Yapsins are key players in the cell surface integrity pathway and cell separation (Albrecht et al., 2006). No other results have been obtained for *EPX1* except a Mcm1 transcription factor binding site, which regulates extracellular matrix remodeling.

The promoter analysis revealed that P_{EPX1} is no strong promoter compared to the constitutive P_{GAP} promoter. The 1,000 bp upstream of ATG were additionally studied concerning growth on three different carbon sources. Population analysis of eGFP expression indicated that the tested promoter sequence is not sufficient for high level recombinant protein production.

Additionally, as mentioned in the introduction, high abundance of Epx1 can still be explained (apart from promoter strength) through either high translation efficiency, high protein stability or several other factors. In general, the total amount of native secreted protein was rather low (500 mg) compared to the high biomass (25 g) obtained at cultivation end, and the therefore high intracellular protein levels.

6.2 Impact of native Epx1 secretion leader pro-sequence on recombinant proteins

The native Epx1 pro-sequence comprises 37 amino acids and includes two distinctive amino acid motifs; one dibasic motif (Lys-Arg) and one monobasic motif (Arg-Thr), whereafter the pro-sequence is cleaved on native Epx1 (Figure 15). Each of the three tested recombinant proteins had severe difficulties in leader processing, resulting in the secretion of unwanted fusion proteins and/or ERAD induced degradation. We assumed that correct processing is disturbed by one or both basic motifs.

Porcine trypsinogen was unsuccessfully secreted by native Epx1pp. In comparison to MFα1pp, half as much trypsinogen had been detected in the supernatant, but the protein smear indicated difficulties in cleaving the pro-peptide (Figure 16). Interestingly, enterokinase activation transformed the smear to correct size trypsin by hexapeptide (and remaining pro-sequence) removal (Figure 17). As SDS-PAGE analysis was accomplished with denatured and reduced samples, it is precluded that the smear appears due to incorrect folding and processing of pTRP. Autocatalytic activities of trypsin itself may also contribute to the protein smear.

Human serum albumin was partially processed after the monobasic, predicted leader end, and at the dibasic motif within the Epx1 pro-sequence

(Figure 21). Here, we suppose that the processing speed, respectively the accessibility to the processing site due to secondary structures (discussed below), is responsible for the double band. Immunofluorescence staining revealed only periplasmatic and bud staining, indicating fast and efficient secretion of HSA with the native Epx1 leader. Monobasic sites are slower cleaved, eventually aided by a co-factor which is lost in time (Schwartz, 1986). Therefore it is likely that Kex2 cleaves at dibasic sites in the meantime, leading to secretion of a dibasic processed fraction.

Enhanced green fluorescent protein failed to be secreted by the native Epx1 secretion leader. Interestingly, intracellular measurement of fluorescence units was higher than expected referring to Western blot analysis. Blotting revealed that eGFP was still attached to the native Epx1 secretion leader (Figure 30). This probably impairs structural requirements for folding, leading to vacuolar degradation, but eGFP is also assumed to contain an intramolecular vacuolar sorting signal (Kunze et al., 1999). Consequently, a strong signal was detected by immunofluorescence staining, indicating vacuolar targeting (Figure 32).

It is accepted widely that every leader has different impact on recombinant proteins. Folding is supported or impaired by the secretion leader, and the cleavage site is better or worse accessible for processing enzymes. The amino acid sequence surrounding the monobasic cleavage site is important and different at the C-terminus for every single protein. Devi (Devi, 1991) reported some common rules and tendencies for cleavage at monobasic sites, apart from the proline-directed and non-proline directed postulation by Schwartz (Schwartz, 1986). For all three tested recombinant proteins within our studies, none of those predictions were completely applicable or not.

6.3 Impact of shortened Epx1 secretion leader pro-sequence on recombinant proteins

The shortened Epx1 secretion leader was expected to improve secretion and processing, by generating the Kex2 favored cleavage site (Figure 22). We suspected that by omitting one of the two potential

processing sites, secretion would be enhanced as no concurring processes would interfere.

As expected, the shortened leader enhanced porcine trypsinogen secretion (Figure 24). The N-terminal hexapeptide (Thr-[Asp]$_4$-Lys) might be exposed to the surface, as it is chipped off leading to activation of trypsin, improving adjacent endoproteolytic cleavage by Kex2. This structural feature might simultaneously enhance Kex2 processing as fungal Kex2-like proteins also show a stretch of four Asp residues subsequent to the Lys-Arg motif (Bader et al., 2008). This is also reported for MFα1pp, where processing is improved by addition of a spacer peptide (Brake et al., 1983).

Under the control of P$_{GAP}$, the shortened Epx1 secretion leader is promising (approximately 100 µg trypsinogen per gram wet cell weight) and screenings under the control of P$_{AOX1}$ might enhance secretion even more. Notably in *P. pastoris*, under the regulated control of P$_{AOX1}$, the expression of MFα1pp fused to shrimp trypsinogen yielded 14 mg/L (Guerrero-Olazarán et al., 2009) respectively 10 mg/L bovine trypsinogen, where high protein levels were reached by using the MFα1pp and HSA secretion signal, and by additional removal of enterokinase cleavage site for inactivating bovine trypsinogen (Hanquier et al., 2003).

However, HSA fails to be secreted by the shortened Epx1 secretion leader (Figure 25). This is unexpected, because native HSA leader also ends with a dibasic (Arg-Arg) motif. The dibasic motif (Lys-Arg) of the shortened Epx1 leader is followed by Asp-Ala at P1' and P2' of HSA, which is in fact an optimal Kex2 cleavage site. HSA was rather subjected to cytosolic localization indicating ER-associated degradation, as observed with immunofluorescence staining (Figure 28). A possible explanation for those results might be the hydrophobicity of the pro-sequence, interfering with folding of HSA.

Enhanced green fluorescent protein was successfully secreted by the shortened leader (Figure 29). Western blot analysis revealed that in contrast to MFα1pp, no N-terminal extensions were observable. Measurement of intracellular fluorescence (FU/wcw) suggested correct folding of eGFP, albeit a fraction of intracellular eGFP was located in the cytosol

(immunofluorescence staining). Therefore we propose that the Epx1 secretion leader secretes eGFP strongly and fast, leading to unfolded fractions and UPR induced degradation. Interestingly, in *P. pastoris* eGFP was secreted efficiently only by the PHA-E (Raemaekers et al., 1999), K28pptox (Eiden-Plach et al., 2004) and MFα1pp secretion leader (Campbell & Choy, 2002) so far. Here we report efficient secretion of eGFP by the MFα1pp secretion leader and the shortened Epx1 secretion leader.

6.4 Influence of secretion leaders on protein stability

Secretion leaders do have significant impact on secretion efficiency in yeast (Sreekrishna et al., 1997). Hydrophobic sequences within the pro-sequence are also known to serve as binding signal for ER-resident chaperones (Otte & Barlowe, 2004). Additionally, the correlation of expression and the thermodynamic stability with various signal sequences has been reported for recombinant protein production. Silkworm lysozyme was unstable when using the MFα1pp, but efficiently secreted by its native signal leader. Just four N-terminal amino acids decrease silkworm lysozyme thermodynamic stability by half. Expression levels of human lysozyme using MFα1pp were more than 100 times higher, even though leading to a mixture of several molecular species which correspond to incomplete processing at the C-terminus (Koganesawa et al., 2001).

Less stable proteins are easily degraded by proteases, which causes a decrease in total expression yield and UPR induction (Patil & Walter, 2001). The possible relationship between protein stability in accordance to its secretion signal is likely and should also be taken into consideration for the Epx1 secretion leader.

6.5 Consensus sequence and structural requirements at monobasic and dibasic processing sequences

As it is difficult to make general statements about leader characteristics on structurally and functionally diverse proteins, we tried to find reasonable explanations on Epx1 itself. The native leader led to secreted recombinant

proteins processed at the predicted monobasic leader end and fractions which were cleaved at the dibasic motif. We therefore hypothesize that the native leader incorporates specific features only necessary for Epx1, supported by improved secretion of recombinant proteins after removing the C-terminal part of the pro-sequence after Lys-Arg.

Since not all mono- or dibasic sites are substrate for endoproteolytic cleavage, it is suggested that secondary and tertiary structures are important in generating a conformation, which is recognized by cleavage enzymes. Structural features within the Epx1 pro-peptide are rather recognized by Kex2 endoprotease as is the mature peptide. 20 different prohormones sequences with 53 dibasic potential cleavage sites have been compared, concluding that dibasic motifs, which are located in or next to ß-turns are cleaved (as they are more likely exposed to the surface), whereas location in or next to ß-sheets or α-helices did not allow cleavage (Brakch et al., 1993).

This prediction is most likely valid for Epx1 pro-sequence. The Lys-Arg motif is located within a β-turn (Figure 33), predicted by an empirical technique for the first 60 N-terminal amino acids, developed by Chou and Fasman (Chou & Fasman, 1974) and its improved algorithm (Garnier, Osguthorpe & Robson, 1978). Both algorithms were applied for structural prediction by using the University of Virginia (http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi?rm=misc1) server.

As we started analyzing Epx1pp's secretion efficiency on recombinant proteins, we observed processing difficulties of the pro-sequence. It failed to be cleaved exclusively at the expected monobasic site, leading to an Epx1pp-pTRP fusion protein smear and approximately half of HSA remaining attached to the pro-sequence. The β-turn accounts for the key feature in proteolytic processing so far and, according to the GOR plot in Figure 33/B, Epx1pp pro-sequence includes two β-turns. One was predicted at the dibasic motif, matching with partially processed native Epx1 secretion leader observed with HSA. However, no prediction for monobasic sites at β-

turns were made, but LKP was determined as N-terminal, extracellular end of Epx1, coinciding with this hypothesis (Figure 33/B; red dot).
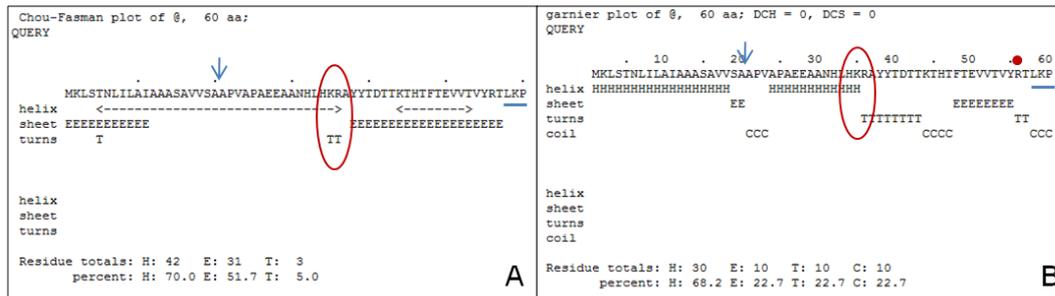
Figure 33: Three-dimensional structure prediction of 60 N-terminal amino acids. A: Chou-Fasman plot. Predicted protein conformation of Epx secretion leader sequence. **B: GOR plot.** Arrow indicates signal peptide cleavage side. Underlined amino acids determine start of mature Epx N-terminus. Red dot indicates monobasic predicted leader end. A dibasic sequence located in turns is predicted to be cleaved (encircled; Brakch et al., 1993).

Additionally, Kex2 processing is favored by His at position P3 and small or negatively charged residues in the P1´, P2´ position; also found in Epx1 pro-sequence. Furthermore, Bader and colleagues reported that denatured substrates were less efficiently processed, suggesting that either the cleavage site is inaccessible or a required secondary structure was lost (Bader et al., 2008).

It was also reported that Kex2 itself (soluble or Golgi persistent) takes part as a saturable limiting sorting component in the secretory pathway, as Kex2 is also involved in sorting to the vacuole via pre-vacuolar compartment. Additionally, secretion of insulin was increased after substituting dibasic sites, as well as by *vps* and *kex2* deletions, suggesting endoproteolytic processing and sorting of Kex2 being the main cause for protein retention and mislocalization to the vacuole (Zhang et al., 2001).

In plants, BiP was suggested to carry an additional vacuolar sorting signal, which enables an adaptor function between ill-defined hydrophobic regions of misfolded proteins with the vacuolar sorting machinery. Vps10 recognizes hydrophobic regions in proteins determined to vacuolar targeting (Pimpl et al., 2006). Furthermore, it has been reported for synthetic leader-mediated insulin precursor (IP) secretion in *S. cerevisiae* that binding of BiP to hydrophobic sequences protects against ERAD and facilitates folding and secretion. The pro-sequence was therefore suspected to function as

intramolecular chaperones mediating pro-peptide assisted folding. An α-helix within the pro-sequence was assumed to fold back upon the IP, thus stabilizing by interaction with the α-helix of the peptide (Kjeldsen et al., 1997). BiP recognizes a motif of bulky hydrophobic and aromatic residues, recognizing a wide variety of polypeptide sequences, which are normally buried in the interior of a folded protein (Blond-Elguindi et al., 1993). Epx1 pro-sequence is highly hydrophobic (Figure 15), but BiP binding probability (according to Blond-Elguindi) needs to be calculated. Apart from that, Epx1 shows an α-helical region within the native pro-sequence (Figure 33), which could also assist in folding.

6.6 Processing possibilities of Epx1 pro-sequence

We suggest two possible roles for the additional dibasic motif within the pro-sequence of Epx1. One possibility might be a two-stage cleavage of the pro-peptide. The first cleavage at the dibasic Lys-Arg motif is performed by Kex2 protease within the late Golgi. Release of Epx1 with pro-sequence into secretory vesicles further allows cleavage by a monobasic specific protease or direct secretion without subsequent cleavage.

The 21 remaining amino acids of the pro-sequence comprise hydrophobic amino acids. Accordingly, this peptide might occupy the hydrophobic cleft within Epx1 during late secretory pathway until dimerization, as also reported for GAPR-1 (human Golgi-associated plant pathogenesis related protein) and yapsins (Gagnon-Arsenault et al., 2006; Serrano et al., 2004b). As reported for pepsin-like aspartic proteinase A, the pro-sequence appears to guide folding of the mature region of PrA (Parr et al., 2007).

Cleavage possibility might be improved by folding the pro-sequence into the hydrophobic cleft of Epx1 and exposing the restriction site to monobasic processing enzymes, such as yapsins. The predicted β-turn at this position is likely to support this activity.

As described in the previous chapter, BiP may bind the N-terminal part of the Epx1 pro-sequence and assist in folding and dimerization. Reported for GAPR-1, the pro-sequence might be removed autocatalytically after

dimerization. Assuming a two-stage cleavage, the C-terminal part of the pro-sequence may also is removed immediately prior to dimerization. The hydrophobic regions of the GAPR-1 monomers are then accessible and promote association. Epx1 is likely to dimerize within the ER, as Epx1 dimers found extracellularly are predicted to be connected via disulfide bonds, but it is unknown whether all of the 4 cysteine residues are involved in dimerization.

A knock out of *EPX1* was attempted during this thesis. Positive clones were indicated through PCR analysis, but could not be isolated. Cells were viable, but had a severely impaired growth. Isolation of *epx1* cells was also impaired as the knock out cassette was (indicated via PCR) sporadically integrated elsewhere, leading to mixed populations. Therefore, the effect of processing in an *epx1*-deletion strain could not be assessed.

Another possibility is that Epx1 is adjusted to cellular needs by differential processing. The specific purpose of Epx1 is not yet known and sequence alignment predicted only C-terminal homology to SCP-family members (Figure 6). Hypothesizing that Epx1 is upregulated under cellular stress (described above) and taken into account that monobasic sites are processed slower compared to dibasic sites, it might be possible that this differentiated processing serves as switch for adjustment.

Aspartic proteases were first described as monobasic and dibasic processing enzymes which additionally compensate dibasic cleavage in *Kex2* deficient yeast (Bourbonnais et al., 1991). A precise distinction between monobasic and dibasic cleavage mechanisms occurs for many precursor proteins in mammalian cells. The same primary translation product can be processed differently in different tissues, e. g. pro-opiomelanocortin (Estivariz, Birch & Loh, 1989), somatostatin (Bourbonnais et al., 1991) and glucagon (Dey et al., 2005). Most prohormones are cleaved in the secretory vesicle at specific sites yielding biologically active peptides (Halban & Irminger, 1994).

On the one hand, Kex2 protease is a saturable enzyme under protein production conditions, leading to partial cleavage at dibasic sites and vacuolar targeting. On the other hand, this might be desirable, as it offers

processing opportunity for monobasic specific endoproteases. Observed for mammalian prohormones, cleavage at either site is tissue specific. Therefore, we do not exclude this hypothesis, even though it is less likely than the assumption described previously.

Notably, the C-terminal amino acids after the Lys-Arg motif of the pro-sequence do not predict a specific sorting signal, as the shortened Epx1 secretion leader also led to secretion.

6.7 Processing of the Epx1 pro-sequence by aspartic proteases (yapsins)

The pro-sequence of Epx1pp as well as of the aspartic-protease family (including yapsins) contain a dibasic and monobasic amino acid motif. Especially for secreting HSA with the native Epx1 secretion leader, two molecular species were observed, and determined to be processed at either side. Yapsins are assumed to process at the monobasic motif. They themselves contain basic motifs within the pro-sequence and the polypeptide. *In vitro* expression analysis of yapsin 1 indicated that yapsin 1 is cleaved and processed additionally after intramolecular monobasic motifs, leading to a heterodimer, stabilized by an intramolecular disulfide bond (Cawley et al., 1998).

Yapsins contain heterogeneous glysocylation and a GPI anchor, as it is a characteristic of yeast periplasmic and cell wall proteins (Ash et al., 1995; Komano et al., 1999). Epx1 is neither glycosylated nor possessing a GPI-anchor. Therefore, processing is suspected to happen in the secretory vesicles. Autocatalytic activation might be possible for Epx1 as it has been proposed that the SCP-domain may functions as endopeptidase (NCBI, conserved domain database, cd05380), but cannot be outlined until studies on native Epx1 have been performed. In contrast, the fungal yapsin Sap9 has been indicated to enclose an autocatalytic processing ability (Albrecht et al., 2006). In comparison to other PR-1 family members, for GAPR-1 it was reported that conserved residues of the putative serine protease catalytic triad are distributed across the two independent molecules of the dimer (Serrano et al., 2004b).

Several studies implicate cleavage of yapsins exclusively C-terminal of a basic residue (Komano et al., 1999). Its negatively charged active pocket tightly interacts with basic residues at P1. As determined by Edman sequencing analysis, Epx1pp pro-sequence was cleaved after Thr at P1, with the basic amino acid Arg at P2. This cleavage is untypical for monobasic motifs, where Arg is dominantly at P1. For yapsin 1, it was reported that a basic residue at P1 is most likely, but cleavage efficiency was enhanced with Arg residues flanking the P1 position (Olsen et al., 1998), as it is the case for Epx1. Therefore we suspect that Epx1 processing at the monobasic motif is mediated by yapsins .

6.8 Conclusion

In this study, we analyzed the effect of the pro-sequence of an unknown extracellular protein on the secretion of three structurally and functionally different recombinant proteins.

Apparently, the results obtained indicate that the native Epx1 secretion leader is not applicable on its own for other proteins as it was not completely chipped off. In this respect, processing of the substrate does not only depend on the amino acid sequence surrounding the processing site, but also on other features such as three dimensional structures.

Within the pro-sequence, a potential Kex2 cleavage site (dibasic Lys-Arg) was identified. Removing 21 amino acids, subsequent to the dibasic Lys-Arg motif, enhanced secretion of two out of three tested proteins without immunogenic N-terminal overhangs. Further studies on Epx1 are necessary to elucidate the function of the function of Epx1 protein and the C-terminal 21 amino acids of its secretion leader. In accordance to the results obtained with the shortened leader, we postulate that the pro-segment is unessential and disturbing for other proteins than Epx1. Moreover, the shortened Epx1 secretion leader is the first described from the SCP-family. Equally important, it enhanced pTRP secretion (14 %) in comparison to the MFα1pp secretion leader and enabled eGFP secretion in *P. pastoris*.

7 References

AHN, J., HONG, J., LEE, H., PARK, M., LEE, E., KIM, C., CHOI, E. & JUNG, J. (2007). Translation elongation factor 1-alpha gene from Pichia pastoris: molecular cloning, sequence, and use of its promoter. *Appl Microbiol Biotechnol* **74,** 601-8.

ALBRECHT, A., FELK, A., PICHOVA, I., NAGLIK, J. R., SCHALLER, M., DE GROOT, P., MACCALLUM, D., ODDS, F. C., SCHÄFER, W., KLIS, F., MONOD, M. & HUBE, B. (2006). Glycosylphosphatidylinositol-anchored proteases of Candida albicans target proteins necessary for both cellular processes and host-pathogen interactions. *J Biol Chem* **281,** 688-94.

ASH, J., DOMINGUEZ, M., BERGERON, J. J., THOMAS, D. Y. & BOURBONNAIS, Y. (1995). The yeast proprotein convertase encoded by YAP3 is a glycophosphatidylinositol-anchored protein that localizes to the plasma membrane. *J Biol Chem* **270,** 20847-54.

BADER, O., KRAUKE, Y. & HUBE, B. (2008). Processing of predicted substrates of fungal Kex2 proteinases from Candida albicans, C. glabrata, Saccharomyces cerevisiae and Pichia pastoris. *BMC Microbiol* **8,** 116.

BERRIN, J. G., WILLIAMSON, G., PUIGSERVER, A., CHAIX, J. C., MCLAUCHLAN, W. R. & JUGE, N. (2000). High-level production of recombinant fungal endo-beta-1,4-xylanase in the methylotrophic yeast Pichia pastoris. In *Protein Expr Purif*, vol. 19, pp. 179-87. 2000 Academic Press., United States.

BLOND-ELGUINDI, S., CWIRLA, S. E., DOWER, W. J., LIPSHUTZ, R. J., SPRANG, S. R., SAMBROOK, J. F. & GETHING, M. J. (1993). Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. In *Cell*, vol. 75, pp. 717-28, United States.

BOURBONNAIS, Y., DANOFF, A., THOMAS, D. Y. & SHIELDS, D. (1991). Heterologous expression of peptide hormone precursors in the yeast Saccharomyces cerevisiae. Evidence for a novel prohormone endoprotease with specificity for monobasic amino acids. *J Biol Chem* **266,** 13203-9.

BOWERS, K. & STEVENS, T. H. (2005). Protein transport from the late Golgi to the vacuole in the yeast Saccharomyces cerevisiae. In *Biochim Biophys Acta*, vol. 1744, pp. 438-54, Netherlands.

BRAKCH, N., RHOLAM, M., BOUSSETTA, H. & COHEN, P. (1993). Role of beta-turn in proteolytic processing of peptide hormone precursors at dibasic sites. *Biochemistry* **32,** 4925-30.

BRAKE, A. J., JULIUS, D. J. & 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, A. J., MERRYWEATHER, J. P., COIT, D. G., HEBERLEIN, U. A., MASIARZ, F. R., MULLENBACH, G. T., URDEA, M. S., VALENZUELA, P. & BARR, P. J. (1984). Alpha-factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **81,** 4642-6.

BRYANT, N. & BOYD, A. (1993). Immunoisolation of Kex2p-containing organelles from yeast demonstrates colocalisation of three processing proteinases to a single Golgi compartment. *J Cell Sci* **106 (Pt 3),** 815-22.

BURCHFIELD, J., LOPEZ, J., MELE, K., VALLOTTON, P. & HUGHES, W. (2010). Exocytotic vesicle behaviour assessed by total internal reflection fluorescence microscopy. *Traffic* **11,** 429-39.

CAMPBELL, T. N. & CHOY, F. Y. (2002). Expression of two green fluorescent protein variants in citrate-buffered media in Pichia pastoris. *Anal Biochem* **311,** 193-5.

CAPLAN, S., GREEN, R., ROCCO, J. & KURJAN, J. (1991). Glycosylation and structure of the yeast MF alpha 1 alpha-factor precursor is important for efficient transport through the secretory pathway. *J Bacteriol* **173,** 627-35.

CAWLEY, N. X., OLSEN, V., ZHANG, C. F., CHEN, H. C., TAN, M. & LOH, Y. P. (1998). Activation and processing of non-anchored yapsin 1 (Yap3p). *J Biol Chem* **273,** 584-91.

CEREGHINO, J. & CREGG, J. (2000). Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiol Rev* **24,** 45-66.

CHAUDHURI, B., LATHAM, S. & STEPHAN, C. (1992). A mutant Kex2 enzyme with a C-terminal HDEL sequence releases correctly folded human insulin-like growth factor-1 from a precursor accumulated in the yeast endoplasmic reticulum. *Eur J Biochem* **210,** 811-22.

CHO, E. Y., CHEONC, S. A., KIM, H., CHOO, J., LEE, D. J., RYU, H. M., RHEE, S. K., CHUNG, B. H., KIM, J.Y. & KANG, A. H. (2010). Multiple-yapsin-deficient mutant strains for high-level production of intact recombinant proteins in Saccharomyces cerevisiae. *Journal of Biotechnology* **149,** 1-7.

CHOU, P. Y. & FASMAN, G. D. (1974). Prediction of protein conformation. *Biochemistry* **13,** 222-45.

CONIBEAR, E. & STEVENS, T. H. (1995). Vacuolar biogenesis in yeast: sorting out the sorting proteins. In *Cell*, vol. 83, pp. 513-6, United States.

CREGG, J., BARRINGER, K., HESSLER, A. & MADDEN, K. (1985). Pichia pastoris as a host system for transformations. *Mol Cell Biol* **5,** 3376-85.

DALY, R. & HEARN, M. (2005). Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production. *J Mol Recognit* **18,** 119-38.

DE ALMEIDA, J. R., DE MORAES, L. M. & TORRES, F. A. (2005). Molecular characterization of the 3-phosphoglycerate kinase gene (PGK1) from the methylotrophic yeast Pichia pastoris. *Yeast* **22,** 725-37.

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.

DEVI, L. (1991). Consensus sequence for processing of peptide precursors at monobasic sites. In *FEBS Lett*, vol. 280, pp. 189-94, Netherlands.

DEY, A., LIPKIND, G. M., ROUILLE, Y., NORRBOM, C., STEIN, J., ZHANG, C., CARROLL, R. & STEINER, D. F. (2005). Significance of prohormone convertase 2, PC2, mediated initial cleavage at the proglucagon interdomain site, Lys70-Arg71, to generate glucagon. In *Endocrinology*, vol. 146, pp. 713-27, United States.

DON, R. H., COX, P. T., WAINWRIGHT, B. J., BAKER, K. & MATTICK, J. S. (1991). 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* **19,** 4008.

DRAGOSITS, M., STADLMANN, J., GRAF, A., GASSER, B., MAURER, M., SAUER, M., KREIL, D. P., ALTMANN, F. & MATTANOVICH, D. (2010). The response to

unfolded protein is involved in osmotolerance of Pichia pastoris. In *BMC Genomics*, vol. 11, pp. 207, England.

EIDEN-PLACH, A., ZAGORC, T., HEINTEL, T., CARIUS, Y., BREINIG, F. & SCHMITT, M. J. (2004). Viral preprotoxin signal sequence allows efficient secretion of green fluorescent protein by Candida glabrata, Pichia pastoris, Saccharomyces cerevisiae, and Schizosaccharomyces pombe. *Appl Environ Microbiol* **70,** 961-6.

ELLIS, S., BRUST, P., KOUTZ, P., WATERS, A., HARPOLD, M. & GINGERAS, T. (1985). Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast Pichia pastoris. *Mol Cell Biol* **5,** 1111-21.

ESTIVARIZ, F. E., BIRCH, N. P. & LOH, Y. P. (1989). Generation of Lys-gamma 3-melanotropin from pro-op... [J Biol Chem. 1989] - PubMed result, vol. 264, J Biol Chem.

GAGNON-ARSENAULT, I., TREMBLAY, J. & BOURBONNAIS, Y. (2006). Fungal yapsins and cell wall: a unique family of aspartic peptidases for a distinctive cellular function. In *FEMS Yeast Res*, vol. 6, pp. 966-78, Netherlands.

GARNIER, J., OSGUTHORPE, D. J. & ROBSON, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of Molecular Biology* **120,** 97-120.

GASSER, B., SALOHEIMO, M., RINAS, U., DRAGOSITS, M., RODRÍGUEZ-CARMONA, E., BAUMANN, K., GIULIANI, M., PARRILLI, E., BRANDUARDI, P., LANG, C., PORRO, D., FERRER, P., TUTINO, M. L., MATTANOVICH, D. & VILLAVERDE, A. (2008). Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microb Cell Fact* **7,** 11.

GETHING, M. & SAMBROOK, J. (1992). Protein folding in the cell. *Nature* **355,** 33-45.

GRAF, A., GASSER, B., DRAGOSITS, M., SAUER, M., LEPARC, G. G., TÜCHLER, T., KREIL, D. P. & MATTANOVICH, D. (2008). Novel insights into the unfolded protein response using Pichia pastoris specific DNA microarrays. *BMC Genomics* **9,** 390.

GUERRERO-OLAZARÁN, M., ESCAMILLA-TREVIÑO, L., CASTILLO-GALVÁN, M., GALLEGOS-LÓPEZ, J. & VIADER-SALVADÓ, J. (2009). Recombinant shrimp (Litopenaeus vannamei) trypsinogen production in Pichia pastoris. *Biotechnol Prog* **25,** 1310-6.

HALBAN, P. A. & IRMINGER, J. C. (1994). Sorting and processing of secretory proteins. *Biochem J* **299,** 1-18.

HAMILTON, S. R., BOBROWICZ, P., BOBROWICZ, B., DAVIDSON, R. C., LI, H., MITCHELL, T., NETT, J. H., RAUSCH, S., STADHEIM, T. A., WISCHNEWSKI, H., WILDT, S. & GERNGROSS, T. U. (2003). Production of complex human glycoproteins in yeast. In *Science*, vol. 301, pp. 1244-6, United States.

HANQUIER, J., SORLET, Y., DESPLANCQ, D., BAROCHE, L., EBTINGER, M., LEFÈVRE, J., PATTUS, F., HERSHBERGER, C. & VERTÈS, A. (2003). A single mutation in the activation site of bovine trypsinogen enhances its accumulation in the fermentation broth of the yeast Pichia pastoris. *Appl Environ Microbiol* **69,** 1108-13.

HASSLACHER, M., SCHALL, M., HAYN, M., BONA, R., RUMBOLD, K., LÜCKL, J., GRIENGL, H., KOHLWEIN, S. & SCHWAB, H. (1997). High-level intracellular expression of hydroxynitrile lyase from the tropical rubber tree Hevea brasiliensis in microbial hosts. *Protein Expr Purif* **11,** 61-71.

HOHENBLUM, H., BORTH, N. & MATTANOVICH, D. (2003). Assessing viability and cell-associated product of recombinant protein producing Pichia pastoris with flow cytometry. *J Biotechnol* **102,** 281-90.

KAWAGUCHI, S., HSU, C. L. & NG, D. T. W. (2010). Interplay of Substrate Retention and Export Signals in Endoplasmic Reticulum Quality Control. *PLoS One* **5**.

KITAJIMA, S. & SATO, F. (1999). Plant pathogenesis-related proteins: molecular mechanisms of gene expression and protein function. *J Biochem* **125,** 1-8.

KJELDSEN, T., PETTERSSON, A., HACH, M., DIERS, I., HAVELUND, S., HANSEN, P. & ANDERSEN, A. (1997). Synthetic leaders with potential BiP binding mediate high-yield secretion of correctly folded insulin precursors from Saccharomyces cerevisiae. *Protein Expr Purif* **9,** 331-6.

KOBAYASHI, K. (2006). Summary of recombinant human serum albumin development. *Biologicals* **34,** 55-9.

KOGANESAWA, N., AIZAWA, T., MASAKI, K., MATSUURA, A., NIMORI, T., BANDO, H., KAWANO, K. & NITTA, K. (2001). Construction of an expression system of insect lysozyme lacking thermal stability: the effect of selection of signal sequence on level of expression in the Pichia pastoris expression system. *Protein Eng* **14,** 705-10.

KOMANO, H., ROCKWELL, N., WANG, G. T., KRAFFT, G. A. & FULLER, R. S. (1999). Purification and characterization of the yeast glycosylphosphatidylinositol-anchored, monobasic-specific aspartyl protease yapsin 2 (Mkc7p). *J Biol Chem* **274,** 24431-7.

KOUTZ, P., DAVIS, G., STILLMAN, C., BARRINGER, K., CREGG, J. & THILL, G. (1989). Structural comparison of the Pichia pastoris alcohol oxidase genes. *Yeast* **5,** 167-77.

KUNZE, I., HENSEL, G., ADLER, K., BERNARD, J., NEUBOHN, B., NILSSON, C., STOLTENBURG, R., KOHLWEIN, S. D. & KUNZE, G. (1999). The green fluorescent protein targets secretory proteins to the yeast vacuole. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1410**.

KURJAN, J. (1985). Alpha-factor structural gene mutations in Saccharomyces cerevisiae: effects on alpha-factor production and mating. *Mol Cell Biol* **5,** 787-96.

KURTZMAN, C. P. (2005). Description of Komagataella phaffii sp. nov. and the transfer of Pichia pseudopastoris to the methylotrophic yeast genus Komagataella. *Int J Syst Evol Microbiol* **55,** 973-6.

KYTE, J. & DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157,** 105-132.

LAROCHE, Y., STORME, V., DE MEUTTER, J., MESSENS, J. & LAUWEREYS, M. (1994). High-level secretion and very efficient isotopic labeling of tick anticoagulant peptide (TAP) expressed in the methylotrophic yeast, Pichia pastoris. *Biotechnology (N Y)* **12,** 1119-24.

LEDUC R. (1992). Activation of human furin precursor processing end... [J Biol Chem. 1992] - PubMed result, vol. 267 (ed. T. B. Molloy SS, Thomas G), pp. 14304-8, J Biol Chem.

LEDUC, R., MOLLOY, S. S., THORNE, B. A. & THOMAS, G. (1992). Activation of human furin precursor processing endoprotease occurs by an intramolecular autoproteolytic cleavage. *J Biol Chem* **267,** 14304-8.

LI, J., XU, H., BENTLEY, W. & RAO, G. (2002). Impediments to secretion of green fluorescent protein and its fusion from Saccharomyces cerevisiae. *Biotechnol Prog* **18,** 831-8.

LI, S. C. & KANE, P. M. (2009). The Yeast Lysosome-like Vacuole: Endpoint and Crossroads. *Biochim Biophys Acta* **1793,** 650-63.

MACAULEY-PATRICK, S., FAZENDA, M. L., MCNEIL, B. & HARVEY, L. M. (2005). Heterologous protein production using the Pichia pastoris expression system. *Yeast* **22,** 249-70.

MARTINEZ-DONATO, G., ACOSTA-RIVERO, N., MORALES-GRILLO, J., MUSACCHIO, A., VINA, A., ALVAREZ, C., FIGUEROA, N., GUERRA, I., GARCIA, J., VARAS, L., MUZIO, V. & DUENAS-CARRERA, S. (2006). Expression and processing of hepatitis C virus structural proteins in Pichia pastoris yeast. In *Biochem Biophys Res Commun*, vol. 342, pp. 625-31, United States.

MARTINEZ-RUIZ, A., MARTINEZ DEL POZO, A., LACADENA, J., MANCHENO, J. M., ONADERRA, M., LOPEZ-OTIN, C. & GAVILANES, J. G. (1998). Secretion of recombinant pro- and mature fungal alpha-sarcin ribotoxin by the methylotrophic yeast Pichia pastoris: the Lys-Arg motif is required for maturation. In *Protein Expr Purif*, vol. 12, pp. 315-22. 1998 Academic Press., United States.

MATTANOVICH, D., CALLEWAERT, N., ROUZÉ, P., LIN, Y. C., GRAF, A., REDL, A., TIELS, P., GASSER, B. & DE SCHUTTER, K. (2009a). Open access to sequence: browsing the Pichia pastoris genome. *Microb Cell Fact* **8,** 53.

MATTANOVICH, D., GASSER, B., HOHENBLUM, H. & SAUER, M. (2004). Stress in recombinant protein producing yeasts. *J Biotechnol* **113,** 121-35.

MATTANOVICH, D., GRAF, A., STADLMANN, J., DRAGOSITS, M., REDL, A., MAURER, M., KLEINHEINZ, M., SAUER, M., ALTMANN, F. & GASSER, B. (2009b). Genome, secretome and glucose transport highlight unique features of the protein production host Pichia pastoris. *Microb Cell Fact* **8,** 29.

MAURER, M., KUHLEITNER, M., GASSER, B. & MATTANOVICH, D. (2006). Versatile modeling and optimization of fed batch processes for the production of secreted heterologous proteins with Pichia pastoris. *Microb Cell Fact* **5,** 37.

MAYER, M., KIES, U., KAMMERMEIER, R. & BUCHNER, J. (2000). BiP and PDI cooperate in the oxidative folding of antibodies in vitro. *J Biol Chem* **275,** 29421-5.

MENENDEZ, J., VALDES, I. & CABRERA, N. (2003). The ICL1 gene of Pichia pastoris, transcriptional regulation and use of its promoter. *Yeast* **20,** 1097-108.

NG, D. T., BROWN, J. D. & WALTER, P. (1996). Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J Cell Biol* **134,** 269-78.

OKA, C., TANAKA, M., MURAKI, M., HARATA, K., SUZUKI, K. & JIGAMI, Y. (1999). Human lysozyme secretion increased by alpha-factor pro-sequence in Pichia pastoris. *Biosci Biotechnol Biochem* **63,** 1977-83.

OLSEN, V., GURUPRASAD, K., CAWLEY, N. X., CHEN, H. C., BLUNDELL, T. L. & LOH, Y. P. (1998). Cleavage efficiency of the novel aspartic protease yapsin 1 (Yap3p) enhanced for substrates with arginine residues flanking the P1 site: correlation with electronegative active-site pockets predicted by molecular modeling. In *Biochemistry*, vol. 37, pp. 2768-77, United States.

OTTE, S. & BARLOWE, C. (2004). Sorting signals can direct receptor-mediated export of soluble proteins into COPII vesicles. In *Nat Cell Biol*, vol. 6, pp. 1189-94, England.

PAREKH, R., FORRESTER, K. & WITTRUP, D. (1995). Multicopy overexpression of bovine pancreatic trypsin inhibitor saturates the protein folding and secretory capacity of Saccharomyces cerevisiae. *Protein Expr Purif* **6,** 537-45.

PARR, C. L., KEATES, R. A., BRYKSA, B. C., OGAWA, M. & YADA, R. Y. (2007). The structure and function of Saccharomyces cerevisiae proteinase A. *Yeast* **24,** 467-80.

PASSOLUNGHI, S., BROCCA, S., CANNIZZARO, L., PORRO, D. & LOTTI, M. (2003). Monitoring the transport of recombinant Candida rugosa lipase by a green fluorescent protein-lipase fusion. *Biotechnol Lett* **25,** 1945-8.

PATIL, C. & WALTER, P. (2001). Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. In *Curr Opin Cell Biol*, vol. 13, pp. 349-55, United States.

PATIL, C. K., LI, H. & WALTER, P. (2004). Gcn4p and Novel Upstream Activating Sequences Regulate Targets of the Unfolded Protein Response. *PLoS Biol* **2**.

PIMPL, P., TAYLOR, J. P., SNOWDEN, C., HILLMER, S., ROBINSON, D. G. & DENECKE, J. (2006). Golgi-Mediated Vacuolar Sorting of the Endoplasmic Reticulum Chaperone BiP May Play an Active Role in Quality Control within the Secretory Pathway. *Plant Cell* **18,** 198-211.

PORRO, D., SAUER, M., BRANDUARDI, P. & MATTANOVICH, D. (2005). Recombinant protein production in yeasts. *Mol Biotechnol* **31,** 245-59.

RABILLOUD, T. (1999). Silver staining of 2-D electrophoresis gels. *Methods Mol Biol* **112,** 297-305.

RAEMAEKERS, R. J., DE MURO, L., GATEHOUSE, J. A. & FORDHAM-SKELTON, A. P. (1999). Functional phytohemagglutinin (PHA) and Galanthus nivalis agglutinin (GNA) expressed in Pichia pastoris correct N-terminal processing and secretion of heterologous proteins expressed using the PHA-E signal peptide. In *Eur J Biochem*, vol. 265, pp. 394-403, Germany.

RAKESTRAW, J., SAZINSKY, S., PIATESI, A., ANTIPOV, E. & WITTRUP, K. (2009). Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in Saccharomyces cerevisiae. *Biotechnol Bioeng* **103,** 1192-201.

ROBERTS, C. J., NOTHWEHR, S. F. & STEVENS, T. H. (1992). Membrane protein sorting in the yeast secretory pathway: evidence that the vacuole may be the default compartment. *J Cell Biol* **119,** 69-83.

SCHWARTZ, T. W. (1986). The processing of peptide precursors. 'Proline-directed arginyl cleavage' and other monobasic processing mechanisms. In *FEBS Lett*, vol. 200, pp. 1-10, Netherlands.

SEARS, I. B., O'CONNOR, J., ROSSANESE, O. W. & GLICK, B. S. (1998). A versatile set of vectors for constitutive and regulated gene expression in Pichia pastoris. In *Yeast*, vol. 14, pp. 783-90, England.

SERRANO, R., KUHN, A., HENDRICKS, A., HELMS, J., SINNING, I. & GROVES, M. (2004a). Structural analysis of the human Golgi-associated plant pathogenesis related protein GAPR-1 implicates dimerization as a regulatory mechanism. *J Mol Biol* **339,** 173-83.

SERRANO, R. L., KUHN, A., HENDRICKS, A., HELMS, J. B., SINNING, I. & GROVES, M. R. (2004b). Structural analysis of the human Golgi-associated plant

pathogenesis related protein GAPR-1 implicates dimerization as a regulatory mechanism. *J Mol Biol* **339,** 173-83.

SHEN, S., SULTER, G., JEFFRIES, T. W. & CREGG, J. M. (1998). A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast Pichia pastoris. In *Gene*, vol. 216, pp. 93-102, Netherlands.

SLEEP, D., BELFIELD, G. P. & GOODEY, A. R. (1990). The secretion of human serum albumin from the yeast Saccharomyces cerevisiae using five different leader sequences. *Biotechnology (N Y)* **8,** 42-6.

SREEKRISHNA, K., BRANKAMP, R. G., KROPP, K. E., BLANKENSHIP, D. T., TSAY, J. T., SMITH, P. L., WIERSCHKE, J. D., SUBRAMANIAM, A. & BIRKENBERGER, L. A. (1997). Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast Pichia pastoris. *Gene* **190,** 55-62.

STADLMAYR, G., MECKLENBRAUKER, A., ROTHMULLER, M., MAURER, M., SAUER, M., MATTANOVICH, D. & GASSER, B. (2010). Identification and characterisation of novel Pichia pastoris promoters for heterologous protein production. In *J Biotechnol*, vol. 150, pp. 519-29. 2010 Elsevier B.V, Netherlands.

STEINER, D. F., SMEEKENS, S. P., OHAGI, S. & CHAN, S. J. (1992). The new enzymology of precursor processing endoproteases. *J Biol Chem* **267,** 23435-8.

TU, B. P. & WEISSMAN, J. S. (2004). Oxidative protein folding in eukaryotes: mechanisms and consequences. *J Cell Biol* **164,** 341-6.

VALDES, I., HERMIDA, L., ZULUETA, A., MARTIN, J., SILVA, R., ALVAREZ, M., GUZMAN, M. G. & GUILLEN, G. (2007). Expression in Pichia pastoris and immunological evaluation of a truncated Dengue envelope protein. In *Mol Biotechnol*, vol. 35, pp. 23-30, United States.

VAN ANKEN, E. & BRAAKMAN, I. (2005). Versatility of the endoplasmic reticulum protein folding factory. *Crit Rev Biochem Mol Biol* **40,** 191-228.

VANDERMEERS. (1972). Simple Automated Spectrophotrometric Method for Assay of Trypsin and Chymotrypsin in Duodenal Juice, vol. CLINICAL CHEMISTRY, Vol. 18. No. 12, 1972.

VOOS, W. & STEVENS, T. H. (1998). Retrieval of resident late-Golgi membrane proteins from the prevacuolar compartment of Saccharomyces cerevisiae is dependent on the function of Grd19p. *J Cell Biol* **140,** 577-90.

WALTER, P. & JOHNSON, A. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu Rev Cell Biol* **10,** 87-119.

WATERHAM, H., DIGAN, M., KOUTZ, P., LAIR, S. & CREGG, J. (1997). Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* **186,** 37-44.

WEBER, K. & OSBORN, M. (1969). The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. *J Biol Chem* **244,** 4406-12.

WILCOX, C. A. & FULLER, R. S. (1991). Posttranslational processing of the prohormone-cleaving Kex2 protease in the Saccharomyces cerevisiae secretory pathway. *J Cell Biol* **115,** 297-307.

XIONG, R. & CHEN, J. (2008). Secreted expression of human lysozyme in the yeast Pichia pastoris under the direction of the signal peptide from human serum albumin. *Biotechnol Appl Biochem* **51,** 129-34.

Yamada, Y., Matsuda, M., Maeda, K. & Mikata, K. (1995). The phylogenetic relationships of methanol-assimilating yeasts based on the partial sequences of 18S and 26S ribosomal RNAs: the proposal of Komagataella gen. nov. (Saccharomycetaceae). *Biosci Biotechnol Biochem* **59,** 439-44.

Yamamoto, Y., Taniyama, Y., Kikuchi, M. & Ikehara, M. (1987). Engineering of the hydrophobic segment of the signal sequence for efficient secretion of human lysozyme by Saccharomyces cerevisiae. *Biochem Biophys Res Commun* **149,** 431-6.

Zhang , B., Chang, A., Kjeldsen, T. & Arvan, P. (2001). Intracellular retention of newly synthesized insulin in yeast is caused by endoproteolytic processing in the Golgi complex. *J Cell Biol* **153,** 1187-98.

Zhao, H., He, Q., Xue, C., Sun, B., Yao, X. & Liu, Z. (2009). Secretory expression of glycosylated and aglycosylated mutein of onconase from Pichia pastoris using different secretion signals and their purification and characterization. *FEMS Yeast Res* **9,** 591-9.

Zhu, H., Yao, S. & Wang, S. (2010). MFalpha signal peptide enhances the expression of cellulase eg1 gene in yeast. *Appl Biochem Biotechnol* **162,** 617-24.

8 Appendix

8.1 Abbreviations

°C	degree centigrade
μ	micro
A	average
AP	alkaline phosphatase
bp	base pair(s)
BSA	bovine serum albumin
ddH$_2$O	reverse osmosis water
dNTP	deoxynucleoside triphosphate
EDTA	ethylenediaminetetraacetic acid
et al.	and others
g	gram
g	gravitational force
h	hour(s)
HEPES	4-(2-hydroxyethyl)-1- piperazineethanesulfonic acid
HRP	horse radish peroxidase
IgG	immunoglobulin G
kb	kilo base(s)
kDa	kilo Dalton
L	liter
LB	lysogeny broth
LDS	lithium dodecyl sulfate
M	mol
m	milli or meter
min	minute(s)
MOPS	3-(N-morpholino)propanesulfonic acid
N	normality (mol/L)
n	nano
OD	optical density

OPD	*o*-phenylenediamine
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS/PBST	phosphate buffered saline/ plus Tween20
pI	isoelectric point
rpm	revolutions per minute
RT	room temperature
s	second(s)
SD	standard deviation
SDS	sodium dodecyl sulfate
SOC	super optimal broth with catabolite repression
TAE	Tris base, acetic acid and EDTA
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
U	unit(s) of activity
UV	ultra violet
wcw	wet cell weight
wt	wild type
x	-fold
YPD	yeast extract peptone dextrose
α	alpha or anti

8.2 List of Figures

8.3 Zusammenfassung

Die methylotrophe Hefe *Pichia pastoris* gilt als beliebtes Expressionssystem zur Produktion von rekombinanten Proteinen. Die Verfügbarkeit sowohl von starken Promotoren (z.B. der Glyceraldehyd-3-Phosphat Dehydrogenase Promoter, P_{GAP} und den Promoter der Alkohol Oxidase, P_{AOX1}) und effizienten Sekretionssignalen (alpha mating factor Sekretionssignal; MF1αpp), als auch eine funktionierende Proteinfaltung und das Anfügen von post-translationalen Modifikationen sind vorteilhaft.

Das extrazelluläre Protein X (Epx1) wurde als stark sekretiertes, jedoch unbekanntes Protein im Überstand von *P. pastoris* bemerkt. Wenig ist bekannt über Epx1, außer dass es eines von 20 endogen sekretierten Proteinen ist und der C-Terminus mit einer SCP-ähnlichen (sekretiert und Cystein-reich) extrazellulären Proteindomäne überein stimmt. Die Anwendbarkeit als neues Sekretionssystem für rekombinante Proteine wurde anhand von intrazellulärer eGFP Expression durch den Promoter (1.000 bp oberhalb des ATG) auf drei unterschiedlichen Kohlenstoffquellen getestet. P_{EPX1} zeigte keine verbesserte eGFP Sekretion, verglichen mit P_{GAP}. Weiters wurde das Epx1 Sekretionssignal mit drei strukturell und funktionell unterschiedlichen Proteinen analysiert (pTRP, porcines Trypsinogen; HSA, humanes Serumalbumin; eGFP, verbessertes grün-fluoreszierendes Protein).

Die Sekretionssequenz von Epx1 wurde teilweise nicht korrekt abgespalten. 21 Aminosäuren der Pro-Sequenz wurden, auf Grund von unerwarteter Spaltung an einer internen Sequenz (zwei basische Aminosäuren), nicht abgespalten und blieben am N-Terminus. Wir verkürzten die Sequenz und erzielten (im Vergleich zu MFα1pp) eine Steigerung der pTRP Sekretion, als auch Sekretion von eGFP ohne N-terminalen Überhang. Zusammenfassend haben wir durch Verkürzung der Epx1-Pro-Sequenz eine neue, effiziente Sekretionssequenz geschaffen.

8.4 Curriculum Vitae

30.11.1985

SILVIA CHRISTINA HEISS

Rum in Tirol

silviaheiss@hotmail.com

EDUCATION

2001 - 2004	BORG Innsbruck (focus on natural sciences)
2004 -2011	**Molecular Biology**, University of Vienna

PRACTICAL EXPERIENCE

09/2009- 12/2010	Diploma thesis at the University of Natural Resources and Life Sciences group: Dr. Diethard Mattanovich; **Establishment and characterization of a novel secretion system in the methylotrophic yeast *Pichia pastoris***
03-05/2009	University of Vienna; Institute of Microbiology and Genetics group: Dr. Angela Witte; **Identification and application of *pha*- genes in *Haloarcula hispanica* and *H. marismortui***
07-09/2008	University of Münster; Institute of Molecular Microbiology and Biotechnology group: Dr. Alexander Steinbüchel; **Proteome analysis of *Ralstonia eutropha* H16 and *R. eutropha* G+1, its glucose utilizing mutant strain**
04-06/2008	Medical University Innsbruck; Division of Hygiene and Medical Microbiology group: Dr. Cornelia Speth; **Examining the pathogenity of exogene proteases of *Pseudallescheria* and *Scedosporium* in human liquor**
02/2007	Böhringer Ingelheim- Austria; **marketing department**
07-09/2006	Laboratories Mühl-Speiser; **medical-diagnostic service**
07-09/2005	Böhringer Ingelheim-Austria; **quality assurance**

PUBLICATIONS

Proteomic and transcriptomic elucidation of the spontaneous mutant *Ralstonia eutropha* G$^+$1 with regard to glucose utilization; Raberg M, Peplinski K, Heiss S, Ehrenreich A, Voigt B, Döring C, Bömeke M, Hecker M, Steinbüchel A; Appl Environ Microbiol. 2011 Mar; 77(6):2058-70.

8.5 Acknowledgements

The most exciting phrase to hear in
science, the one that heralds new
discoveries, is not "Eureka!" ("I found
it!") but rather, "Hmmm… That's
funny…"
ISAAC ASIMOV