Multistep processing of the secretion leader of the extracellular protein Epx1 in *Pichia pastoris* and implications for protein localization

Silvia Heiss,1 Verena Puxbaum,1,2 Clemens Gruber,2,3 Friedrich Altmann,2,3 Diethard Mattanovich1,2 and Brigitte Gasser1,2

1Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria
2Austrian Centre of Industrial Biotechnology (ACIB GmbH), Vienna, Austria
3Department of Chemistry, University of Natural Resources and Life Sciences, Vienna, Austria

Secretion leaders are required to direct nascent proteins to the secretory pathway. They are of interest in the study of intracellular protein transport, and are required for the production of secretory recombinant proteins. Secretion leaders are processed in two steps in the endoplasmic reticulum and Golgi. Although yeast cells typically contain about 150 proteins entering the secretory pathway, only a low number of proteins are actually secreted to the cell supernatant. Analysis of the secretome of the yeast *Pichia pastoris* revealed that the most abundant secretory protein, which we named Epx1, belongs to the cysteine-rich secretory protein family CRISP. Surprisingly, the Epx1 secretion leader undergoes a three-step processing on its way to the cell exterior instead of the usual two-step processing. The Kex2 cleavage site within the *P. pastoris* Epx1 leader is not conserved in the homologues of most other yeasts. We studied the effect of exchanging the Kex2-cleavage motif on the secretory behaviour of reporter proteins fused to variants of the Epx1 leader sequence, and observed mistargeting for some but not all of the variants using fluorescence microscopy. By targeting several recombinant human proteins for secretion, we revealed that a short variant of the leader sequence, as well as the Epx1 signal sequence alone, resulted in the correct N-termini of the secreted proteins. Both leader variants proved to be very efficient, even exceeding the secretion levels obtained with commonly used secretion leaders. Taken together, the novel Epx1 secretion leader sequences are a valuable tool for recombinant protein production as well as basic research of intracellular transport.

INTRODUCTION

Polypeptides, which are to be secreted, are translocated into the endoplasmic reticulum (ER). This is facilitated by a short hydrophobic N-terminal signal peptide. After cleavage of the signal peptide by an ER-resident signal peptidase, and correct folding by chaperones and foldases, the proteins are then transported to the Golgi network. Subsequently, proteins are delivered to their final cellular location, which may be the ER, Golgi, secretory vesicles, peroxisomes, endosomes, vacuole, cell wall or the cell exterior (recently reviewed by Delic et al., 2013). The route through the secretory pathway requires as a minimum an N-terminal signal peptide at the nascent polypeptide, which allows for co- or post-translational translocation from the cytosol to the ER lumen and typically consists of 13 to 36 mostly hydrophobic amino acids (pre-sequence) (Ng et al., 1996; Zimmermann et al., 2011).

A portion of secretory proteins carry a more complex type of secretion leaders consisting of pre-prosequences (signal peptide and leader sequence), e.g. the precursors of the *Saccharomyces cerevisiae* mating factor alpha (MFα) or carboxypeptidase Y (CPY) (Johnson et al., 1987; Julius et al., 1984; Valls et al., 1987; Waters et al., 1988). These pro-peptides have been reported to possess chaperone activity in some cases and are often present to prevent pre-activation of proteins before they reach their final cellular destinations (as is the case for vacuolar proteases, which are produced as inactive pre-forms and become activated by cleavage of their pro-peptides) (Bryant & Stevens, 1998; Klionsky et al., 1990; Parr et al., 2007).

Abbreviations: AP, alkaline phosphatase; ER, endoplasmic reticulum; HC, heavy chain; HSA, human serum albumin; LC, light chain; pTRP, porcine trypsinogen.

One supplementary figure is available with the online Supplementary Material.
Although there is no consensus sequence, signal peptides have a common structure consisting of a short positively charged N-terminal region, a central hydrophobic region (h-region) and a more polar C-terminal region containing the site that is cleaved by the ER-resident signal peptidase (Martoglio & Dobberstein, 1998; von Heijne, 1984). After passing ER quality control, correctly folded and processed proteins are allowed to exit the ER. Secreted proteins are then transported from the Golgi network to the plasma membrane for final secretion to the cell exterior. The leader pro-peptide is cleaved off the protein by (presumably) late Golgi-resident proteases, such as Kex2 protease of S. cerevisiae (Bryant & Boyd, 1993; Fuller et al., 1989).

During the passage through the secretory pathway at least two quality control systems are in place, assuring that only correctly folded and stable proteins are allowed to reach their final destinations. The first quality control system governs the exit from the ER towards the Golgi (reviewed by Delic et al., 2013 and Kleizen & Braakman, 2004), and redirects terminally misfolded proteins to the proteasome for degradation (ER-associated protein degradation) (recently reviewed by Ruggiggio et al., 2014 and Thibault & Ng, 2012). The second quality control step is located in the late Golgi, where proteins that do not fulfil the criteria are targeted to the vacuole for subsequent degradation (Bryant & Stevens, 1998; Holkeri & Makarow, 1998). At present it is not yet clear which criteria determine this distinction, although both thermodynamic instability as well as prolonged binding to Kex2-like enzymes were reported to contribute to vacuolar sorting (Coughlan et al., 2004; Zhang et al., 2001).

Secretion leaders are not only of interest in the study of intracellular protein transport, but are also required for the production of secretory recombinant proteins (Emr et al., 1983; Zsebo et al., 1986). Yeasts are commonly used host organisms for production of a wide range of heterologous proteins, such as enzymes, vaccines, hormones and biopharmaceutical proteins. Most of these proteins are secreted in their native environment; thus, should preferably also be secreted by the recombinant expression host (Macaully-Patrick et al., 2005; Mattanovich et al., 2012). The methylotrophic yeast Pichia pastoris (synonym Komagataella sp.; Kurtzman, 2005) 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 organelle biosynthesis (Dunn et al., 2005; Suda & Nakano, 2012). Unravelling secretion signals, ER folding and transport to the plasma membrane are, therefore, of paramount interest; Damasceno et al., 2012). Although yeast cells typically contain about 150 proteins carrying a signal peptide, only a low number (fewer than 50 proteins) are actually secreted to the cell supernatant (Mattanovich et al., 2009). Even though P. pastoris is a very efficient host for a diverse range of heterologous proteins, there is a deficiency in useful secretion leaders (Damasceno et al., 2012; Gasser et al., 2013).

By far the most-used and best-studied secretion leader in yeast is the MFz pre-pro-leader originating from S. cerevisiae, which has been used for production of secretory proteins with and without an EAEA (Glu-Ala)2 overhang in various yeast species (Daly & Hearn, 2005; Ghosalkar et al., 2008; Mattanovich et al., 2012). Unfortunately, both variants of MFz suffer from inefficient processing by Kex2 and/or Ste13, which leaves unwanted overhangs at the N-termini of the produced recombinant proteins and leads to heterogeneous products (among others: Almeida et al., 2001; Hashimoto et al., 1998; Steinlein et al., 1995; Zhao et al., 2009). While in some cases those overhangs resulted in improved protein stability (Schaefer & Plückthun, 2012), mostly they are regarded as undesirable and unacceptable contaminants, especially for biopharmaceutical products. Further secretion leaders (mostly derived from heterologous eukaryotic systems) have been tested in P. pastoris (Gasser et al., 2013), but do not achieve similar levels of secretion when compared to the MFz leader, apart from some particular exceptions where native signal sequences are preferred (e.g. for human serum albumin (HSA), human lysozyme, some fungal enzymes – see Gasser et al. 2013). Earlier this year, two publications employing signal sequences only (Fitzgerald & Glick, 2014; Govindappa et al., 2014), as well as a mutant version of MFz (Lin-Cereghino et al., 2013), have been reported. Reasons for differences in secretion rates are manifold, but processing of the secretion leader, thermodynamic instability, misfolding, as well as misdirecting, are obvious; however, not yet understood in full detail.

Analysis of the secretome of P. pastoris revealed that the most abundant secretory protein Epx1 (extracellular protein X) belongs to the secretory cysteine-rich protein family (CRISP), which is conserved among fungal species and other eukaryotes (Heiss et al., 2013; Mattanovich et al., 2009). Recently the function of the Epx1 homologues in S. cerevisiae (Pry1 and Pry2) was identified to be required for secretion of cholesteryl acetate from the cell in order to keep the bound lipid soluble in the extracellular environment (Choudhary & Schneiter, 2012). We have previously shown that the knockout of Epx1 in P. pastoris has no effect on protein secretion or growth in standard conditions, and that the high secretion levels of the Epx1 protein are not due to high transcription of the EPX1 gene (Heiss et al., 2013). Therefore, we assumed that the Epx1 leader sequence must be responsible for the strong extracellular protein abundance.

In the present work, we have characterized different variants of the pre-prosequence of Epx1 for their efficiency to secrete reporter proteins and their impact on intracellular protein targeting. Little is known about how the combination of secretion leader and peptide influences the targeting within or outside the cell. Therefore, we

http://mic.sgmjournals.org
combined different detection methods, such as fluorescence microscopy, immunoblotting and quantitative analysis of extracellular reporter proteins, to gain better understanding about how *P. pastoris* targets recombinant proteins, based upon their secretion leader, within the cell. Elucidating the underlying mechanisms of secretion processes helps to better understand the model organism *P. pastoris* and leads to improved secretion of target proteins.

**METHODS**

**Construction of expression vectors using EpxL-RT for secretion.** Based on the experimentally determined N-terminus of the natively secreted Epx1 protein (Heiss et al., 2013), we constructed expression plasmids carrying the putative Epx1 secretion leader sequence (consisting of the Epx1 signal sequence and prosequence up to the experimentally determined N-terminus of the mature Epx1 protein) fused to different reporter proteins. As the last amino acids preceding the experimentally verified N-terminus of mature Epx1 were Arg-Thr (RT), the putative secretion leader sequence was termed EpxL-RT (Fig. 1a). This secretion leader was amplified from *P. pastoris* genomic DNA using the primers EpxL-RT_F and EpxL-RT_R (Table 1), and cloned into a linearized pPM2dZ10_pGAP (glyceraldehyde 3-phosphate dehydrogenase) vector (Stadlmayr et al., 2010). Subsequently, genes encoding (EGFP, enhanced green fluorescent protein), HSA or porcine trypsinogen (pTRP) were amplified by PCR from vectors described by Stadlmayr et al. (2010) and Baumann et al. (2008) using primers shown in Table 1, and cloned in-frame with EpxL-RT. After sequence verification, the obtained vectors pPM2dZ10_pGAP_EpxL-RT-EGFP, pPM2dZ10_pGAP_EpxL-RT_pTRP and pPM2dZ10_pGAP_EpxL-RT-HSA were linearized with the restriction enzyme AvaII in the GAP promoter region to allow targeted integration into the respective locus in the *P. pastoris* genome. Transformation of electrocompetent *P. pastoris* was performed according to our previous protocol (Gasser et al., 2013). Correct insertion was tested by colony PCR for random samples. As a reference, a construct with the native HSA leader (Stadlmayr et al., 2010) was used for HSA, while GFP and pTRP were fused to the *S. cerevisiae* MF2 pre-pro-leader (Table 2).

**Construction of expression vectors containing variants of EpxL.** The importance and influence of the Lys-Arg dibasic motif on secretory behaviour was investigated by exchanging the amino acids KR for different pairs of amino acids (either QN, TV or SH; Fig. 1c). The exchange within the prosequence was performed with the primers EpxL-KR :: QN_F and EpxL-KR :: TV_F and EGFP_R (Table 1) into the *Pvu*I and *Sfi*I digested vector pPM2dZ10_pGAP_EpxL-RT_x_EGFP. The HSA and pTRP constructs were established by using the *Acc*I and *Sfi*I cleavage sites to replace EGFP.

Additionally, vectors were constructed using the EpxL signal sequence and prosequence up to the Lys-Arg (KR) dibasic motif (Fig. 1d) using the following primers (Table 1). HSA was amplified using EpxL-KR_HSA_F and HSA_R, and ligated into the *Bgl*II and *Sfi*I digested vector. pTRP was amplified by using EpxL-KR_pTRP_F and pTRP_R, and ligated into the *Pvu*I and *Sfi*I digested vector. The expression vector pPM2dZ10_pGAP_EpxL-KR_EGFP was generated in the same way with EpxL-KR_EGFP_F and EGFP_R (Table 1). The light chain (LC) and the heavy chain (HC) of the HyHEL-10 antibody were amplified using EpxL-KR_LC_F and LC_R or EpxL-KR_HC_F and HC_R, and ligated into the *Bgl*II and *Sfi*I digested vector.

We also constructed vectors containing just the short signal peptide of the native Epx1 leader (Fig. 1e), termed EpxL-SP. Antibody HyHEL-10, Fab-HC and EGFP were amplified using EpxL-SP_LC_F, EpxL-SP_eGFP_F and EpxL-SP_FabHC_F (Table 1), and ligated into the *Bgl*II and *Sfi*I digested expression vector. All plasmids were linearized in the GAP promoter prior to transformation into *P. pastoris*.

**Cultivation of P. pastoris.** After transformation, positive *P. pastoris* transformants were selected on YPD agar plates containing 25 μg Zeocin ml⁻¹. For analysis of protein secretion, *P. pastoris* was inoculated overnight in 5 mlYP medium (in 1 l: 10 g yeast extract, 20 g peptone, 20 g glucose and 25 μg Zeocin ml⁻¹) at 28 °C. Aliquots of these cultures (corresponding to a final OD₆₀₀ of 0.1) were transferred to 10 ml expression culture medium and incubated for 48 h at 28 °C at 170 r.p.m. in 100 ml Erlenmeyer flasks. Every 12 h cells were fed with 100 μl 50 × glucose (in total 3.5 % glucose). Alternatively, 2 ml expression culture medium was used for cultivation in 24-deep-well plates.

Synthetic-screening medium for pTRP and HSA contained in 1 l: 22 g glucose monohydrate, 22 g citric acid, 3.15 g (NH₄)₂HPO₄, 0.027 g

---

**Fig. 1.** Amino acid sequence of Epx1 secretion leader and variants. Name-giving amino acids are indicated in bold. (a) Experimentally determined native Epx1 secretion leader and prosequence, the first N-terminal amino acids of the mature Epx1 protein are in italics. (b) EpxL-RT full native Epx1 secretion leader, (c) EpxL-KR :: QN, EpxL-KR :: TV, full secretion leader where KR (Lys-Arg) is exchanged to QN (Gln-Asn) or TV (Thr-Val). (d) EpxL-KR shortened leader with only prosequence 1. (e) EpxL-SP signal peptide. The signal peptidase cleavage sites are indicated by arrows 1 and the predicted Kex2 cleavage site indicated by arrow 2. The determined end of the Epx1 secretion leader is marked by arrow 3.
CaCl\textsubscript{2}·2H\textsubscript{2}O, 0.9 g KCl, 0.5 g MgSO\textsubscript{4}·7H\textsubscript{2}O, 2 ml 500 × biotin and 1.47 ml trace salts stock solution (in 1 l: 6 g CuSO\textsubscript{4}·5H\textsubscript{2}O, 0.08 g NaI, 3 g MnSO\textsubscript{4}·H\textsubscript{2}O, 0.2 g Na\textsubscript{2}MoO\textsubscript{4}·2H\textsubscript{2}O, 0.02 g H\textsubscript{2}BO\textsubscript{3}, 0.5 g CoCl\textsubscript{2}, 20 g ZnCl\textsubscript{2}, 5 g FeSO\textsubscript{4}·7H\textsubscript{2}O and 5 ml H\textsubscript{2}SO\textsubscript{4}). The pH was adjusted to 5 for (pTRP) or 6 (for HSA) with 5 M KOH, followed by sterilization through filtration. Complex screening medium used for GFP and antibody fragment secretion contained in 1 l: 22 g glucose monohydrate, 10 g yeast extract, 10 g peptone, 100 mM potassium phosphate buffer pH 6.0, 13.4 g yeast nitrogen base with ammonium sulfate, 0.4 mg biotin. Every 12 h cells were fed with 100 μl 50% glucose, before cells were harvested by centrifugation at 2500 g until analysis.

Quantification of recombinant proteins. The amount of secreted pTRP was quantified using the enzymic p-toluene sulfonyl-L-arginine ethyl ester-HCl (TAME) assay after desalting of the supernatants, a human albumin standard was used with a starting concentration of 400 ng ml\textsuperscript{-1}. Supernatant samples were diluted accordingly. Detection was carried out with para-nitrophenylphosphate (pNPP) substrate. Coating, dilution and washing buffer were based on PBS and completed with 1% (w/v) BSA and 0.1% (v/v) Tween20 accordingly.

**SDS-PAGE, Western blot analysis and immunoassay.** For protein gel analysis the NuPAGE Novex Bis-Tris system was used. After electrophoresis, the proteins were either visualized by silver staining or electroblotted onto a nitrocellulose membrane for Western blot detection. HSA was detected with anti-HSA-horseradish peroxidase (HRP) conjugate (Bethyl Laboratories), for detection of IgG LC the chemiluminescent Super Signal West chemiluminescent substrate for HRP conjugates was used.

**Determination of protein N-termini by Edman degradation or HPLC-MS.** The N-terminus of HSA, secreted as a double band by pTRP, was analysed by N-terminal Edman sequencing. Therefore, 500 μl respective supernatant was loaded onto a centrifugal filter for concentrating the protein (Amicon Ultra-0.5 ml 10 kDa centrifugal filter), centrifuged for 5 min and a 15 μl sample was recovered by reverse spin. Thereafter, samples were separated by a 4–12% Bis-Tris NuPAGE gel and a Western blot with borate buffer (in 1 l: 3.09 g boric acid, 100 ml methanol at pH 9 adjusted with 1M NaOH) was performed for 2 h at 25 V using a PVDF membrane. 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
RESULTS AND DISCUSSION

Secretion of recombinant proteins by testing the Exp1 native leader sequence

The sequence preceding the secreted form of the *P. pastoris* EpXL-RT was investigated in more detail and tested for its suitability to secrete recombinant reporter proteins. As can be seen in Fig. 1, the native EpXL-RT sequence contains a pre-prosequence and has a rather unconventional cleavage motif VYRT–LKPG (Fig. 1a). Firstly, EGFP was fused to the C-terminus of the native EpXL-RT sequence (EpXL-RT; Fig. 1b) for secretion and was used to transform *P. pastoris* X-33. After cultivation, supernatants as well as cell extracts were analysed with SDS-PAGE (Fig. 2a) and Western blotting. No secretion of EGFP using the native EpXL-RT was observed (Fig. 2a, right panel), in contrast to MFx (Fig. 2a, left panel). In cell extracts, EGFP was found to be still attached to parts of the EpXL-RT sequence; thus, ruling out defects in protein expression itself (data not shown). Fluorescence microscopy (Fig. 3a) revealed that EGFP was mainly located in the vacuole (stained with Coomassie R250 solution) after 15 min, indicating that the protein was not correctly processed.

**Fluorescence microscopy.** GFP-expressing clones were inoculated in minimal medium with an initial OD_{600} of 0.1 and grown for approximately 18 h. Vacuoles of the cells were stained using FM4-64 (Vida & Emr, 1995) and the nucleus was stained using DAPI as described previously (Delic et al., 2012). Briefly, the cells were stained with 15 μM FM4-64 (Invitrogen) diluted in culture medium for 15 min at 30 °C with shaking in the dark. The cells were washed and incubated for a further 60 min in minimal medium. After a final washing step, the cells were viewed on a Leica DMi6000 SPS confocal microscope using a HCX PL APO CS 63.0 × 1.30 NA glycerol-immersion objective. Sequential scans were performed using 488 and 561 nm lasers. For DAPI staining, GFP-expressing clones were fixed in 4 % (v/v) paraformaldehyde solution (Thermo Scientific) for 1 h, washed with PBS and resuspended in PBS. A few microtubules fixed cells were washed with an equal amount of Vectashield mounting medium (Vector Laboratories) containing 50 ng DAPI ml^{-1} (Invitrogen), put on a microscope slide and covered with a coverslip. The cells were viewed on a Leica DMi6000B fluorescence microscope using a HCX PL APO 100 × 1.4 NA oil-immersion objective, and appropriate filters for DAPI and GFP.

**Fig. 2.** Reporter protein secretion using the native Exp1 leader (EpXL-RT), variants thereof (ExpL-KR::QN, ExpL-KR::TV) or a reference leader (MFx leader or native HSA leader). SDS-PAGE of supernatant of *P. pastoris* secreting: (a) EGFP with native ExpL-RT leader or MFx leader; (b) pTRP with native ExpL-RT leader or MFx leader. (c) Supernatant of *P. pastoris* secreting HSA with native ExpL-RT leader or HSA leader: (i) silver stained SDS-PAGE; (ii) Western blot. (d) SDS-PAGE of up-concentrated supernatant of (c) prepared for Edman sequencing. (e) Analyses of *P. pastoris* secreting EGFP by EpXL-RT leader variants, where the dibasic amino acids Lys-Arg (KR) were replaced with Gln-Asn (ExpL-KR::QN) and Thr-Val (ExpL-KR::TV): (i) silver stain of supernatant; (ii) Western blot of up-concentrated supernatant; (iii) Western blot of cell lysate. SDS-PAGE of supernatant of *P. pastoris* secreting: (f) pTRP by ExpL-KR::QN and ExpL-KR::TV; (g) HSA by ExpL-KR::QN and ExpL-KR::TV. Arrows represent reporter protein, while asterisks indicate fusions of the reporter with incorrectly processed leader sequences. The results of two independent clones per construct are shown.
using FM4-64; Vida & Emr, 1995) in these cells when using the native EpxL-RT leader or MFz. Both vacuolar and cytosolic location has been previously reported by other groups for MFz (Kottmeier et al., 2011; Kunze et al., 1999).

As EGFP is not secreted in its native form, it may not be a perfectly suited reporter to study secretion. Therefore, we decided to investigate the ability of EpxL-RT to secrete model proteins that are secreted by their natural hosts. We chose two proteins with different N-termini (Table 2), namely pTRP and HSA, which both were shown to be secreted by P. pastoris before, albeit at different levels (Baumann et al., 2008; Kobayashi et al., 2000). HSA is one of the highly produced proteins in P. pastoris (Kobayashi, 2006); thus, serving as a good model for efficiently secreted proteins. Usually, HSA is secreted using its native signal sequence. Analysis by (immuno-) fluorescence microscopy shows a distinct ER pattern (prominent perinuclear ring; Rossanese et al., 1999) within the cell for efficiently secreted proteins such as HSA secreted by the HSA leader [Fig. 3b(i)]. Using the EpxL-RT leader, HSA has only a weak ER pattern [Fig. 3b(ii)], again pointing to some deficiencies in secretion or leader processing. It should be noted that when using a Δpep4 background (deficient in vacuolar proteinase A), HSA can also be found in the vacuole (as well as EGFP, data not shown), but it seems to be rapidly degraded in wild-type cells.

Cells expressing pTRP using EpxL-RT or MFz for secretion were cultivated and analysed using enzymic trypsin activity assay and SDS-PAGE (Fig. 2b). Unexpectedly, pTRP expression using the EpxL-RT secretion leader led to a protein smear of approximately 30 kDa (Fig. 2b, right panel), whereas MFz led to secreted pTRP of the correct size (25 kDa; Fig. 2b, left panel). Parts of secreted pTRP with EpxL-RT also led to correctly sized pTRP, albeit to a lesser extent. The smeared band probably represents an EpxL-RT-pTRP fusion protein, which might derive from incorrect processing (cleavage) of the Epx-leader sequence (Fig. 1a).

A distinct double band pattern was visible for HSA secreted by EpxL-RT in the Western blot (Fig. 2c, right panel). In comparison with HSA secreted by its native leader (Fig. 2c, left panel) and purified HSA (not shown), the lower band represents correctly processed HSA. The upper band with slightly higher molecular mass may represent an EpxL-RT-HSA fusion protein with incorrectly processed EpxL-RT. To investigate the N-terminal amino acids of these two bands, Edman sequencing as well as

Fig. 3. Intracellular localization of reporter proteins secreted by the native Exp1 leader (EpxL-RT), variants thereof (ExpL-KR::QN, ExpL-KR::TV) or a reference leader. (a) P. pastoris expressing EGFP using (i) MFz leader, (ii) native Exp1 leader (EpxL-RT) and EpxL-RT leader with the KR-exchanges (iii) QN and (iv) TV were stained with FM4-64 and observed with a confocal microscope. Left panel, bright-field illumination; right panel, fluorescence images of EGFP (green) and FM4-64 (red, vacuolar membrane). (b) Immunofluorescence images of P. pastoris expressing HSA with (i) HSA leader, (ii) native Exp1 leader (EpxL-RT) and EpxL-RT leader with the KR-exchanges (iii) QN and (iv) TV. Left panel, fluorescence images of HSA; right panel, fluorescence images of HSA (green) and DAPI (blue, nucleus). Bars, 3 μm.
HPLC-MS were performed. For the lower band (Fig. 2d), N-terminal sequencing confirmed Asp-Ala (DA) as the N-terminal amino acids, conclusively with the first amino acids of HSA (N-terminus of HSA; Table 2), while for the upper band (Fig. 2d) Ala-Tyr-Tyr-Thr (AYYT) were determined as N-terminus of the secreted protein. These amino acids are part of the EpxL-RT prosequence (Fig. 1b), leaving an unwanted 21 amino acid overhang on HSA (designated prosequence 2). The amino acid sequence preceding AYYT in EpxL-RT is Leu-His-Lys-Arg (LHKR), which might be part of a dibasic Lys-Arg peptidase cleavage motif.

Analysing the impact of the dibasic motif KR in the prosequence of the native Epx1 leader

BLAST and CLUSTAL W sequence alignment of P. pastoris EpxL-RT with the leader sequences of Epx1 homologues of other yeasts (e.g. S. cerevisiae, Kluyveromyces lactis, Candida ssp., Yarrowia lipolytica, Hansenula polymorpha) revealed that Lys-Arg (KR) in the leader sequence is not conserved in most other yeast species (Fig. S1, available in the online Supplementary Material). Exceptions are the closely related methylotrophic yeast H. polymorpha (synonym Ogatea parapolymorpha), the phylogenetically close species Y. lipolytica (Dujon, 2010) and some filamentous fungi; however, in these cases the surrounding sequences seem not to be conserved. Processing of the KR motif by proteases such as Kex2 cannot be assumed a priori based on sequence analysis, since the amino acids surrounding the processing site and the 3D structure are key determinants (Bader et al. 2008). To investigate whether the presence of the KR motif is influencing secretion using EpxL-RT, we exchanged those two amino acids against other pairs of amino acids (Fig. 1c). Thr-Val (TV) was selected as it is the consensus sequence at this position of leaders of Epx1 homologues (SCP-like proteins) in other yeasts including S. cerevisiae. Additionally, the charged amino acids Lys-Arg were exchanged for two polar amino acids Gln-Asn (QN) and Ser-His (SH) to test for a pair of hydrophobic or positively charged amino acids; thus, generating the modified secretion leader EpxL-KR :: TV, ExpL-KR :: QN and ExpL-KR :: SH, respectively. As ExpL-KR :: QN and ExpL-KR :: SH resulted in a similar outcome, in the following we further report just on ExpL-KR :: QN and EpxL-KR :: TV.

The exchange of KR in the secretion leader for the abovementioned amino acids did not affect EGFP secretion significantly. The native EpxL-RT leader failed to secrete GFP and so did the exchanges, as can be seen in the SDS-PAGE in Fig. 2e(ii). Western blot analysis of concentrated supernatants revealed a low level of EGFP secretion for both exchange constructs. With EpxL-KR :: QN-GFP slightly higher secretion levels were observed, but based on the band pattern we concluded that the Epx1-leader was still attached [Fig. 2e(ii), left panel]. Also, Western blots of cell extracts showed that intracellular leader processing is

<table>
<thead>
<tr>
<th>Protein</th>
<th>N-terminal sequence of mature protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSA</td>
<td>DAHKSEVAHRKDFDLGEENFKALVLI ...</td>
</tr>
<tr>
<td>EGFP</td>
<td>VSKEELFTGVVPLHELQDGVNGH ...</td>
</tr>
<tr>
<td>pTRP</td>
<td>TDDDDKIVGGYTCANISIYPYQSLN ...</td>
</tr>
<tr>
<td>HyHel LC</td>
<td>DIVLTQPSATLSTVPGNSVLSCRA ...</td>
</tr>
<tr>
<td>HyHel Fab</td>
<td>DVLQIESGSLVKPSQTLSTCSVT ...</td>
</tr>
</tbody>
</table>
| Reporter proteins used in this study

Table 2.

(a) GFP + (EA) 2 MFα EpxL-KR pTRP HSA EpxL-KR (b) MFα EpxL-KR pTRP HSA EpxL-KR (c) Western blot of supernatant of P. pastoris secreting: (a) EGFP with MFα leader or EpxL-KR leader (signal peptide and prosequence 1), the use of MFα leader leads to EAEA overhangs attached to GFP as determined by N-terminal sequencing; (b) pTRP with MFα leader or EpxL-KR leader. (c) Western blot of supernatant of P. pastoris secreting: (d) antibody LC with MFα leader or ExpL-KR leader; (e) EGFP with MFα leader or EpxL-SP signal peptide; (f) LC with MFα leader or EpxL-SP. Arrows represent reporter protein, while asterisks indicate fusions of the reporter with incorrectly processed leader sequences. The results of two independent clones per construct are shown.

Fig. 4. Recombinant-protein secretion with truncated Epx1 leader variants and a reference leader (MFα leader or native HSA leader). SDS-PAGE of supernatant of P. pastoris secreting: (a) EGFP with MFα leader or EpxL-KR leader (signal peptide and prosequence 1), the use of MFα leader leads to EAEA overhangs attached to GFP as determined by N-terminal sequencing; (b) pTRP with MFα leader or EpxL-KR leader. (c) Western blot of supernatant of P. pastoris secreting: (d) antibody LC with MFα leader or ExpL-KR leader; (e) EGFP with MFα leader or EpxL-SP signal peptide; (f) LC with MFα leader or EpxL-SP. Arrows represent reporter protein, while asterisks indicate fusions of the reporter with incorrectly processed leader sequences. The results of two independent clones per construct are shown.
Intracellular localization of reporter proteins secreted by truncated Epx1 leader variants. (a) *P. pastoris* expressing EGFP using the prosequence of the Epx1 leader (EpxL-KR) and the signal peptide of the Epx1 leader (EpxL-SP) were stained with FM4-64 (red, vacuolar membrane) and viewed directly with a confocal microscope (upper and middle panel). *P. pastoris* expressing EGFP using the signal peptide of the Epx1 leader (EpxL-SP) was formaldehyde-fixed, stained with DAPI (blue, nucleus) and viewed with an epifluorescence microscope (lower panel). (b) *P. pastoris* expressing HSA using the prosequence of the Epx1 leader (EpxL-KR) was immunostained (green) and the nucleus was stained using DAPI (blue). The cells were viewed with an epifluorescence microscope. Bars, 3 μm.

**Fig. 5.** Intracellular localization of reporter proteins secreted by truncated Epx1 leader variants. (a) *P. pastoris* expressing EGFP using the prosequence of the Epx1 leader (EpxL-KR) and the signal peptide of the Epx1 leader (EpxL-SP) were stained with FM4-64 (red, vacuolar membrane) and viewed directly with a confocal microscope (upper and middle panel). *P. pastoris* expressing EGFP using the signal peptide of the Epx1 leader (EpxL-SP) was formaldehyde-fixed, stained with DAPI (blue, nucleus) and viewed with an epifluorescence microscope (lower panel). (b) *P. pastoris* expressing HSA using the prosequence of the Epx1 leader (EpxL-KR) was immunostained (green) and the nucleus was stained using DAPI (blue). The cells were viewed with an epifluorescence microscope. Bars, 3 μm.

Processing of the Epx1 secretion leader in *P. pastoris*

For pTRP and HSA, a different picture arose. Exchange of KR to QN, SH or TV led to improved Epx1-leader processing and increased levels of correctly secreted recombinant protein (Fig. 2f, g). Based on silver stain analysis, a positive effect of eliminating the KR motif in the EpxL-RT on targeting pTRP to the extracellular space was observed. In contrast to the protein smear observed with native EpxL-RT, each exchange improved protein secretion of correctly processed recombinant pTRP (Fig. 2f). Similarly, processing of HSA was improved by exchanging KR in the secretion leader. The double band which arose for HSA secreted by the native EpxL-RT leader also appeared with the KR-exchanges, but HSA was correctly processed to a much higher extent (> 80 %; Fig. 2g). Immunofluorescence microscopy showed that HSA, which was secreted with EpxL-KR :: QN or EpxL-KR :: TV, led to an ER pattern with punctate structures, most likely indicating ER exit sites [Fig. 3b(iii, iv)].

Furthermore, we also tested the secretion of the native Epx1 protein using the KR exchanges in the Δepx1 knockout background (Heiss et al. 2013). Thereby, we observed that Epx1 secretion is still occurring (data not shown), the internal KR cleavage is not required for processing and secretion is unaffected by this consensus sequence. It seems that protease cleavage at KR (presumably by Kex2) is not required for driving secretion of Epx1 or reporter proteins, but this cleavage negatively affects processing, maybe by steric hindrance.

**Generation of leader variants for examining intracellular processing**

Next, we wanted to clarify whether the KR motif in the Epx1 leader represents a functional cleavage site of Kex2 or related proteases. Therefore, vectors for the expression of recombinant proteins using the EpxL-KR sequence (signal peptide and prosequence) for secretion of the reporter proteins were constructed (Fig. 1d). Surprisingly, we observed that processing was indeed enhanced by using the shortened Epx1 secretion leader EpxL-KR (Fig. 4a–d). The shortened version of the Epx1 leader, which ends with the KR site, facilitated secretion of HSA. The best results were obtained with EpxL-KR :: QN or EpxL-KR :: TV, each exchange improved protein secretion of HSA by 80 %; Fig. 2g). Similarly, processing of HSA was improved by exchanging KR in the secretion leader. The double band which arose for HSA secreted by the native EpxL-RT leader also appeared with the KR-exchanges, but HSA was correctly processed to a much higher extent (> 80 %; Fig. 2g). Immunofluorescence microscopy showed that HSA, which was secreted with EpxL-KR :: QN or EpxL-KR :: TV, led to an ER pattern with punctate structures, most likely indicating ER exit sites [Fig. 3b(iii, iv)].

Furthermore, we also tested the secretion of the native Epx1 protein using the KR exchanges in the Δepx1 knockout background (Heiss et al. 2013). Thereby, we observed that Epx1 secretion is still occurring (data not shown), the internal KR cleavage is not required for processing and secretion is unaffected by this consensus sequence. It seems that protease cleavage at KR (presumably by Kex2) is not required for driving secretion of Epx1 or reporter proteins, but this cleavage negatively affects processing, maybe by steric hindrance.
leader [based on the different band sizes in Fig. 4a (left panel) as well as HPLC-MS analyses]. Fluorescence microscopy confirmed that EGFP was not targeted to the vacuole anymore when using EpXL-KR (Fig. 5a). Cytosolic targeting of EGFP and secretion was also achieved with the viral K28 pptox leader peptide (Eiden-Plach et al., 2004). Correct processing of the EpXL-KR leader and increased secretion levels were also observed for pTRP. Contrary to the protein smear observed when using the native EpXL-RT leader for secretion (Fig. 2b, right panel), secretion of HSA with EpXL-RT was still lower compared to the native HSA leader. In order to check whether this is due to the aspartate (D) being an unpreferred amino acid at position +1 after the KR cleavage site, we tested for another readily secreted protein starting with the same amino acid residue, which is common in many mammalian proteins. Contrary to HSA, HyHEL antibody LC (HyHEL-LC) was secreted in equal amounts by EpXL-KR and MFza-secretion leaders (Fig. 4d), indicating that KRD is not leading to inefficient processing. The correct N-terminus of pTRP secreted using EpXL-KR was verified by mass spectrometric analysis (HPLC-MS). Moreover, the amount of pTRP secreted using EpXL-KR was more than 20 % higher compared to the MFza-secretion leader as measured by trypsin activity assay (data not shown).

Correct processing of the EpXL-KR leader was also shown for HSA. Contrary to the double band pattern observed when using EpXL-RT for secretion (Fig. 2c, right panel), secretion of HSA using EpXL-KR yielded a single band of the correct size (Fig. 4c, right panel) and fluorescence microscopy imaging showed a weak ER pattern together with more intense ER exit sites for EpXL-KR-HSA (Fig. 5b). Despite the correctly processed N-terminus, secretion levels of HSA with EpXL-KR were still lower compared to the native HSA leader. In order to check whether this is due to the aspartate (D) being an unpreferred amino acid at position +1 after the KR cleavage site, we tested for another readily secreted protein starting with the same amino acid residue, which is common in many mammalian proteins. Contrary to HSA, HyHEL antibody LC (HyHEL-LC) was secreted in equal amounts by EpXL-KR and MFza-secretion leaders (Fig. 4d), indicating that KRD is not leading to inefficient processing. The correct N-terminus of pTRP secreted using EpXL-KR was verified by mass spectrometric analysis (HPLC-MS). Moreover, the amount of pTRP secreted using EpXL-KR was more than 20 % higher compared to the MFza-secretion leader as measured by trypsin activity assay (data not shown).

### Table 3. Summary of results

<table>
<thead>
<tr>
<th>Leader</th>
<th>Secretion &amp; processing of leader</th>
<th>Intracellular localization</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFza</td>
<td>EGFP: medium, incorrect processing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pTRP: strong</td>
<td>EGFP: vacuole</td>
<td>Reference leader</td>
</tr>
<tr>
<td>HSA</td>
<td>EGFP: strong</td>
<td>HSA: intense ER (circular structure around nucleus), no vacuole</td>
<td>Reference leader for HSA</td>
</tr>
<tr>
<td>EpXL-RT</td>
<td>EGFP: very weak (&lt;5 % of MFza)</td>
<td>EGFP: vacuole</td>
<td>Incorrect processing, parts of leader still attached</td>
</tr>
<tr>
<td></td>
<td>HSA: medium (50 % of HSA), &gt;50 % incorrect processing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EpXL-KR:: QN</td>
<td>pTRP: weak (&lt;50 % of MFza), smear</td>
<td>EGFP: vacuole and cytosol</td>
<td>Correct processing of pTRP</td>
</tr>
<tr>
<td></td>
<td>EGFP: very weak (&lt;5 % of MFza)</td>
<td>HSA: punctate structures (maybe ER exit sites or Golgi)</td>
<td>More efficient processing of HSA</td>
</tr>
<tr>
<td></td>
<td>HSA: medium (50 % of HSA), &lt;20 % incorrect processing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EpXL-KR:: TV</td>
<td>pTRP: medium (75 % of MFza), correct processing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EpXL-KR (prosequence 1)</td>
<td>EGFP: strong (120 % of MFza), correct</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HSA: weak (15 % of HSA), correct</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LC: strong, correct</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pTRP: strong (120 % of MFza), correct</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EpXL-SP (signal peptide)</td>
<td>EGFP: strong (120 % of MFza), correct</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LC: strong (&gt; than MFza), correct</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

S. Heiss and others

Microbiology 161

Downloaded from www.microbiologyresearch.org by IP: 54.70.40.11
On: Tue, 18 Dec 2018 22:31:46
terminus of HyHEL-LC secreted by EpxL-KR was verified by HPLC-MS analysis and immunofluorescence microscopy showed a weak ER pattern with punctate structures paired with cytosolic staining (Fig. 5c).

Together, this confirms that KR is active, meaning that the Epx1 leader consists of a bipartite prosequence. The second part of this prosequence seems to be responsible for influencing intracellular targeting of Epx1, as reporter proteins without it are targeted for secretion to the cell exterior, which is considered to be the default pathway for secretory proteins unless other targeting signals are present.

**Truncated minimal Epx1 leader sequences efficiently secrete recombinant proteins**

Given that the EpxL-KR leader (consisting of the signal peptide and prosequence 1) was sufficient and even better for secretion than the native full-length Epx1 leader, in addition, we decided to investigate whether the signal peptide alone would be sufficient for secretion, or whether prosequence 1 and KR processing are needed. Reporter proteins were fused to the first 20 amino acids of the native Epx1 leader, termed EpxL-SP (Fig. 1e). Alanine (A) represents the last amino acid C-terminal of the predicted signal peptidase cleavage site, whereas on the N-terminal side of the signal peptidase cleavage site the recombinant protein starts immediately. We tested the secretion of three recombinant proteins with different N-terminal amino acids to demonstrate that the N-terminus of the product does not influence cleavage by signal peptidase. Therefore, we chose EGFP (starting with the hydrophobic amino acid Val), HyHEL-LC (starting with the negatively charged amino acid Asp) and IFN2x [starting with Cys, a very unusual N-terminal amino acid (Devi, 1991)]. HSA was excluded at this point.

As can be seen in Fig. 4(e, f), secretion levels of GFP and HyHEL-LC using EpxL-SP were comparable to those obtained with MFx and EpxL-KR (Fig. 4a, d). Using just the signal peptide for secretion, both strains exhibit an intracellular product distribution representative for proteins present in the secretory pathway (mainly ER pattern, Fig. 5a, c). IFN2x possesses an N-terminal cysteine residue that is engaged in a disulfide bond. For leader sequences that require cleavage by Kex2, incorrect processing has been reported as cleavage is sterically hindered at this cysteine (Salunkhe et al., 2010). Using EpxL-SP, which is cleaved by signal peptidase and does not require Kex2, we obtained correctly cleaved IFN2x regardless of the N-terminal Cys (correct cleavage was confirmed by N-terminal sequencing). Therefore, we concluded that the EpxL-SP signal peptide not only is sufficient for the secretion of recombinant proteins, but also has beneficial properties that lead to the observed ‘secretory phenotype’.

**Conclusions**

**Unusual three-step processing.** In this study, the secretion leader of the most abundant secretory protein of *P. pastoris*, Epx1 (Heiss et al., 2013), was investigated. We tested the full-length leader and several fragments thereof for their capacity to direct secretion of reporter proteins as well as for correct processing of the leader (summarized in Table 3). Additionally, we monitored the intracellular distribution of the reporter proteins by live cell or immunofluorescence microscopy.

We revealed that the Epx1 leader consists of a signal peptide and a bipartite prosequence, which is separated by a dibasic Lys-Arg motif resembling the cleavage site of Golgi endoprotease Kex2/furin (Rockwell et al., 2002). Processing of dibasic Lys-Arg motifs by proteases such as Kex2 depends not only on the motif itself, but also on the 3D structure and the surrounding amino acid environment (Bader et al., 2008). We studied the effect of exchanging the Kex2 cleavage motif on the secretory behaviour of reporter proteins fused to variants of the Epx1 leader sequence. Previous research has shown that the folding, secretion and processing ability of the cell is different for each protein and that the N-terminal amino acids may influence cleavage of the secretion leader (Wang et al., 2014). We analysed several reporter proteins, including natively non-secreted EGFP and readily secreted proteins such as pTRP and HSA, each starting with a different amino acid.

In summary, the leader of Epx1 consists of a signal peptide and an unconventional bipartite prosequence with an internal Kex2 cleavage site. The Kex2 cleavage site within the *P. pastoris* Epx1 leader is not conserved in homologues in most other yeasts (Fig. S1, see also Bader et al., 2008). This particular dibasic site does not seem not to be required for further processing of the Epx1 leader as determined by exchanging the KR cleavage site to non-cleavable amino acids SH, QN or TV. In contrast, the presence of prosequence 2 in full-length Epx1 leader seems to influence intracellular targeting. This might be related to the presumed function of the Epx1 protein. The *P. pastoris* Epx1 protein is homologous to the *S. cerevisiae* Pry proteins, which function in the export of acetylated sterols along the secretory pathway during lipid detoxification (Choudhary & Schneiter, 2012). As mentioned before, the *S. cerevisiae* Pry proteins do not seem to contain a bipartite leader sequence and the N-terminus of secreted Pry1/2 proteins is unknown. However, N-terminal truncation studies confirmed that the N-terminal signal sequence of Pry1 fused to the CAP domain (containing the sterol-binding region) is sufficient to drive secretion of acetylated sterols, and thus also Pry1 itself, in *S. cerevisiae* (Choudhary & Schneiter, 2012). In *P. pastoris*, the cellular content and distribution of ergosterol is different to *S. cerevisiae*. The intracellular localization of Epx1 might thus convey a species-specific layer of regulation in ergosterol transport and detoxification. As the absence of the KR cleavage
motif and thus the constant presence of prosequence 2 results in vacuolar targeting of the reporter proteins, we assume that the post-Golgi sorting of the native Epx1 protein might be affected in a similar manner.

The enzyme responsible for prosequence 2 cleavage remains to be identified. The secretory aspartyl protease Bar1 or the glycosylphosphatidylinositol (GPI)-anchored aspartyl proteases (yapsins) might be potential candidates. The first amino acids of the mature Epx1 protein are similar to the sequence of \textit{S. cerevisiae} MFz, which is subject to the cleavage by Bar1 (Ballensiefen & Schmitt, 1997), while yapsins are described to recognize and cleave mainly at monobasic residues (reviewed by Gagnon-Arsenault et al., 2006). Homologues of Bar1, as well as seven members of the yapsin family, can be found in the \textit{P. pastoris} genome. However, the specific protease or cleavage pathway has to be identified in future studies.

**Two novel efficient secretion leaders with correct N-termini.** When using the EpxL-RT sequence for secretion, a smear or double band pattern is visible for the secreted recombinant protein, which is indicative of incorrect or no processing of the long leader EpxL-RT. This feature renders the full-length Epx1 leader sequence unsuitable and not useful for recombinant protein secretion. By targeting several recombinant human proteins for secretion, we revealed that a short variant of the Epx1 leader sequence is functional, and thus the constant presence of prosequence 2 remains to be identified. The secretory aspartyl protease might be affected in a similar manner.

When using the EpxL-RT sequence for secretion, a smear or double band pattern is visible for the secreted recombinant protein, which is indicative of incorrect or no processing of the long leader EpxL-RT. This feature renders the full-length Epx1 leader sequence unsuitable and not useful for recombinant protein secretion. By targeting several recombinant human proteins for secretion, we revealed that a short variant of the Epx1 leader sequence is functional, and thus the constant presence of prosequence 2 remains to be identified. The secretory aspartyl protease might be affected in a similar manner.

In summary, our study has revealed that the secretory protein Epx1 possesses an unusual bipartite prosequence that seems to affect intracellular targeting. Furthermore, we identified two shortened variants of the Epx1 secretion leader as novel and valuable tools for the production of recombinant secretory proteins.

**ACKNOWLEDGEMENTS**

The authors thank Marion Rothmueller, Rebecca Goenrich and Christine Sanystra for their help with \textit{P. pastoris} cultivations and analyses, and Richard Zahl for critically reading the manuscript. This project was financially supported by LONZA (Visp, Switzerland) and is covered by WO 2014067926. V. P. and C. G. were funded by the Federal Ministry of Science, Research and Economy (BMWFW), the Federal Ministry of Traffic, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol and ZIT – the Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG. EQ BOKU VIBT GmbH is acknowledged for providing HPLC MS-/MS instrumentation. Finally, we thank the BOKU-VIBT Imaging Center for access and expertise with fluorescence microscopy equipment. Some of the authors (S. H., D. M. and B. G.) are inventors (but not owners) of a patent application related to the described results.

**REFERENCES**


Stadlmayr, G., Mecklenbräucker, A., Rothmüller, M., Maurer, M., Sauer, M., Mattanovich, D. & Gasser, B. (2010). Identification and


Edited by: I. Van der Klei