Appl Microbiol Biotechnol (2015) 99:327–336
DOI 10.1007/s00253-014-6052-5

Enhancement in production of recombinant two-chain Insulin Glargine by over-expression of Kex2 protease in *Pichia pastoris*

Suma Sreenivas · Sateesh M. Krishnaiah · Nagaraja Govindappa · Yogesh Basavaraju · Komal Kanojia · Niveditha Mallikarjun · Jayaprakash Natarajan · Amarnath Chatterjee · Kedarnath N. Sastry

Received: 24 July 2014 / Revised: 20 August 2014 / Accepted: 26 August 2014 / Published online: 20 September 2014

Abstract Glargine is an analog of Insulin currently being produced by recombinant DNA technology using two different hosts namely *Escherichia coli* and *Pichia pastoris*. Production from *E. coli* involves the steps of extraction of inclusion bodies by cell lysis, refolding, proteolytic cleavage and purification. In *P. pastoris*, a single-chain precursor with appropriate disulfide bonding is secreted to the medium. Downstream processing currently involves use of trypsin which converts the precursor into two-chain final product. The use of trypsin in the process generates additional impurities due to presence of Lys and Arg residues in the Glargine molecule. In this study, we describe an alternate approach involving over-expression of endogenous Kex2 proprotein convertase, taking advantage of dibasic amino acid sequence (Arg-Arg) at the end of B-chain of Glargine. *KEX2* gene over-expression in *Pichia* was accomplished by using promoters of varying strengths to ensure production of greater levels of fully functional two-chain Glargine product, confirmed by HPLC and mass analysis. In conclusion, this new production process involving Kex2 protease over-expression improves the downstream process efficiency, reduces the levels of impurities generated and decreases the use of raw materials.

S. Sreenivas (✉) · N. Govindappa · Y. Basavaraju · K. Kanojia · N. Mallikarjun · J. Natarajan · A. Chatterjee · K. N. Sastry
Biocon Research Limited, Plot No.2&3, Phase IV, Bommasandra-Jigani Link Road, Bangalore 560099, Karnataka, India
e-mail: suma.sreenivas@biocon.com

S. M. Krishnaiah
Molecular Diagnostics Laboratory, Dept. of Microbiology & Biotechnology, Bangalore University, JnanaBharathi Campus, Bangalore 560056, Karnataka, India

Keywords Two-chain Glargine · Kex2 protease · Over-expression · *Pichia pastoris*

Introduction

Insulin Glargine (also referred as Glargine) is a slow acting analog of Insulin. It is similar in structure to Insulin, except for the replacement of A21 amino acid from asparagine to glycine in A-chain and the addition of two positively charged arginine (RR) residues at the C-terminus of B-chain. These changes shift the isoelectric point towards neutrality and result in micro-precipitation of Glargine at the site of injection. Glargine, administered by subcutaneous route, has duration of action of 24 h with no peak and gives good glycaemic control. It has lesser incidence of nocturnal hypoglycaemia (Poon and King 2010). Two recombinant microbial hosts are being employed for commercial manufacturing namely *Escherichia coli* and *Pichia pastoris*. In *E. coli*, the manufacturing process involves lysis of cells, recovery of inclusion bodies, refolding and cleavage of the peptide connecting A- and B-chains to yield two-chain product (Zimmerman et al. 2012).

In *P. pastoris*, Glargine is expressed as a secreted single-chain precursor consisting of N-terminal B-chain followed by A-chain with no connecting peptide. The Glargine precursor is treated with trypsin to separate A- and B-chains and is converted to the two-chain product (Hazra et al. 2009). The A and B-chains are connected by two disulfide bridges. Trypsin has the specificity to cleave at the C-terminus of both lysine and arginine residues (Olsen et al. 2004). Since, Glargine B-chain sequence has internal lysine and arginine residues, a number of fragmented impurities are generated following trypsin treatment. In an attempt to improve the downstream process efficiency, we have explored the advantage of using Kex2 protease (Kex2p), which cleaves specifically at the C-terminus of

dibasic amino acid residues KR and RR (Germain et al. 1992). Glargine has a unique primary amino acid sequence where the B-chain ends with dibasic amino acid sequence 'RR' which can be cleaved specifically by Kex2p converting the single-chain precursor to the two-chain product. Engineering the *P. pastoris* strain with increased levels of Kex2p would enable the in vivo processing and could lead to increased levels of two-chain Glargine in the medium.

The *Saccharomyces cerevisiae* Kex2 protease (also known as Kexin; EC 3.4.21.61) is a Ca^{2+}-dependent serine protease of subtilisin-like proprotein convertase family (Mizuno et al. 1989). It is encoded by the *KEX2* gene and is an endoprotease. It is well known that the yeast endoprotease Kex2p removes the signal peptides from the pre-proteins and releases the mature form of secreted proteins and thus plays a pivotal role in yeast secretory pathways. The high specificity of Kex2p is in contrast to the other members of the subtilisin family which are known to be degradative in their action. Kex2p and its mammalian homologues Furin (van den Ouweland et al. 1990), PC2 (Smeekens and Steiner 1990; Seidah et al. 1990) and PC3 (Smeekens et al. 1991; Seidah et al. 1991) are similar to each other than to the degradative subtilisins. The Kex2p can be secreted by deleting the transmembrane domain and the C-terminal tail region without affecting the activity of the protease (Brenner and Fuller 1992). In our study, we have used *P. pastoris* Kex2 proprotein convertase (UniProtKB/Swiss-Prot Q5J881). The gene encoding Kex2p has been identified in *Pichia* genome and has been annotated based on homology to the *S. cerevisiae KEX2* gene (De Schutter et al. 2009). We have used a shorter version of the *KEX2* gene for our study after deleting the C-terminal region.

When Glargine was expressed in *P. pastoris*, approximately 3 % of the recombinant protein is secreted as two-chain Glargine, after endogenous Kex2p cleavage at C-terminus of 'RR' sequence, while the rest is secreted as a single-chain precursor. We hypothesized that the level of Kex2p was not sufficient to carryout complete conversion of all the expressed molecules. The aim of this study was to determine if the majority of single-chain precursor can be converted to two-chain Glargine in vivo. Here, we report engineering of the *P. pastoris* genome by inserting additional copies of *KEX2* gene expression cassettes with promoters of varying strengths to increase the process efficiency. Plasmid constructs were developed by fusing the *KEX2* gene in frame with the strong constitutive glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter (Waterham et al. 1997) and inducible formaldehyde dehydrogenase (FLD1) promoter (Duan et al. 2009). These constructs were introduced to the *P. pastoris* host expressing Glargine precursor. The conversion of the precursor to the two-chain product was monitored by HPLC analysis of the supernatant, and the authenticity of the product was confirmed by peptide mass fingerprint analysis.

Materials and methods

Strains, plasmids and media

The *P. pastoris* strain GS115 (ATCC 20864) was used for expression of Glargine. The expression vector used was pPIC9K (Invitrogen, CA, USA) that provides the *AOX1* promoter and terminator sequences and the *S. cerevisiae* Mat-α signal sequence for secretion. The pPICZA vector (Invitrogen, CA, USA) was modified for the cloning of the Kex2p coding sequence in frame with different promoters. All yeast cultures were grown at 30 °C. The *E. coli* strain DH5α was used for routine cloning and propagation of plasmids. Yeast extract-peptone-dextrose (YPD) medium (containing per liter: 10 g yeast extract, 20 g peptone, and 20 g dextrose) was used for routine growth and sub-culturing of *P. pastoris* strains. YNBD agar plates (containing per liter: 13.4 g yeast nitrogen base without amino acids, 20 g dextrose and 20 g agar) was used for selection of the transformants. Luria broth/agar was used for culturing *E. coli*. Media components used were either from Himedia (Mumbai, India) or Difco (Becton, Dickinson and Company, MD, USA).

Construction of recombinant plasmids for expression in *P. pastoris*

The Glargine nucleic acid sequence was codon optimized for expression in *P. pastoris*. The synthetic Glargine coding gene was obtained from Geneart Inc., Germany. The gene was sub-cloned into the *Pichia* expression vector pPIC9K between *Xho*I and *Eco*RI sites. The resulting construct Glargine/pPIC9K (Fig. 1a) was verified by sequencing the gene of interest.

The Kex2p expression cassettes were developed by cloning *KEX2* gene 660 amino acid coding sequence (UniProtKB/Swiss-Prot Q5J881, M1 to D660) in frame with different promoters and cloned in pPICZA vector. The *P. pastoris* GS115 (ATCC 20864) genomic DNA was used as template for PCR amplifications. Briefly, the 477 bp GAP promoter was amplified by PCR using primers GAPFP and GAPRP. The 1980 bp *KEX2* gene (NCBI Reference Sequence: XM_002491154, +1 to +1980) was amplified by PCR using primers KexFP1 and KexRP1. The amplified products of GAP promoter and *KEX2* gene were gel purified (Qiagen, Germany) and used as template to generate the GAP-*KEX2* gene construct (Fig. 1b) by overlap PCR (Yon and Fried 1989) using primers GAPFP and KexRP1. It was cloned into pTZ57R vector (MBI Fermentas, Germany) and sequence verified. The GAP promoter-*KEX2* gene-fused DNA fragment was later cloned into the pPICZA vector between the *Bgl*II and *Eco*RI restriction sites. Similarly, using the GS115 genomic DNA, the FLD1 promoter was amplified by PCR using primers FLDFP and FLDRP and the *KEX2* gene was

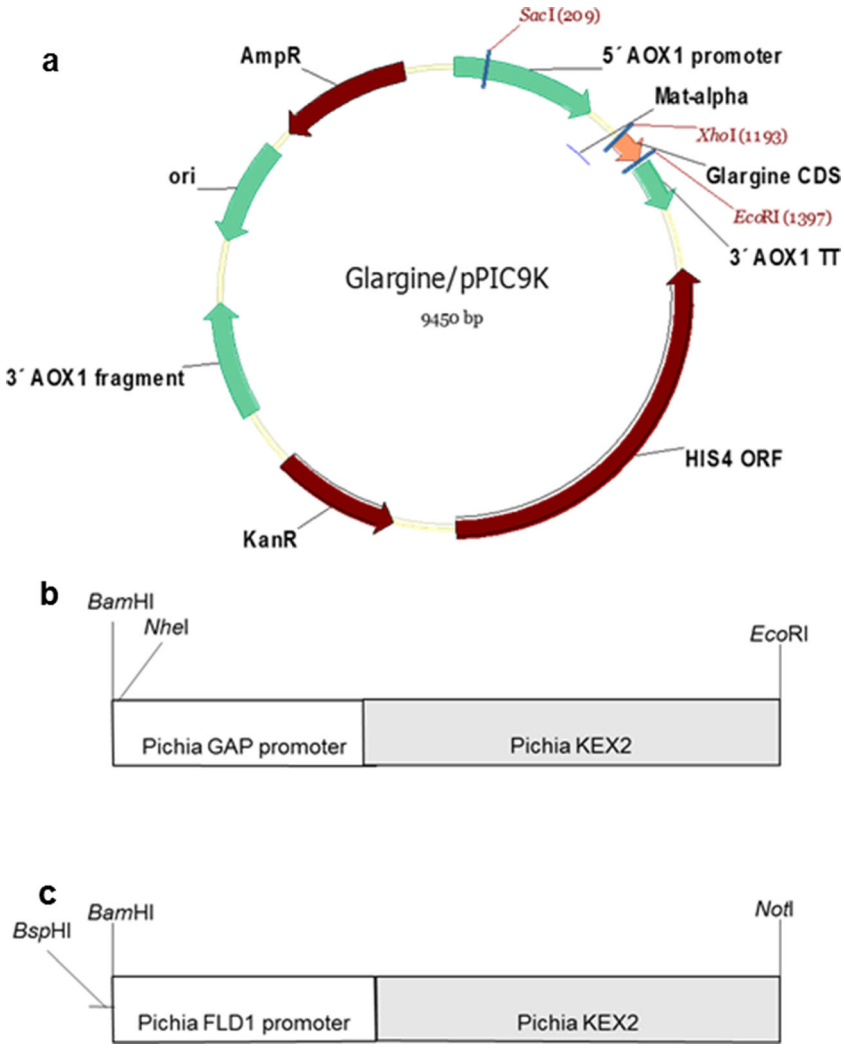

Fig. 1 **a** Map of Glargine/pPIC9K vector showing the Glargine nucleic acid sequence cloned between the *Xho*I and *Eco*RI restriction sites in frame with the AOX1 promoter. It contains the HIS4 ORF for metabolic selection in *P. pastoris* host. The unique *Sac*I site was used for linearizing the vector prior to transformation. **b** The schematic representation of GAP promoter and the *KEX2* gene products assembled by PCR and was sub-cloned to pPICZA vector between the *Bgl*II and *Eco*RI restriction sites. The restriction enzyme *Nhe*I was used for linearizing prior to transformation. **b** The caption of FLD1 promoter and the *KEX2* gene products which were assembled by PCR and further sub-cloned to pPICZA vector between *Bgl*II and *Not*I restriction sites. The restriction enzyme *Bsp*HI was used for linearizing the construct before transformation to *P. pastoris* host

amplified using the primers KexFP2 and KexRP2. The amplified gene products were purified using gel extraction kit and used as template to carry out overlapping PCR using primers FLDFP and KexRP2 (Fig. 1c). The FLD1 promoter-*KEX2* gene-fused DNA fragment was cloned to the pPICZA vector between the *Bgl*II and *Not*I sites. The primers sequence details are provided in Table 1. Expand High fidelity PCR system (Roche Diagnostics, GmbH, Germany) was used for all the PCR amplification steps.

Transformation of Glargine/pPIC9K plasmid into *P. pastoris*

The expression construct Glargine/pPIC9K (Fig. 1a) was linearized with *Sac*I restriction enzyme (New England Biolabs, USA) and transformed into histidine auxotrophic *P. pastoris* strain GS115. Transformation was carried out by electroporation of freshly grown cells in 0.2-cm cuvette. The pulse was delivered by Gene Pulser (Bio-Rad Laboratories Inc., USA) at 2000 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

Table 1 List of primers used in the study

Primer	Sequence (5′ to 3′)
GAPFP	GGA TCC GCT AGC GCT AGC TTT TTG TAG AAA TGT C
GAPRP	GGC AAA TAC ATA TAG TTG TTC AAT TGA TTG AAA T
KexFP1	TTG AAC AAC TAT ATG TAT TTG CCA GCA CTT CGC
KexRP1	GAA TTC CTT AAT CAC CGG TTG TAG AAG TCT C
FLDFP	GCG GAT CCG CAT GCA GGA ATC TCT GGC ACG G
FLDRP	CTG GCA AAT ACA TTG TGA ATA TCA AGA ATT GTA T
KexFP2	CTT GAT ATT CAC AAT GTA TTT GCC AGC ACT TCG C
KexRP2	GCG GCC GCT TAA TCA CCG GTT GTA GAA GTC TC
IGPFP	CAT TGT TCA ACA ATA CCT CTT CTA G
IGPRP	CAT TTG TGT GGT TCT CAT TTG G
GAPQFP	ATG ACC GCC ACT CAA AAG AC
GAPQRP	GCA CCA GTG GAA GAT GGA AT

spread on YNBD agar plates. The resulting transformants were screened on YPD plates containing progressively increasing concentrations of G418 (0.5–2.0 mg/ml) to identify multicopy clones. Several clones that were viable at 2 mg/ml concentration were selected for productivity check in small scale induction experiments. Culture supernatant samples were analysed by HPLC to identify the best expressing clone.

Induction of Glargine expression in *P. pastoris*

The selected transformants were grown in 10 ml YNBD medium (1.34 % yeast nitrogen base w/o amino acids and 2 % dextrose) overnight. It was next sub-cultured to 50 ml BGMY (1 % yeast extract, 2 % peptone, 1.34 % YNB, 100 mM potassium phosphate at pH 6.0 and 1 % glycerol) at a concentration of 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 5000 rpm for 5 min at room temperature and the cell pellet was resuspended in induction medium (1 % yeast extract, 2 % peptone, 1.34 % YNB and 100 mM potassium phosphate at pH 6.0). Methanol (20 % stock) was added to the culture medium for induction of protein expression at a final concentration of 1 %. The induction was carried out for 4 days with methanol being added to the medium once every 24 h. The cell-free supernatant samples were collected every day and analysed by HPLC.

Transformation of the Kex2p/pPICZA constructs to Glargine-producing *P. pastoris* host

The constructs corresponding to promoter-*KEX2* gene fusions (Fig. 1b, c) were linearized by digesting with the restriction enzymes *Nhe*I and *Bsp*H1 respectively and were independently transformed to the Glargine-producing *P. pastoris* host. The transformation procedure used was the same as described above, except that after allowing cells to recover in 1 M Sorbitol, one volume of YPD medium was added and the culture was incubated for 1 h at 30 °C. Cells were then plated on YPD medium supplemented with 1 M sorbitol and 100 μg/ml Zeocin (Life technologies, USA). The transformants obtained were screened for higher Zeocin resistance. The colonies resistant to 1 mg/ml Zeocin were chosen for induction and evaluation of two-chain Glargine secretion.

Estimation of copies of *Glargine* gene using *GAP* gene as a single-copy housekeeping reference

The *Glargine* gene copy was determined by quantitative real-time PCR (QPCR) using relative quantitation method (Li et al. 2004). The single-copy *GAP* gene in *Pichia* genome was used as the reference gene to determine the integrated *Glargine* gene copies. The QPCR was set up using primers GAPQFP and GAPQRP for amplification of the *GAP* gene and IGPFP and IGPRP for amplification of the *Glargine* gene. The genomic DNA template concentration of 1, 10, 100, 1000 and 10,000 picograms were used to obtain consensus results. SYBR green chemistry was used for monitoring the amplification and data acquisition.

HPLC assay

HPLC assay for measuring the levels of Glargine precursor as well as the two-chain Glargine product were performed using C18 column (250×4.6 mm; 5 μm from ACE) under a gradient run. The mobile phase consisted of A 0.1 % trifluoroacetic acid (TFA) (75 %) and B 25 % acetonitrile. The runtime was 18 min where the component B composition decreased from 75 to 60 % in 15 min at a flow rate of 1.0 ml/min. A sample volume of 50 μl was injected into the column maintained at 40 °C, and the sample was monitored at 220 nm.

Liquid chromatography/mass spectrometry

The Glargine sample from shake flask fermentation was purified on semi-preparative reverse phase HPLC method in Agilent 1200 chromatography system (Agilent, CA, USA) using Supelco Discovery BIO Wide Pore C18 reverse phase column (250×10 mm, 5 μm). About 200 μg of crude protein was loaded on to the column, and temperature was maintained at 40 °C. The eluents were monitored at 214 nm, and the peak corresponding to Glargine was purified up to 98 %. The purified sample was lyophilized under vacuum and used for analysis.

From the lyophilized powder, about 1 mg/ml of Glargine was dissolved in 0.01 N HCl and complete physico-chemical characterization was done in comparison with commercially available Lantus®. Identification of Glargine was carried out on Thermo Scientific LTQ Orbitrap coupled with chromatography system (Shimadzu, DGU-20A model, Japan) using ACE symmetry C18 column (250×4.6 mm, 5 μm). The mobile phase solvents consisted of A 0.1 % TFA and B 100 % acetonitrile. The purified Glargine sample was eluted using a linear gradient program with initial 25 % of B and eluted with 37 % of acetonitrile with the flow rate of 0.7 ml/min. About 5 μl of sample (Glargine or Lantus®) was injected to the column maintained at 40 °C, and the eluents were monitored at 214 nm. The mass spectrometer was operated in electrospray ionization (ESI) positive ionization mode with the following mass spectrometry parameters: source voltage of 3.5 kV and capillary temperature of 340 °C. The sheath gas flow was kept at 40, aux gas flow at 20, and the scan was performed in the mass range 700–2200 Da, and the data acquired from Orbitrap (FT) was analysed.

Peptide mass fingerprint analysis

Both Glargine and Lantus® were digested with endoprotease Glu-C with a sample: enzyme *w/w* ratio of 1:40. For this, 100 μl (1 mg/ml) of protein was first pH adjusted to 7.5 using Tris buffer (pH 8.0) followed by the addition of Glu-C protease. The reaction mixture was incubated at 37 °C for 4 h. The digested samples were reduced using 5 μl of 1 M DTT and incubated at 37 °C for 30 min. Further analysis was performed on both non-reduced and reduced digested sample, to confirm disulfide bridge connectivity and sequence information.

Digested samples were separated on a Waters Symmetry C18 column (250×4.6 mm, 5 μm) at a flow rate of 0.7 ml/min using the gradient program 0–60 min, 5–80 % B, solvent A 0.1 % aqueous TFA and solvent B 95 % acetonitrile, and column temperature was maintained at 40 °C. A sample

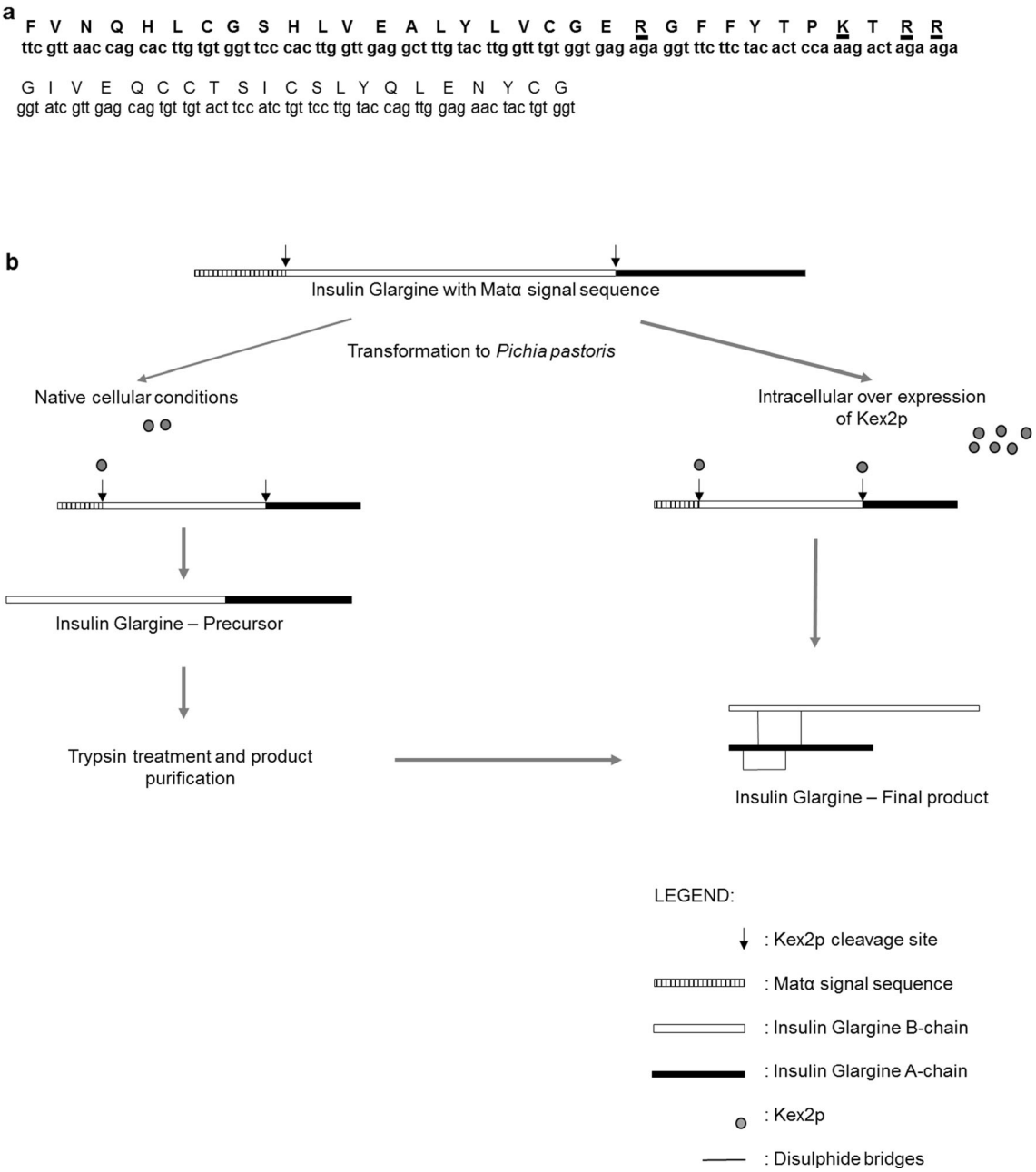

Fig. 2 Strategy for direct secretion of two-chain Insulin Glargine in *P. pastoris*. **a** The amino acid sequence of Insulin Glargine used in this study where the B-chain (shown in *bold*) is followed by the A-chain sequence. The amino acids underlined are the lysines and arginines susceptible to trypsin cleavage. The codon-optimized nucleotide sequence for expression in *P. pastoris* is shown below the corresponding amino acids. **b** Cartoon representation of the steps in the conversion of precursor to final product by conventional method using trypsin in comparison to the use of Kex2p

volume of 25 μl was injected, and peaks eluted were monitored at 214 nm. The thus separated Glu-C-digested peptides were further analysed by ESI-MS in order to generate peptide mass fingerprint data.

Results

Strategy and construction of expression plasmids

The amino acid sequence and the codon-optimized nucleotide sequence of Glargine used in the study are shown in Fig. 2a where the B-chain (marked in bold) ends with RR and is followed by the A-chain without any connecting C-peptide. Also marked are the trypsin-susceptible lysines and arginines present in the B-chain, which have been underlined. Figure 2b represents the outline of the strategy. Marked in the figure with arrows are the potential Kex2p cleavage sites. The Glargine coding sequence was cloned in the pPIC9K vector in frame with the Mat-α signal sequence to enable secretion of the protein. The *KEX2* gene was fused to the C-terminus of different promoters by PCR and cloned into the pPICZA vector. All the constructs were verified by sequencing and were linearized prior to transformation to the *P. pastoris* host.

Expression of Glargine in *P. pastoris* host

The Glargine/pPIC9K construct (Fig. 1a) was transformed into the *P. pastoris* GS115 strain. The colonies which showed resistance to 2 mg/ml of G418 were confirmed by PCR for the presence of the Glargine coding sequence integrated into the genome. The clones selected were screened for the Glargine productivity by methanol induction as described previously in shake flasks. Clone #9F12 consistently produced higher quantity of Glargine. Copy number analysis of this clone was performed by QPCR as described earlier. The QPCR experiment raw data was analysed for the linearity of the standard curve for both *Glargine* and endogenous *P. pastoris GAP* genes (Fig. 3a). The correlation coefficients for the *GAP* and the *Glargine* genes were 0.9994 and 0.9998, respectively. On analysis of the melt curve, the Tm for GAP and Glargine gene products are 81.05 °C and 76.61 °C, respectively. The gene copy index (GCI) was found to be 3. The Ct values obtained using various DNA concentrations are shown in Table 2. The Glargine secreted to the medium was analysed using HPLC method as described previously, and the chromatogram obtained is shown in Fig. 4a. The major peak observed at the retention time (RT) of 9.8–10 min corresponds to the Glargine precursor species. The productivity of Glargine precursor was found to be 56.3 mg/l. The two-chain Glargine product secreted to the medium was only 1.25 mg/l.

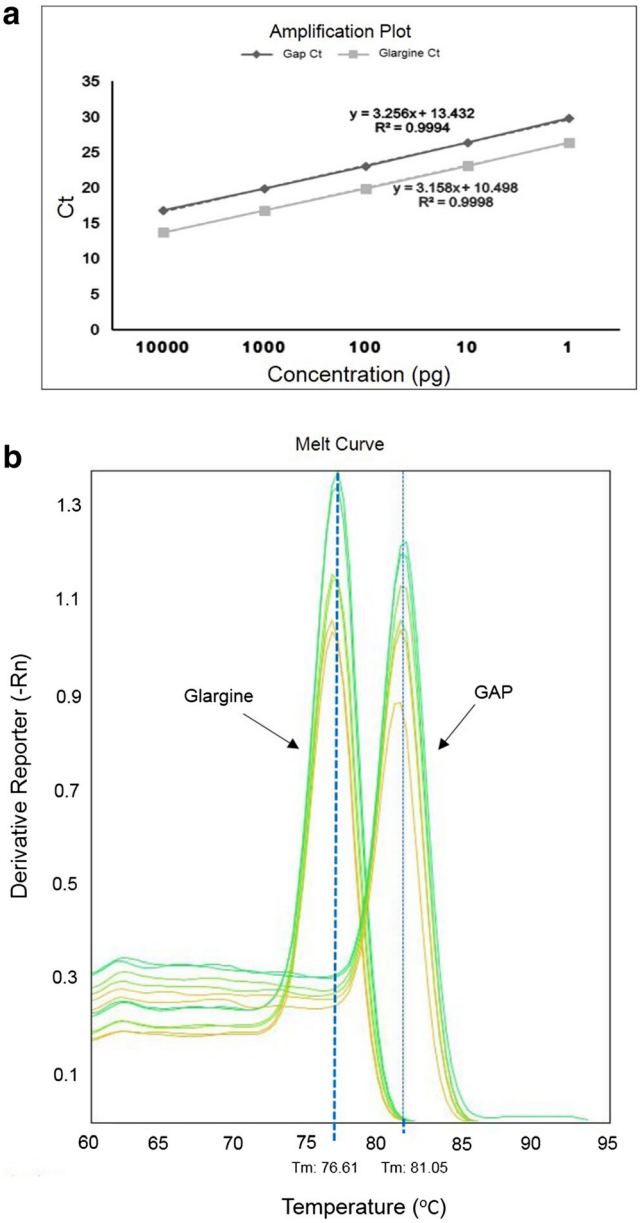

Fig. 3 QPCR data analysis for determination of the GCI of Glargine (clone #9F12). **a** The standard curve for *Glargine* and *GAP* genes. The R^2 of the standard curve obtained for both the genes are 0.9994 and 0.9998, respectively. **b** The melt curve obtained for the GAP and the Glargine genes shows a specific melting temperature (*Tm*) for each gene indicating specific product formation

Intracellular expression of Kex2p using GAP promoter

The *P. pastoris* clone #9F12 producing Glargine was transformed with Kex2p construct (Fig. 1b) where a constitutive GAP promoter was used for the expression of *KEX2* gene. Colonies which grew on 1 mg/ml of Zeocin were selected for evaluation of the levels of two-chain Glargine in the medium. Cell-free supernatants of induced cultures were analysed by HPLC. The chromatogram showed the presence of both

Table 2 Results of the GCI analysis for *Glargine* gene by QPCR

Concentrations of DNA	Ct Glargine	Ct GAP	Δ Ct	Average Δ Ct
10 ng	13.72	16.82	3.1	3.12
1 ng	16.81	19.88	3.07	
100 pg	19.88	23.06	3.18	
10 pg	23.17	26.30	3.13	

The Ct values obtained for the PCR amplification of *Glargine* gene and the single-copy endogenous *GAP* gene using different genomic DNA concentrations of *P. pastoris* clone #9F12. Δ Ct is the difference in the Ct values between *Glargine* and *GAP* gene. Average Δ Ct is the average of the Δ Cts obtained for each DNA concentration

precursor and two-chain Glargine species (Fig. 4b). The peak which is between RT of 9.8–10 min was the precursor Glargine species, and the peak between RT of 11.5–11.8 min corresponds to the processed two-chain Glargine.

Intracellular expression of Kex2p using FLD1 promoter

The Glargine-secreting *P. pastoris* clone #9F12 was transformed with Kex2p construct (Fig. 1c) where strong inducible FLD1 promoter was fused in frame with the *KEX2* gene. Colonies which grew on 1 mg/ml of Zeocin were evaluated for two-chain Glargine levels in the medium. The clones were induced using methanol, and the day 4 samples were analysed by HPLC. The HPLC profiles of the cell-free supernatant indicated that the major peak at RT of 11.5–11.8 min corresponds to the two-chain Glargine species (Fig. 4c).

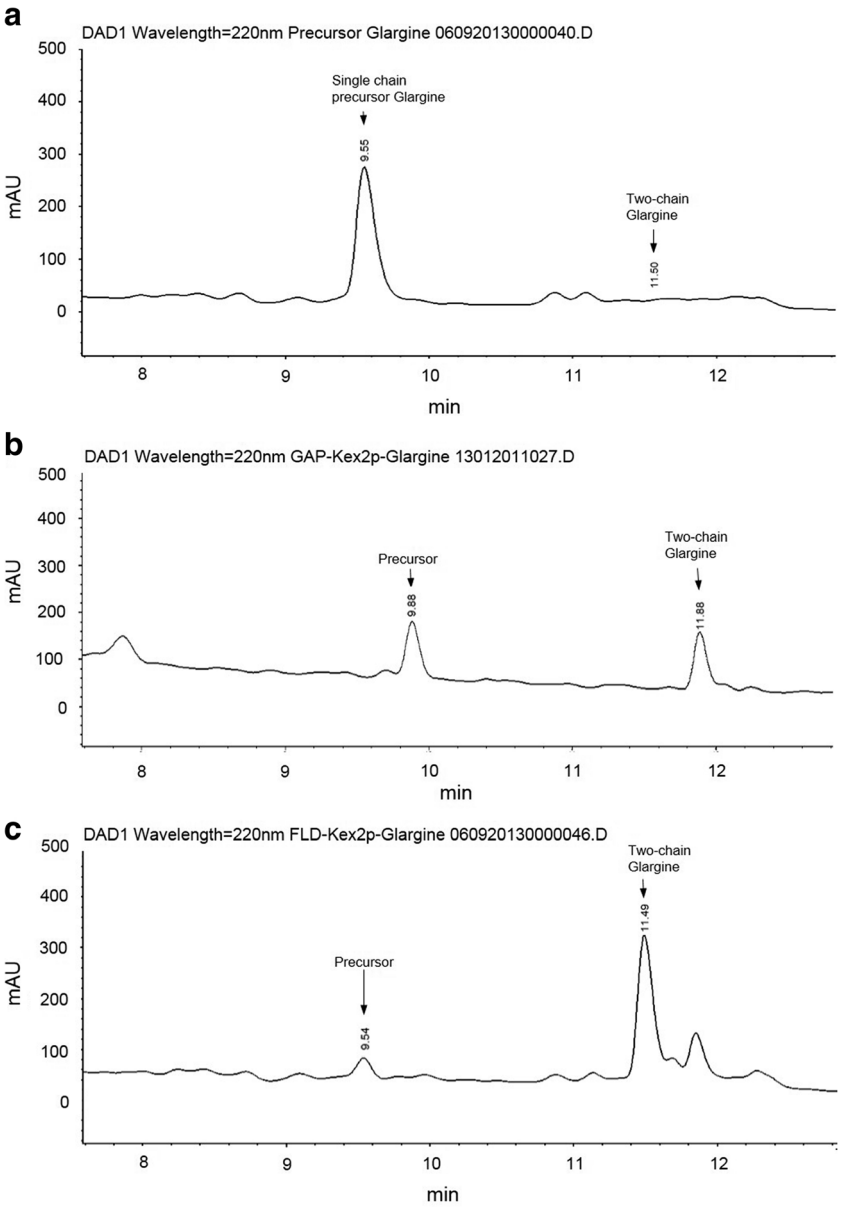

Fig. 4 **a** HPLC chromatogram of cell-free supernatant of *P. pastoris* strain expressing Glargine (clone #9F12). The peak marked by an arrow at the left shows the prominent single-chain precursor being secreted by the host. The peak marked at the right shows very low levels of the two-chain Glargine. **b** HPLC chromatogram of cell-free supernatant of *P. pastoris* strain (clone #9F12) with intracellular co-expression of Kex2p using GAP promoter which shows conversion of ~50 % precursor to two-chain Glargine**. c** HPLC profile of the cell-free supernatant of *P. pastoris* strain (clone #9F12) with intracellular co-expression of Kex2p using strong FLD1 promoter, showing >90 % two-chain Glargine

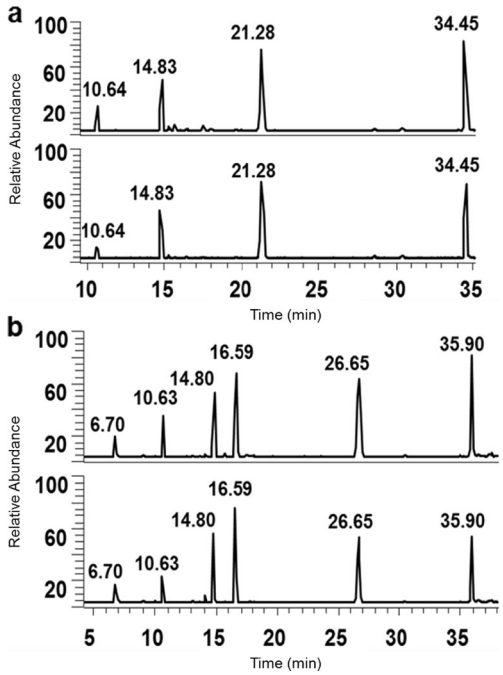

Fig. 5 The total ion chromatogram of peptide mass fingerprint analysis of Glargine in comparison to Lantus®, showing the elution of different peptides with respective RTs. **a** The endoprotease Glu-C digest of the samples shows four expected peptides for Glargine (*top*) and Lantus® (*bottom*). **b** DTT reduced endoprotease Glu-C digest which shows the presence of six expected peptides of Glargine (*top*) and Lantus® (*bottom*). The sequence was further confirmed by low energy CID MS/MS, and the data is shown in Tables 3 and 4, respectively

Mass spectroscopy analysis

The two-chain Glargine secreted from the clone #9F12 with Kex2p over-expressed using FLD1 promoter was analysed by mass spectroscopy in comparison with Lantus®. The deconvoluted mass spectrum of the sample showed the mass to be 6063.9 Da $(M+H)^+$.

The peptide mass fingerprint analysis results for Glargine sample was compared to that of Lantus®. It was observed that the converted two-chain Glargine sequence matches exactly with the Lantus® in terms of mass-based identities of the Glu-C-digested peptides. This confirms that once the Kex2p cleaves at the C-terminal of RR sequence at the end of B-chain, the completely processed two-chain Glargine is secreted to the medium. Figure 5 shows the results of the peptide mass fingerprint data, where the Glargine precursor sample from the shake flask is compared to Lantus®. The peaks corresponding to the different peptides from Glargine correlated well with that of Lantus®. The data of the identity of the different peptides under non-reduced and reduced conditions when the samples are digested with Glu-C is depicted in Tables 3 and 4, respectively.

Discussion

The methylotrophic yeast *P. pastoris* has been proven to be a suitable host for production of recombinant proteins (Cereghino and Cregg 2000; Cregg et al. 2000). We expressed single-chain Glargine precursor in the yeast *P. pastoris* as a secreted protein using the Mat-α signal sequence. It was secreted mostly as a single-chain precursor and was converted to the two-chain product by a process which involved the use of trypsin. We have explored the use of Kex2p for the processing of the Glargine precursor in vivo to secrete the two-chain molecule to the medium. The Glargine amino acid sequence presents an opportunity where the enzymatic activity of Kex2p can convert the precursor to the final two-chain Glargine molecule. It was observed that when Glargine precursor was expressed in the *P. pastoris* host (clone #9F12); only about 3 % of the precursor was processed into two-chain Glargine, and the rest of the secreted recombinant protein was the single-chain precursor (Fig. 4a). This result indicated that endogenous Kex2p produced is insufficient to complete this process. Limitations in the amount of Kex2p on intracellular expression were also reported (Seeboth and Hein 1991). Co-expression of the truncated version of Kex2p resulted in the correct processing of the *Rhizopus oryzae* lipase in

Table 3 Theoretical and observed peptides following endoprotease Glu-C digest of Glargine and Lantus®

Peptide no.	Peptide range	RT of Glargine and Lantus® (min)	Theoretical mass $(M+H)^+$ (Da)	Observed mass $(M+H)^+$ (Da)		Sequence	Linked peptide
				Glargine	Lantus®		
I	A (1–4)	10.64	417.2	417.24	417.23	GIVE	–
II	B (22–32)	14.83	1428.8	1428.79	1428.79	RGFFYTPKTRR	–
III	A (18–21) and B (14–21)	21.28	1320.6	1320.57	1320.56	(NYCG) and (ALYLVCGE)	Disulfide linkage between A20 and B19
IV	A (5–17) and B (1–13)	34.45	2968.30	2968.34	2968.34	(QCCTSICSLYQLE) and (FVNQHLCGSHLVE)	Disulfide linkage between A6-A11 and A7-B7

The linked peptide confirms the appropriate disulfide connectivity. The mass in the table represents monoisotopic mass

Table 4 Theoretical and observed peptides of endoprotease Glu-C digestion followed by DTT reduction of Glargine and Lantus®

Peptide no.	Peptide range	RT of Glargine and Lantus® (min)	Theoretical mass $(M+H)^+$ (Da)	Observed mass $(M+H)^+$ (Da)		Sequence
				Glargine	Lantus®	
I	A (18–21)	6.70	456.1	456.16	456.16	NYCG
II	A (1–4)	10.63	417.2	417.24	417.24	GIVE
III	B (1–13)	16.59	1482.7	1482.73	1482.73	FVNQHLCGSHLVE
IV	B (22–32)	14.80	1428.8	1428.80	1428.81	RGFFYTKPTRR
V	B (14–21)	26.65	867.4	867.43	867.43	ALYLVCGE
VI	A (5–17)	35.90	1490.6	1490.64	1490.64	QCCTSICSLYQLE

The sequence of all the peptides was confirmed by CID MS/MS. The mass in the table represents monoisotopic mass

S. cerevisiae (Takahashi et al. 2000). Therefore, an effort to improve the conversion efficiency was initiated by over-expression of Kex2p in the *P. pastoris* clone #9F12. We expressed the truncated version of the Kex2p (M1 to D660) secreting into the medium in Glargine-producing *P. pastoris* clone #9F12 under the regulatory control of GAP promoter. The secreted Glargine sample was analysed by HPLC and found that the conversion of the precursor to the two-chain Glargine product was not enhanced (data not shown). This might be possibly due to accessibility, and conditions prevalent in the medium may not be optimal for the functional activity of the Kex2p. So, the over-expression of the Kex2p intracellularly to improve the processing efficiency was attempted. It has been reported that an incorrectly processed IFNα-2a molecules were secreted from the precursor with N-terminal extensions of the mature protein when expressed in *Hansenula polymorpha* (Muller et al. 2002). Correct processing was observed when *KEX2* gene was over-expressed in the same host using its native promoter. The yields of many recombinant proteins were greatly influenced by Kex2 P1′ site residues, and the optimized residue could determine the final amount of secretory protein. A further improvement of the yield can be accomplished by integration of additional copies of *KEX2* gene (Yang et al. 2013).

In this study, we initially used a constitutive *GAP* promoter for intracellular expression of Kex2p in the *P. pastoris* host secreting Glargine. Approximately 40–60 % of two-chain Glargine species was detected extracellularly (Fig. 4b) indicating that the intracellular over-expression of Kex2p was enhancing the conversion. But the quantity of protease produced using GAP promoter was still not sufficient to achieve complete processing. Hence, in order to improve the conversion efficiency, the Kex2p was co-expressed intracellularly with strong inducible FLD1 promoter which is also induced by methanol in the Glargine-producing host. In this case, >90 % of the secreted recombinant protein was found to be the two-chain Glargine. The copy number of the *KEX2* gene integrated using different promoter's influenced the production of recombinant protein; hence, neglecting the gene copy numbers can lead to incorrect interpretation of the results (Abad et al. 2010). Therefore, an attempt was made to understand whether the improvement in efficiency of conversion using FLD1 promoter-driven *KEX2* gene is due to higher copy numbers integrated or due to the relative promoter strength. GCI of *KEX2* gene cassettes integrated using the GAP and FLD1 promoters was determined. Relative quantitation was performed using protocol similar to the one used for determining the copy number of Glargine gene. The GCI of *KEX2* gene was found to be two for both GAP and FLD1 promoter-driven Kex2p over-expressing host strains (data not shown). The Kex2p over-expressing clones have same copies of the cassette integrated in the genome. This result indicates that the increase in the conversion rates is due to the difference in the promoter's strength.

The two-chain Glargine species that was secreted from FLD1 promoter-*KEX2* gene over-expressing host was characterized by mass analysis and peptide mass fingerprinting. It was compared to the commercially available Lantus®. The

Table 5 Quantitative estimation of % of two-chain Glargine levels from Glargine-expressing *P. pastoris* hosts with Kex2p over-expression

Clone	Precursor peak area (A)	Two-chain Glargine peak area (B)	Total area C=(A+B)	% two-chain Glargine B/(C)×100
Clone #9F12 expressing Glargine in *P. pastoris* host (Fig. 4a)	2887	64	2951	2.27
GAP-KEX2 expressed in clone #9F12 (Fig. 4b)	807	797	1604	49.68
FLD1-KEX2 expressed in clone #9F12 (Fig. 4c)	231	2611	2842	91.87

The peak areas are obtained when the samples were analysed by HPLC

Data shown is average of two different experiments

mass obtained was 6063.9 Da which compares well with the theoretical mass and the mass of Lantus®. The peptide mass fingerprinting data also showed that the Glargine secreted by the *P. pastoris* host has appropriate disulfide linkages. The percentage conversions of precursor to two-chain Glargine using various constructs are tabulated in Table 5. The results obtained clearly indicate that the endogenous Kex2p produced is not sufficient to cleave the entire quantity of expressed single-chain precursor. Only upon co-expression of Kex2p under strong promoter, the two-chain Glargine is being generated with a high efficiency.

We envisage that using the aforementioned strategy for generating a two-chain-processed molecule directly into the medium clears the hurdles involved in removing many product-related impurities arising during downstream processing in the earlier process employing trypsin. This new approach is anticipated to simplify the purification process and reduce the cost of commercial production.

Acknowledgments We thank Molecular Biology group of R&D for their support and suggestions. We thank Biocon Research Limited for supporting this work.

References

Abad S, Kitz K, Hormann A, Schreiner U, Hartner FS, Glieder A (2010) Real-time PCR-based determination of gene copy numbers in *Pichia pastoris*. Biotechnol J 5(4):413–420

Brenner C, Fuller RS (1992) Structural and enzymatic characterization of a purified prohormone-processing enzyme: secreted soluble kex2 protease. Proc Natl Acad Sci U S A 89(3):922–926

Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol Rev 24(1): 45–66

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

De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouze P, Van de Peer Y, Callewaert N (2009) Genome sequence of the recombinant protein production host *Pichia pastoris*. Nat Biotechnol 27(6):561–566

Duan H, Umar S, Hu Y, Chen J (2009) Both the AOX1 promoter and the FLD1 promoter work together in a *Pichia pastoris* expression vector. World J Microbiol Biotechnol 25:1779–1783

Germain D, Dumas F, Vernet T, Bourbonnais Y, Thomas DY, Boileau G (1992) The pro-region of the Kex2 endoprotease of *Saccharomyces cerevisiae* is removed by self-processing. FEBS Lett 299(3):283–286

Hazra P, Dave N, Kannan V, Tiwari S, Goel A, Iyer H, Roy N, Venkatesan K, Vasudevan A, Jagadish A, Sachdev G, Patale MB (2009) A method of obtaining purified heterologous insulins expressed in yeast. WO/2009/104199 A1

Li Z, Hansen JL, Liu Y, Zemetra ZS, Berger PH (2004) Using real-time PCR to determine transgene copy number in wheat. Plant Mol Biol Rep 22:179–188

Mizuno K, Nakamura T, Ohshima T, Tanaka S, Matsuo H (1989) Characterization of KEX2-encoded endopeptidase from yeast *Saccharomyces cerevisiae*. Biochem Biophys Res Commun 159(1):305–311

Muller F, Tieke A, Waschk D, Muhle C, Seigelchifer M, Pesce A, Jenzelewski V, Gellissen G (2002) Production of IFNα in *Hansenula polymorpha*. Process Biochem 38(1):15–25

Olsen JV, Ong SE, Mann M (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol Cell Proteomics 3(6): 608–614

Poon K, King AB (2010) Glargine and detemir: safety and efficacy profiles of the long-acting basal insulin analogs. Drug Healthc Patient Saf 2:213–223

Seeboth PG, Hein J (1991) In-vitro processing of yeast alpha factor leader fusion proteins using a soluble yscF (Kex2) variant. Appl Microbiol Biotechnol 35:771–776

Seidah NG, Gaspar L, Mion P, Marcinkiewicz M, Mbikay M, Chrttien M (1990) cDNA sequence of two distinct pituitary proteins homologous to kex2 and furin gene products: tissue specific mRNAs encoding candidates for prohormone processing proteinases. DNA Cell Biol 9:415–424

Seidah NG, Marcinkiewicz M, Benjannet S, Gaspar L, Beaubien G, Mattei MG, Lazure C, Mbikay M, Chrdtien M (1991) Cloning and primary sequence of a mouse candidate prohormone convertase PC1 homologous to PC2, Furin and kex2: distinct chromosomal localization and messenger RNA distribution in brain and pituitary compared to PC2. Mol Endocrinol 5:111–122

Smeekens SP, Steiner DF (1990) Identification of human insulinoma cDNA encoding a novel mammalian protein structurally related to the yeast dibasic processing protease kex2. J Biol Chem 265:2997–3000

Smeekens SP, Avruch AS, LaMendola J, Chan SJ, Steiner DF (1991) Identification of a cDNA encoding a second putative prohormone convertase related to PC2 in AtT20 cells and islets of Langerhans. Proc Natl Acad Sci U S A 88:340–344

Takahashi S, Ueda M, Tanaka A (2000) Effect of truncation of the C-terminal region of Kex2 endoprotease on processing of the recombinant *Rhizopus oryzae* lipase precursor in the co-expression system in yeast. J Mol Catal B Enzym 10:233–240

van den Ouweland AM, van Duijnhoven HL, Keizer GD, Dorssers LC, van de Ven WJ (1990) Structural homology between the human fur gene product and the subtilisin-like protease encoded by yeast KEX2. Nucleic Acids Res 18(5):1332

Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM (1997) Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene 186:37–44

Yang S, Kuang Y, Li H, Liu Y, Hui X, Li P, Jiang Z, Zhou Y, Wang Y, Xu A, Li S, Liu P, Wu D (2013) Enhanced production of recombinant secretory proteins in *Pichia pastoris* by optimizing Kex2 P1' site. PLoS ONE 8(9):e75347. doi: 10.1371/journal.pone.0075347

Yon J, Fried M (1989) Precise gene fusion by PCR. Nucleic Acids Res 17(12):4895

Zimmerman RE, Stokell DJ, Akers MP (2012) Glargine proinsulin and methods of producing glargine insulin analogs therefrom. US 20120214965 A1