jmb

A New Signal Sequence for Recombinant Protein Secretion in *Pichia pastoris*

Nagaraj Govindappa[1]*, Manjunatha Hanumanthappa[2], Krishna Venkatarangaiah[2], Sankar Periyasamy[1], Suma Sreenivas[1], Rajeev Soni[1], and Kedarnath Sastry[1]

[1]*Biocon Research Limited - SEZ Unit, Bengaluru 560 099, India*
[2]*Department of PG Studies and Research in Biotechnology, Jnanasahyadri, Kuvempu University, Shankaraghatta, Karnataka 577 451, India*

Pichia pastoris is one of the most widely used expression systems for the secretory expression of recombinant proteins. The secretory expression in *P. pastoris* usually makes use of the pre-pro MATα sequence from *Saccharomyces cerevisiae,* which has a dibasic amino acid cleavage site at the end of the signal sequence. This is efficiently processed by Kex2 protease, resulting in the secretion of high levels of proteins to the medium. However, the proteins that are having the internal accessible dibasic amino acids such as KR and RR in the coding region cannot be expressed using this signal sequence, as the protein will be fragmented. We have identified a new signal sequence of 18 amino acids from a *P. pastoris* protein that can secrete proteins to the medium efficiently. The *PMT1*-gene-inactivated *P. pastoris* strain secretes a ~30 kDa protein into the extracellular medium. We have identified this protein by determining its N-terminal amino acid sequence. The protein secreted has four DDDK concatameric internal repeats. This protein was not secreted in the wild-type *P. pastoris* under normal culture conditions. We show that the 18-amino-acid signal peptide at the N-terminal of this protein is useful for secretion of heterologous proteins in *Pichia*.

Keywords: DDDK protein, *PMT1* gene in activation, *P. pastoris* strain BICC 9450, protein secretion

Introduction

Signal peptides control the entry of virtually all proteins to the secretory pathway, both in eukaryotes and prokaryotes [12, 16, 26]. They comprise the N-terminal part of the amino acid (aa) chain and are cleaved off while the protein is translocated through the membrane. In some cases, the presence of N-terminal signal sequences leads to strong stimulation of the expression of recombinant proteins [20]. With very few exceptions, all the proteins secreted from eukaryotic cells require a N-terminal signal peptide. There are a couple of online tools available for predicting signal peptides and corresponding cleavage sites in protein constructs based on their aa sequence [3]. Stephan Klatt and Zoltan Konthur have successfully demonstrated that minor modifications in the aa sequence

of the natural SP from SAP1 based on *in silico* predictions, following the (-3, -1) rule, resulted in different expression-secretion rates of the protein of interest [21].

Pichia pastoris (*P. pastoris*) has very few known secretion leaders for secretion of heterologous proteins. The most widely used signal sequence is the *S. cerevisiae* MATα signal sequence. That apart, the *P. pastoris* acid phosphatase *PHO1* signal sequence is being used in a few cases [7, 17] and the *S. cerevisiae SUC2* gene signal sequence has been used occasionally [15]. The classical mammalian bovine β-casein signal peptide was evaluated for the secretory expression of recombinant xylanase in *Pichia*. It was found that more recombinant xylanase was trapped intracellularly with the β-casein signal sequence when compared with the MATα signal sequence [6].

Insulin and insulin analog expressed in *P. pastoris* have
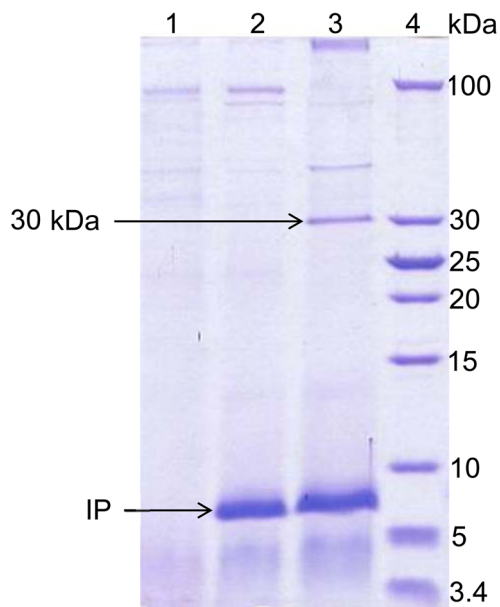
Fig. 1. Secretion of a 30 kDa DDDK protein in the *PMT1* gene inactivated host.

The *P. pastoris* strains were grown and induced with methanol using the standard protocol described in the Invitrogen manual under identical conditions. Equivalent volumes (15 μl) of secreted samples were analyzed on Tricine SDS-PAGE. Lane 1: *P. pastoris* GS115; lane 2: Insulin precursor (IP)-secreting *P. pastoris* clone; lane 3: *PMT1* gene inactivated IP-producing clone; and lane 4: low molecular weight protein marker (Thermo Scientific, Part No. 26632). It is evident from the gel that a 30 kDa protein is being secreted to the medium in the *PMT1*-gene-inactivated host strain.

O-glycosylated species [4, 10]. The inactivation of various protein mannosyltransferase (*PMT*) genes present in the *Pichia* genome helps in reduction of *O*-mannosylated species, which is considered an impurity. We have reported previously that the *PMT1* gene knockout in *P. pastoris* insulin strain reduces the *O*-glycosylation of insulin precursor by ~60% [4]. Analysis of the extracellular protein profile of all these *PMT*-gene-inactivated strains revealed that a ~30 kDa protein was being secreted into the medium from the *PMT1*-gene-inactivated strain (Fig. 1). No other *PMT*-gene-inactivated strains showed this protein band. All other *PMT*-gene-inactivated strains were similar to the parent strain. The protein band was subjected to N-terminal sequencing by Edman degradation method. The amino acid sequence deduced was "DQTFGVLLIRSG". This sequence was matching with the mature portion of a hypothetical protein (XP_002492554.1). The 18 amino acids at the N-terminus were not part of the secreted protein. We were keen to understand why this protein is being secreted

only upon *PMT1* gene inactivation. We attempted to investigate whether these 18 amino acids have any signal sequence property and can be used for heterologous protein secretion in *P. pastoris*. We have first cloned this 30 kDa protein under the control of the strong *AOX1* promoter inframe with this new signal sequence and checked for secretion in the wild-type *P. pastoris* strain. Later, we developed various expression constructs and transformed them into *P. pastoris* host to evaluate and characterize the signal sequence. The resulting strains were induced with methanol in shake flask experiments. Our results show that the 18-amino-acid signal peptide indeed has potential to secrete the heterologous proteins efficiently into the medium. Moreover, this signal sequence is cleaved off independently as efficiently as with the use of Kex2p cleavage sites. Hence, this signal sequence is more useful for expressing recombinant proteins with internal dibasic amino acids such as KR and RR (Kex2p cleavage sites). We also studied the expression of porcine carboxypeptidase B (CPB) [14] and *Erythrina* trypsin inhibitor (ETI) [23] from plant origin to show that this signal sequence is robust and can be used for secretion of many recombinant proteins that can be expressed in *P. pastoris*.

Materials and Methods

Strains, Plasmids, and Media

The *P. pastoris* strain BICC 9450 (MTCC 5835) [5] and its derivative BICC 9682 (*his*⁻) hosts were used in the current studies. The modified pMBL 208 expression vector, which provides the promoter and terminator sequences of the *P. pastoris AOX1* gene, were employed for expression. The *S. cerevisiae* MATα signal sequence was replaced with the "MFNLKTILISTLASIAVA" signal sequence for secretion of proteins into the medium. *E. coli* strain DH5α was used for routine cloning and propagation of plasmids. Yeast extract-peptone-dextrose (YPD) medium containing 10 g/l yeast extract, 20 g/l peptone, and 20 g/l dextrose was used for routine growth and subculturing of *Pichia* strains. YNBD medium used for selection contained 13.4 g/l yeast nitrogen base (YNB) without amino acids and 20 g/l dextrose (D). Luria broth/agar was used for culturing *E. coli*. Media components used were either from Himedia (Mumbai, India) or Difco (USA).

Amino Acid Sequencing

The 30 kDa protein secreted into the medium by the *PMT1*-gene-inactivated *P. pastoris* strain was separated on 10% SDS-PAGE using a standard procedure [18]. The protein bands were immobilized on to Sequi-blot-PVDF membrane, 0.2 μM (Bio-Rad Laboratories, USA) by transfer using 25 mM Tris, 192 mM glycine, and 20% methanol [24]. After transfer, the membrane was rinsed thoroughly with distilled water, followed by 50% methanol

Table 1. Primers used for genetic manipulation of the strains developed in this study.

Primer name	Nucleotide sequence in 5' to 3' direction
DDKFP	CGG GAT CCA TGT TCA ACC TGA AAA CTA TTC TC
DDKRP	CGG AAT TCT TAT TTC TCC CAT ACT TTA AGG CCA AT
DDK(-NS)FP	GAA GAT CTG ACC AAA CCT TCG GTG TCC TT
DDDKKRFP	ATC GAT CGC TGT TGC CAA AAG AGA CCA AAC CTT CGG TGT C
NSSFP1	GGA TCC ATG TTC AAC CTG AAA AC
NSSFP2	TAT TCT CAT CTC AAC ACT TGC ATC GAT CGC
NSSFP3	TGT TGC CCT CGA GCG G
NSSRP1	CCG CTC GAG GGC AAC AGC GAT CGA TGC AAG
NSSRP2	TGT TGA GAT GAG AAT AGT TTT CAG GTT GA
NSSRP3	ACA TAT GGA ATT CC
NSSVACPB	CCA TCG ATC GCT GTT GCC GAA GCT GAG GCA CAT CAC TC
NSSVAETI	CCA TCG ATC GCT GTT GCC GTT TTG TTG GAT GGT AAT GG

solution and once again with distilled water to remove any glycine traces that would be carried over from the transfer buffer. The protein bands were visualized on the PVDF membrane by staining with 0.1% Coomassie blue in 1% acetic acid and 40% methanol. The membrane was destained using 50% methanol. The 30 kDa band was excised and sequenced at ProSeq (3 King George Drive, Boxford, MA 01921, USA).

**Construction of the Vectors for Secretory Expression in *P. pastoris***

The gene corresponding to the DDDK protein ORF was amplified along with the signal sequence using primers DDKFP and DDKRP (Table 1) from *P. pastoris* genomic DNA. The 870 bp PCR product was cloned into pTZ57R vector and the sequence verified. The gene was excised using *Bam*HI and *Eco*RI restriction enzymes and subcloned into pMBLNSS208 vector in *Bam*HI and *Eco*RI sites to obtain DDK/pMBLNSS208 vector, which replaces the MATα signal sequence. The DDDK gene was now under the regulatory control of the *AOX1* promoter (Fig. 2B). The DDDK protein ORF without the signal sequence was amplified and cloned similarly to have a control for the above (Fig. 2A), but the forward primer used was DDK(-SS)FP in place of DDKFP. These two constructs were used to study whether the strong promoter can lead to secretion of this protein in the wild-type *Pichia*. The constructs in Fig. 2C are similar to those in Fig. 2B but with the Kex2p site (KR) at the end of the signal sequence. This construct was developed to evaluate the level of secretion with and without Kex2p cleavage site.

Synthesis of Signal Sequence and Development of CPB and ETI Clones

The signal sequence was synthesized using the primers NSSFP1, NSSFP2, NSSFP3, NSSRP1, NSSRP2, and NSSRP3 (Table 1). The primers were phosphorylated by treating with T4 polynucleotide kinase, annealed, and ligated. The full-length signal sequence was amplified from the ligated template using NSSFP1 and NSSRP1. The PCR product was cloned into pTZ57R vector and the

sequence verified. The signal sequence was cloned into pMBL208 vector to replace the MATα signal sequence.

The CPB and ETI coding sequences were amplified using the forward primers NSSVACPB and NSSVAETI and pPIC9K reverse primer from the plasmids CPB/pPIC9K and ETI/pPIC9K, which we had expressed previously in our laboratory. The forward primer has the *Cla*I restriction site. The PCR product was digested
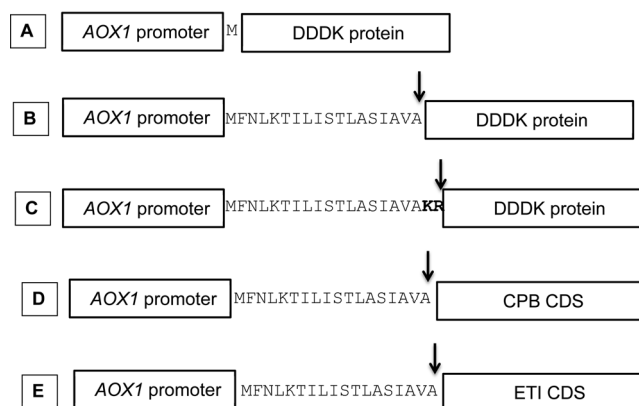
Fig. 2. Constructs developed to evaluate the signal sequence properties of the 18 amino acids at the N-terminus of DDDK protein under the regulatory control of the *AOX1* promoter. The signal sequence part is designated with a single letter amino acid code. (**A**) The control construct with DDDK protein mature sequence without any signal sequence under the *AOX1* promoter. (**B**) Complete open reading frame of the DDDK protein with native signal sequence. (**C**) Same as (**B**) but with Kex2p cleavage site 'KR' introduced after the native signal sequence, which is generally being used in the *P. pastoris* system for efficient secretion of proteins to the medium. (**D**) The heterologous protein porcine CPB coding sequence cloned inframe with the signal sequence. (**E**) The heterologous protein ETI coding sequence cloned inframe with the signal sequence.

with *Cla*I and *Eco*RI restriction enzymes and cloned into the DDK/pMBLNSS208 vector in the identical sites to yield the constructs in Figs. 2D and 2E.

Transformation of *P. pastoris*

The plasmid constructs were digested with *Sac*I and transformed into the *P. pastoris (his⁻)* strain. Transformation was carried out by electroporation of freshly prepared competent cells in 0.2 cm cuvettes. The pulses were delivered by a Gene Pulser (BioRad) at 1,500 V, 25 µF, and 200 Ω. The electroporated cells were allowed to recover for 1 h in 1 M Sorbitol at 30°C and then spread on to YNBD agar plates. The resulting transformants were selected on YPD plates, containing 0.5 mg/ml of G418 concentration, to identify multicopy clones. Several clones that were viable at 0.5 mg/ml concentration were tested in small-scale induction experiments. The proteins secreted to the extracellular medium were analyzed on 10% SDS-PAGE.

Induction of Protein Expression in *P. pastoris*

The clones were grown on 10 ml of YNB medium (1.34% yeast nitrogen base w/o amino acids and 2% dextrose) overnight. It was subcultured into 50 ml of BMGY (1% yeast extract, 2% peptone, 1.34% YNB, 100 mM potassium phosphate at pH 6.0, and 1% glycerol) to 0.5 OD$_{600}$. This culture was incubated at 30°C for 48 h with shaking at 220 rpm. The cells were harvested by centrifugation at 5,000 rpm for 5 min at room temperature and resuspended in induction medium (1% yeast extract, 2% peptone, 1.34% YNB, and 100 mM potassium phosphate at pH 6.0). Methanol (20%) was added to the culture medium to give a final concentration of 1.0% methanol as an inducer. Samples were collected every day and analyzed by SDS-PAGE.

Western Blotting

Western blotting was carried out as per the method described by Towbin *et al.* [25] with a few modifications. The protein samples were separated on 12% SDS-PAGE and then transferred to a Nitrocellulose sheet, 0.2 µM (Bio-Rad Laboratories, USA), as described above. After transfer, the membrane was soaked in 1% BSA in Tris-NaCl buffer (20 mM Tris, pH 7.2, 150 mM NaCl) and incubated overnight to block the nonspecific binding. The membrane was washed with wash buffer (20 mM Tris, pH 7.2, 150 mM NaCl, 0.05% Tween-20) thoroughly, and then probed with an anti-DDK monoclonal antibody (OriGene Technologies, Inc, Rockville, MD, USA) at 1:100 dilution with Tris-NaCl buffer containing 0.1% BSA. The DDK protein was identified with a polyclonal sheep anti-mouse IgG-HRP conjugate (product code: NA931; GE Healthcare, UK) using TMB substrate.

Results

Amino Acid Analysis and Identification of Signal Sequence

The 30 kDa protein secreted in the *PMT1*-gene-inactivated *P. pastoris* strain was immobilized onto Sequi-blot-PVDF membrane and submitted for N-terminal aa sequencing at ProSeq Inc., USA. The first 12 aa were found to be "DQTFGVLLIRSG". This sequence was searched against the NCBI database using the BLASTp program to find out the probable homologous protein. It resulted in the identification of a 289-aa-containing hypothetical protein in the *P. pastoris* GS115 genome database [19]. Interestingly these 12 aa are matching from the aa 19D to 27G in the complete ORF of the hypothetical protein (XP_002492554.1). There are 18 aa at the N-terminal of the protein acting as a signal sequence (Fig. 3), which seem to be cleaved off during the secretion of the mature protein into the medium. The theoretical mass of this protein was found to

```
          10         20         30         40         50         60
MFNLKTILIS TLASIAVADQ TFGVLLIRSG SPYHYSTLTN RDEKIVAGGG NKKVTLTDEG

          70         80         90        100        110        120
ALKYDGGKWI GLDDDGYAVQ TDKPVTGWST NGGYLYFDQG LIVCTEDYIG YVKKHGECKG

         130        140        150        160        170        180
DSYGMAWKVL PADDDKDDDK DDDKDDDKDY DDDNDHGDGD YYCSITGTYA IKSKGSKHQY

         190        200        210        220        230        240
EAIKKVDAHP HVFSVGGDQG NDLIVTFQKD CSLVDQDNRG VYVDPNSGEV GNVDPWGELT

         250        260        270        280
PSVKWDIDDG YLIFNGESNF RSCPSGNGYS LSIKDCVGGT DIGLKVWEK
```

Fig. 3. Amino acid sequence of the 30 kDa DDDK protein identified by searching against the *P. pastoris* database.
The N-terminal sequencing of the 30 kDa band resulted in the amino acids D19 to G30, which are underlined. The protein has four repeats of DDDK in a stretch from D133 to K148, which are italicized and underlined. Hence, we named this as the DDDK protein. The amino acids M1 to A18, which are in bold, correspond to the signal sequence studied for heterologous protein expression in this study.
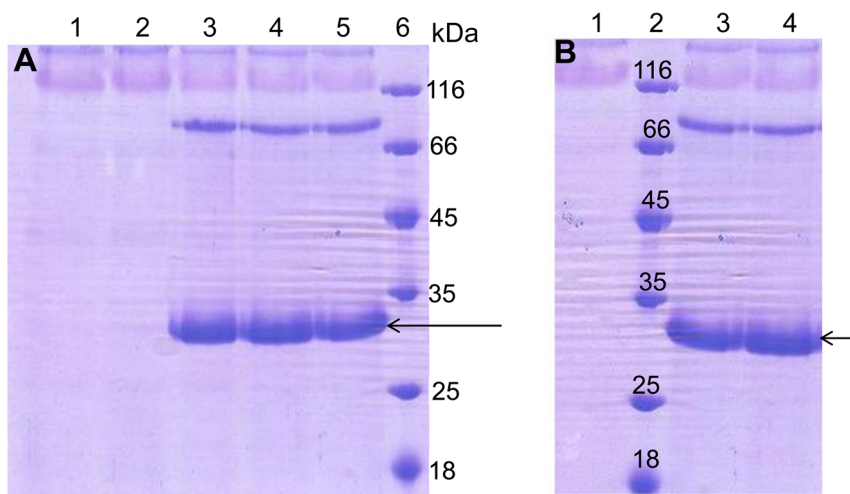
Fig. 4. SDS-PAGE analysis of the secreted 30 kDa DDDK protein.
(**A**) The 30 kDa DDDK protein secreted into the medium when expressed under the regulatory control of *AOX1* promoter in wild-type *P. pastoris*. Lane 1: *P. pastoris* host strain; lane 2: *P. pastoris* host carrying DDDK gene CDS without the 18-amino-acid signal peptide construct (Fig. 2A); lanes 3–5: Clones #1, 2, and 3 of *P. pastoris* host carrying construct of Fig. 2B that has the DDDK gene CDS along with 18-amino-acid signal sequence. (**B**) Comparison of 30 kDa DDDK protein secretion when Kex2p cleavage site was inserted after the 18-amino-acid signal sequence. Lane 1: *P. pastoris* host carrying construct that has DDDK gene CDS without the 18-amino-acid signal peptide (Fig. 2A); lane 2: protein molecular weight marker; lane 3: *P. pastoris* host carrying construct that has DDDK gene CDS along with the 18-amino-acid signal sequence (Fig. 2B); and lane 4: *P. pastoris* host carrying construct that has DDDK gene CDS along with the 18-amino-acid signal sequence plus KR (Fig. 2C), the Kex2p cleavage site. There was no difference in the level of protein secretion in lanes 3 and 4. This indicates that the signal sequence can secrete the protein efficiently without the need for Kex2p.

be 31.66 kDa with a pI of 4.44 when calculated using ExPASY protein parameter tools (http://web.expasy.org/protparam/). It is a highly acidic protein with four 'DDDK' aa repeats at a stretch in the primary aa sequence.

Secretion of DDDK Protein Under *AOX1* Promoter in Wild-Type *P. pastoris*

In order to evaluate whether this protein will be secreted if fused with the strong promoter *AOX1*, the 870 bp complete DDDK ORF (which includes the coding sequence of 18 aa signal peptide) was cloned under the regulatory control of the *AOX1* promoter (Fig. 2B). This construct was transformed into the *P. pastoris* wild-type strain. The clones obtained were screened for the secretory production of the DDDK protein. We found good secretory expression of the DDDK protein in the wild-type *P. pastoris* strain (Fig. 4). This indicates that the promoter of the native protein is indeed very weak and may not be able to produce a sufficient amount of protein to secrete into the medium, although it has a potential signal sequence at the N-terminus. The results shown in Fig. 4B indicate that the proteins secreted with and without the Kex2p cleavage site are similar.

Identification of Production of DDDK Protein Using Antibody in Wild-Type *P. pastoris* Host

To understand whether the DDDK protein is being secreted only in strains with *PMT1*-gene-inactivated host and also to identify the presence of any intracellular proteins, a western blot experiment using a monoclonal antibody that binds to proteins that have the DDDK motif was carried out. The protein band that corresponds to DDDK protein was not detected in the supernatants of *Pichia* wild-type strain. (lane 1, Fig. 5), indicating that the protein is being secreted only when the *PMT1* gene is inactivated. It was also noticed that there was no band observed in the intracellular fraction of *Pichia* (lanes 5 and 6, Fig. 5), indicating that all the protein expressed had been successfully secreted to the medium with use of the DDDK protein signal sequence.

Secretion of CPB and ETI Using DDDK Protein Signal Sequence

In an attempt to determine whether the signal sequence functions well for secreting heterologous proteins, we developed constructs having CPB and ETI coding sequences fused inframe with this signal sequence (Figs. 2D and 2E).
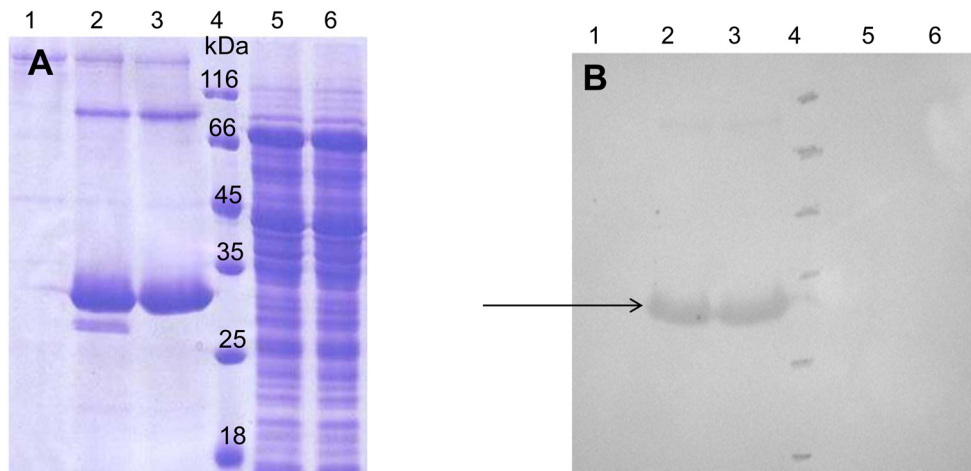
Fig. 5. Western blot analysis of the 30 kDa DDDK protein to identify whether any protein accumulates intracellularly as a result of non-processing of signal sequence.

(**A**) Day 4 methanol-induced culture supernatant samples were loaded (15 μl) in each lane, except lanes 5 and 6 where 0.5 OD_{600} cell lysates were loaded. Lane 1: *P. pastoris* wild-type strain; lanes 2 and 3: *P. pastoris* host carrying construct shown in Figs. 2B and 2C, respectively; lane 4: Protein molecular weight marker; and lanes 5 and 6: Intracellular proteins of *P. pastoris* host carrying construct shown in Figs. 2B and 2C, respectively. (**B**) The proteins from the replicate of the gel shown in Fig. 5A was transferred to a nitrocellulose sheet and probed using anti-DDDK antibody as described in the Methods. The detection of 30 kDa bands in lanes 2 and 3 indicates that the antibody is binding specifically to DDDK protein. The absence of the corresponding bands in lanes 5 and 6 indicates that all the protein made has been secreted efficiently to the medium.

The constructs were transformed into the *P. pastoris* wild-type strain. The clones obtained were screened for the secretion of these proteins. We found the good secretory expression of CPB and ETI proteins, which were comparable to the quantity of the same proteins produced using the MATα signal sequence (Fig. 6).
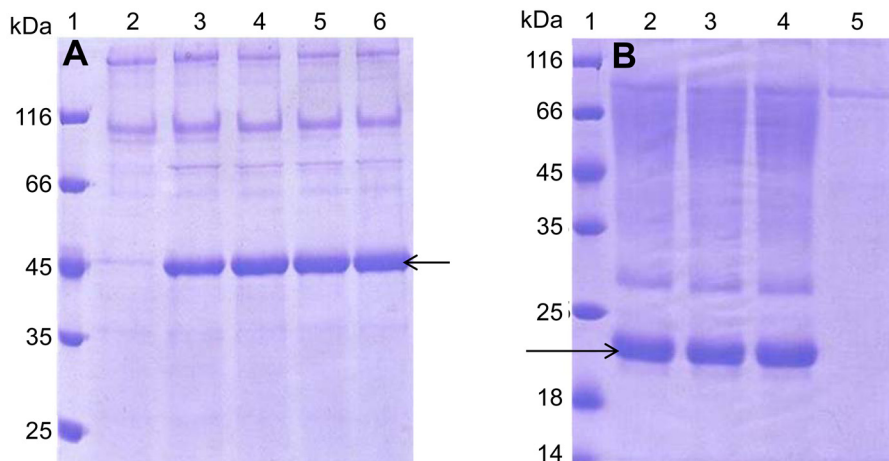
Fig. 6. Porcine carboxypeptidase B expressed using the DDDK signal sequence has been secreted into the medium as efficiently as that secreted with the MATα signal sequence.

Lane 1: Protein molecular weight marker; lane 2: *P. pastoris* wild-type strain; lanes 3, 4, and 5: Clones #1, 2, and 3 of *P. pastoris* host carrying construct that has Pro-CPB CDS with the 18-amino-acid signal peptide (Fig. 2D); and lane 6: *P. pastoris* host carrying Pro-CPB CDS with the MATα signal sequence. (**B**) The level of *Erythrina* trypsin inhibitor (ETI) expressed using the DDDK signal sequence secreted into the medium is comparable to the one secreted with the MATα signal sequence. Lane 1: Protein molecular weight marker; lane 2: *P. pastoris* host carrying ETI CDS with the MATα signal sequence; lanes 3 and 4: Clones #1 and 2 of *P. pastoris* host carrying construct that has ETI CDS with the 18-amino-acid signal peptide (Fig. 2E); and lane 5: *P. pastoris* host strain. All lanes carry the equivalent volume of secreted samples (15 μl).

Discussion

The DDDK protein is secreted into the culture medium only in the *PMT1*-gene-inactivated *P. pastoris* strain and not in any other *PMT*-gene (PMT4, 5, and 6)-inactivated strains. It was not clear why it was secreted in this strain alone. PMT1 is an integral membrane protein [22]. Although the doubling time of the *PMT1*-gene-inactivated insulin precursor producing clone was slightly lower in YPD broth, the cell mass accumulation was not affected and similar results were observed for all the other *PMT*-gene-inactivated strains [4]. It was also noticed from the tricine SDS-PAGE [8] gel picture that the cells were not lysed, owing to the *PMT1* gene inactivation (Fig. 1). If there was cell lysis, one could find other cellular proteins, as shown in Fig. 5, lanes 5 and 6. We sequenced the N-terminal 12 amino acids of the 30 kDa protein band and found the *PIR* gene family of proteins [11] having four concatameric repeats of amino acids "DDDK". This protein is found to be unique to *P. pastoris* when analyzed using the protein databank. The *PMT1* gene was previously inactivated in *S. cerevisiae* and shown by Strahl-Bolsinger *et al*. [22] to be a key enzyme in the protein *O*-glycosylation. When we analyzed the 12-amino-acid sequence from the N-terminus using protein BLAST searches, homology to cell wall protein CPW1 from the *Pichia* genome was identified based on the *S. cerevisiae* database. When the full-length DDDK protein was compared with the *Saccharomyces* cell wall protein (*CPW1/YKL096W*) sequence using NCBI BLAST2p protein search, however, we found only 33% identity and 58% similarity (data not shown). The DDDK repeats were not present in the *S. cerevisiae* cell wall protein. However, when the stretch of four DDDK repeat sequence was used for search using the NCBI BLAST2p program, a few sequences (GenBank Accession No. ELA36046.1, EKG21897.1, CCH44912.1, and XP_001349335.1) were identified. All these deposits are part of sequencing projects and were reported as hypothetical proteins. The N-terminal 18 aa (Fig. 3) was not part of the secreted protein, emphasizing once again that the DDDK protein has not come out of the cell by cell lysis but because of the presence of the signal sequence. We were curious to understand why this signal sequence is not secreting the DDDK protein in the wild-type strains and also to characterize the same sequence for heterologous protein expression in *P. pastoris*.

Currently, the *P. pastoris* expression system to secrete heterologous proteins is mainly based on the 85-amino-acid *S. cerevisiae* MATα signal sequence [1, 2]. A Kex2p cleavage site "KR" at the end of the MATα signal sequence

cleaved off during the export of recombinant proteins to the medium [9]. One of the drawbacks of this powerful expression technology is that, if there are any accessible internal dibasic amino acids present in the recombinant protein sequence, one cannot use the current system. The cleavage and productivity will also depend on the protein N-terminus [13]. The signal sequence identified will be useful to express recombinant proteins having internal Kex2p cleavage sites using the Kex2p-disrupted *P. pastoris* strain. The Kex2 protease-inactivated *Pichia* strain is viable [27].

In the current study, we have demonstrated that the signal sequence of a DDDK protein can be used for the expression of recombinant proteins. Initially, we cloned the DDDK protein along with its signal sequence under the *AOX1* promoter (Fig. 2B) and successfully secreted the protein in wild-type *P. pastoris* strain (Fig. 4). The results of this experiment indicate three things. First, the 18-amino-acid signal sequence is able to secrete the proteins into the medium; secondly the strong promoter is able secrete the protein in wild-type *P. pastoris* using the native DDDK signal sequence; and thirdly, the levels of protein secreted with and without the Kex2p sites are similar (Fig. 4B lanes 3 and 4). It was further demonstrated with the use of anti-DDK antibody that the protein is being secreted to the medium only if the *PMT1* gene is inactivated, as there was no detectable DDDK protein band observed in the culture supernatant of wild-type *Pichia*, as shown in lane 1, Fig 5. The induced cells were lysed using glass beads and separated by SDS-PAGE along with the corresponding secretory protein. The western blot of the same gel demonstrated that the DDDK protein signal sequence was very efficiently processed, and all protein expressed was secreted to the medium as there was no intracellular protein accumulation, as shown in Fig. 5 lanes 5 and 6. There was another prominent protein band that showed up between 116 and 66 kDa. We have not yet characterized this protein band. This could be a protein complex, being formed by interaction between the mature DDDK protein and PMT1 protein. The mRNA transcripts levels of DDDK protein in wild-type *Pichia* and PMT1-inactivated *Pichia* strains are similar (data not shown). However, the mature DDDK protein is not being secreted from the wild-type strain, possibly due to interaction between DDDK and PMT1 proteins. This hypothesis regarding the interaction between these two proteins as a controlling step preventing the secretion of DDDK protein could be an interesting role that could be the object of a later study.

We also successfully expressed the heterologous recombinant

proteins such as porcine CPB and ETI (Fig. 6). This is an indication that the signal sequence can be used for heterologous protein expression. We have determined the intact mass of the ETI after gel-eluting the protein band from clone #1 (Fig. 6B lane 3) and found that the intact mass is matching with the theoretical mass (data not shown). This indicates the signal sequence is cleaved off at the appropriate position. The reported new signal sequence will be a useful addition to efficient signal sequences available for heterologous protein expression in *Pichia*.

Acknowledgments

References

1. Cereghino GPL, Cereghino JL, Ilgen C, Cregg JM. 2002. Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. *Curr. Opin. Biotechnol.* **13:** 329-332.

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

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

4. Govindappa N, Hanumanthappa M, Venkatarangaiah K, Kanojia K, Venkatesan K, Chatterjee A, *et al.* 2013. *PMT1* gene plays a major role in *O*-mannosylation of insulin precursor in *Pichia pastoris*. *Protein Expr. Purif.* **88:** 164-171.

5. Govindappa N, Nataraj N, Tiwari S, Hazra P, Patale M, Jothiraman G, *et al.* 2011. Novel prolipase-bovine trypsinogen fusion proteins. Patent No. WO/2011/030347 2011.

6. He Z, Huang Y, Qin Y, Liu Z, Mo D, *et al.* 2012. Comparison of alpha-factor preprosequence and a classical mammalian signal peptide for secretion of recombinant xylanase *xynB* from yeast *Pichia pastoris*. *J. Microbiol. Biotechnol.* **22:** 479-483.

7. Heimo H, Palmu K, Suominen I. 1997. Expression in *Pichia pastoris* and purification of *Aspergillus awamori* glucoamylase catalytic domain. *Protein Expr. Purif.* **10:** 70-79.

8. Hermann S, Gebhard von J. 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166:** 368-379.

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

10. Kannan V, Narayanaswamy P, Gadamsetty D, Hazra P, Khedkar A, Iyer H. 2009. A tandem mass spectrometric approach to the identification of *O*-glycosylated glargine glycoforms in active pharmaceutical ingredient expressed in *Pichia pastoris*. *Rapid Commun. Mass Spectrom.* **23:** 1035-1042.

11. Khasa YP, Conrad S, Sengul M, Plautz S, Meagher MM, Inan M. 2011. Isolation of *Pichia pastoris PIR* genes and their utilization for cell surface display and recombinant protein secretion. *Yeast* **28:** 213-226.

12. Killian JA, Ph. de Jong AM, Bijvelt J, Verklei AJ, de Kruijfff B. 1990. Induction of non-bilayer lipid structures by functional signal peptides. *EMBO J.* **9:** 815-819.

13. Kjeldsen T, Pettersson AF, Hach M. 1999. Secretory expression and characterization of insulin in *Pichia pastoris. Biotechnol. Appl. Biochem.* **29:** 79-86.

14. Mansur M, Martinez L, Perez M, Alonso-del-Rivero M, Marquez I, Proenza Y, *et al.* 2007. Expression, purification and characterization of porcine pancreatic carboxypeptidase B from *Pichia pastoris* for the conversion of recombinant human insulin. *Enzyme Microb. Technol.* **40:** 476-480.

15. Paifer E, Margolles E, Cremata J, Montesino R, Herrera L, Delgado JM. 1994. Efficient expression and secretion of recombinant alpha amylase in *Pichia pastoris* using two different signal sequences. *Yeast* **10:** 1415-1419.

16. Rapoport TA. 1992. Transport of proteins across the endoplasmic reticulum membrane. *Science* **258:** 931-936.

17. Romero PA, Lussier M, Sdicu AM, Bussey H, Herscovics A. 1997. Ktr1p is an **a**-1, 2-mannosyltransferase of *Saccharomyces cerevisiae*: comparison of the enzymic properties of soluble recombinant Ktr1p and Kre2p/Mnt1p produced in *Pichia pastoris*. *Biochem. J.* **321:** 289-295.

18. Sambrook J, Russell DW. 2001. *Molecular Cloning, A Laboratory Manual*, 3rd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

19. Schutter KD, Lin Y, Tiels P, Hecke AV, Glinka S, Weber-Lehmann J *et al.* 2009. Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nat. Biotechnol.* **27:** 561-566.

20. Sletta H, Tondervik A, Hakvag S, Vee Aune TE, Nedal A, Aune R, *et al.* 2007. The presence of N-terminal secretion signal sequences leads to strong stimulation of the total expression levels of three tested medically important proteins during high-cell-density cultivation of *Escherichia coli*. *Appl. Environ. Microbiol.* **73:** 906-912.

21. Stephan K, Zoltán K. 2012. Secretory signal peptide modification for optimized antibody-fragment expression-secretion in *Leishmania tarentolae*. *Microb. Cell Fact*. **11:** 97-106.

22. Strahl-Bolsinger S, Immervoll T, Deutzmann R, Tanner W. 1993. *PMT1*, the gene for a key enzyme of protein *O*-glycosylation in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **90:** 8164-8168.

23. Teixeira AV, Dowdle EBD, Botes DP. 1994. Syntheisis and

expression of a gene coding for *Erythrina* trypsin inhibitor (ETI). *Biochem. Biophys. Acta* **121:** 716-722.

24. Towbin H, Staehelin T, Gordon J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA* **76:** 4350-4354.

25. Towbin H, Kurien BT, Scofield RH. 2009. Protein Blotting and detection: methods and protocols. *Methods Mol. Biol.* **536:** 1-3.

26. Von Heijne G. 1990. The signal peptide. *J. Membr. Biol.* **115:** 195-201.

27. Werten MWT, de Wolf FA. 2005. Reduced proteolysis of secreted gelatin and Yps1-mediated α-factor leader processing in a *Pichia pastoris kex2* disruptant. *Appl. Environ. Microbiol.* **71:** 2310-2317.