5
Expression of Insulin in Yeast: The Importance of Molecular Adaptation for Secretion and Conversion

THOMAS KJELDSEN*, PER BALSCHMIDT, IVAN DIERS, MORTEN HACH, NIELS C. KAARSHOLM AND SVEND LUDVIGSEN

Novo Nordisk A/S, Novo Alle 6B S.90, 2880 Bagsværd, Denmark

Introduction

The globular, two-chain and 51 amino acid residue peptide-hormone insulin is produced and secreted by the β-cells of the pancreatic islets of Langerhans. Insulin is synthesized as preproinsulin (110 amino acids). The pre-peptide (signal peptide) is removed upon entrance into the endoplasmic reticulum. Proinsulin folds in the endoplasmic reticulum, is transported to the Golgi apparatus and subsequently processed into the mature insulin molecule that is stored in well-defined storage vesicles (*Figure 5.1*) (Steiner *et al.*, 1967, 1986; Dodson and Steiner, 1998). Proinsulin and insulin have self-assembling properties that play an important role in processing and storage in the β-cell's secretory pathway and both associate to dimers and in the presence of zinc these further assemble into hexamers (Dodson and Steiner, 1998). In the late Golgi apparatus proinsulin is targeted to acidifying secretory granules and conversion of proinsulin to insulin occurs by removal of the C-peptide by cleavage at dibasic processing sites by the endoproteases PC3 (or PC1) and PC2 (mammalian

*To whom correspondence should be addressed (thk@novonordisk.com)

Abbreviations: α-factor, *Saccharomyces cerevisiae* mating factor α, a 13-amino-acid-residue peptide pheromone; BiP/Kar2 protein, chaperone encoded by the *Saccharomyces cerevisiae KAR2*; CD, circular dichroism; ES–MS, electro-spray mass spectrometry; Δ, deletion; Da, Daltons; des(B30) insulin, human insulin lacking amino acid threonine^{B30}; *FLP*, 2 μm plasmid gene encoding a site-specific recombinase; GuHCl, guanidine hydrochloride; Kex2 endoprotease, endopeptidase (EC 3.4.21.61) processing *Saccharomyces cerevisiae* α-factor precursor; MALDI–MS, matrix assisted laser desorption ionization mass spectrometry; NMR, nuclear magnetic resonance spectroscopy; *POT*, *Schizosaccharomyces pombe* triose phosphate isomerase gene; RP-HPLC, reverse phase high performance liquid chromatography; SDS-PAGE, sodium dodecyl sulphate polyacrylamide gel electrophoresis; *STE13*, *Saccharomyces cerevisiae* gene encoding dipeptidyl aminopeptidase A; TPI, triose phosphate isomerase; *TPI1*, *Saccharomyces cerevisiae* gene encoding TPI; Yap3 endoprotease, yeast aspartyl endoprotease 3 (Yapsin 1); YPD, 1% yeast extract/2% peptone/2% glucose.

Biotechnology and Genetic Engineering Reviews – Vol. 18, July 2001
0264–8725/01/18/89–121 $20.00 + $0.00 © Intercept Ltd, P.O. Box 716, Andover, Hampshire SP10 1YG, U.K.

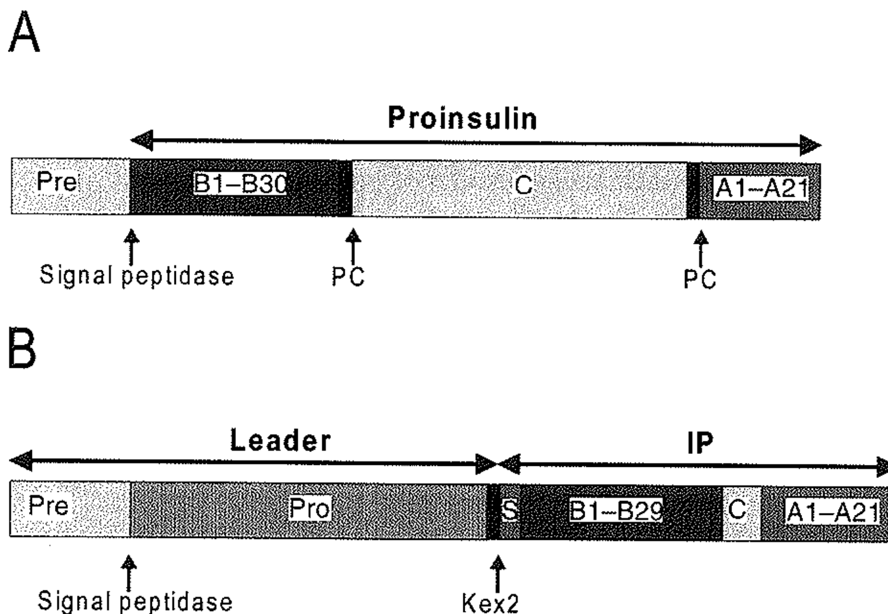

Figure 5.1. Schematic comparison of preproinsulin (A) and the insulin molecule adapted to expression and secretion in yeast (B). A: In the β-cells insulin is synthesized as preproinsulin. The signal peptide (Pre) is removed upon entrance into the endoplasmic reticulum, and subsequently proinsulin is folded and the C-peptide (C) removed by prohormone convertase (PC) cleavage at dibasic processing sites (shown in black) located at the C-peptides' termini. The B-chain 'B1–B30' and the A-chain 'A1–A21' are indicated. Sites for processing by the signal peptidase and for processing by the prohormone convertases (PC) are shown. B: Proinsulin was not readily expressed in yeast (Thim *et al.*, 1986). However, insulin could be expressed in yeast as a single-chain insulin precursor fused to a leader. The leader consists of a pre-peptide and of a pro-peptide. Here, a genetically engineered leader (leader) lacking N-linked glycosylation is indicated. 'Pre' indicates the pre-peptide (the Yap3 signal peptide) and 'Pro' indicates the genetically engineered pro-peptide. The dibasic Kex2 endoprotease (a functional equivalent of mammalian prohormone convertases) processing site (KR) located at the C-terminus of the pro-peptide is shown in black. The single-chain insulin precursor (IP) comprises the first 29 amino acids of the human insulin B-chain (B1–B29) joined to the 21 amino acids of the human insulin A-chain (A1–A21) by the mini C-peptide (C) which connects Lys^{B29} and Gly^{A1}. The fusion protein also features a spacer peptide (S) for optimizing Kex2 endoprotease processing. Sites for processing by the signal peptidase and for processing by the Kex2 endoprotease are indicated.

functional equivalents of the *Saccharomyces cerevisiae* Kex2 endoprotease). Subsequently, a carboxypeptidase removes the two basic amino acid residues from the C-terminus of the B-chains. Removal of the C-peptide changes the solubility properties and the insulin hexamer forms micro-crystals in the storage vesicles.

Insulin is essential for maintaining glucose homeostasis and following an increase in the blood glucose level, insulin is secreted by the β-cells in a highly regulated fashion. Insulin binds to specific receptors in various target tissues, such as muscle, liver and adipose tissue, increasing glucose uptake and subsequently metabolism, as well as suppressing hepatic glucose output. Diabetes mellitus is a metabolic disorder characterized by chronic hyperglycaemia, in which glucose is not taken up and metabolized normally by the cells because of inadequate insulin concentration. This

may be due to β-cell destruction (in Type 1 diabetes) or to a combination of β-cell failure and resistance of target tissues to insulin action (in Type 2 diabetes).

Since the early 1920's diabetes mellitus has been treated with insulin purified from porcine or bovine pancreas. Unfortunately, the incidence of diabetes mellitus is increasing rapidly and it is estimated that the number of persons with diabetes will double, from ≈150 millions to ≈300 millions in the next 25 years. However, only a minor fraction (approximately 10%) of these will have Type 1 diabetes (insulin-dependent diabetes) requiring insulin for survival. The vast majority will have Type 2 diabetes (non-insulin-dependent diabetes) and are only relatively insulin-deficient and therefore not dependent on insulin for survival. However, frequently these patients will nevertheless progress to a state that requires insulin treatment for metabolic control. As a result, the medical need for insulin is increasing. Additionally, novel routes for administration of insulin, such as the intrapulmonary route where bio-availability is significantly lower than with subcutaneous administration, further emphasizes the need for efficient production. More than 7000 kilograms of insulin per year are currently used for the treatment of diabetes mellitus patients. 100–150 mg of insulin can be purified from one kilogram of porcine pancreas (Brange *et al.*, 1987) and accordingly the amount of insulin needed to satisfy current and future requirements will be difficult to provide by pancreatic extraction. In the 1970's the developments in molecular biology and biotechnology provided new opportunities for insulin production and since the 1980's genetically engineered human insulin for treatment of diabetes mellitus has been produced in *S. cerevisiae* and *Escherichia coli*. From an economical perspective insulin production is also an interesting challenge within biotechnology and the US market alone represents a value of US$1 billion.

Many of the biochemical and structural properties of the insulin molecule have evolved in response to the requirements of biosynthesis, processing, transport and storage in the β-cells of Langerhans islets (Blundell *et al.*, 1972; Dodson and Steiner, 1998). However, these properties are incompatible with efficient expression/secretion of insulin in *S. cerevisiae*. In this review we describe the adaptation of the insulin molecule to efficient biosynthesis, processing, transport and secretion in yeast as well as subsequent *in vitro* conversion to human insulin.

Secretory expression of insulin in *S. cerevisiae*

S. cerevisiae or 'Baker's yeast' has a highly specialized metabolism for alcoholic fermentation, converting mono- and some disaccharides into carbon dioxide and alcohol by means of enzymes and this has been extensively used for preparation of wine and beer. *S. cerevisiae* has also been developed for efficient secretory eukaryotic expression (for a review see e.g. Glick and Pasternak, 1998). Hitzeman *et al.* (1983) successfully expressed human interferon in *S. cerevisiae* and in 1982 expression of hepatitis B virus surface antigen particles in yeast formed the basis for the first genetically engineered vaccine (Valenzuela *et al.*, 1982). The basic organization of the *S. cerevisiae* secretory pathway is similar to that of higher eukaryotes, i.e. endoplasmic reticulum → Golgi complex → vesicles → plasma membrane or vacuole, making possible the expression and secretion of certain mammalian proteins (Schekman and Norvick, 1982). The secretion process is multifunctional, consisting

of translocation across the endoplasmic reticulum membrane, attachment of N-linked carbohydrate chains and folding in the lumen, transport from the endoplasmic reticulum to the Golgi apparatus and further post-translational modification in the Golgi apparatus, transport by secretory vesicles to the cell membrane and finally exit to the extracellular space.

Haploid *S. cerevisiae* cells of mating type α secrete a 13-residue peptide pheromone (α-factor) essential for mating with cells of mating type **a** (Thorner, 1981). The α-factor is the product of the *MFα1* gene, which encodes a 165-residue polypeptide (prepro-α-factor) featuring a pre-pro-peptide (leader) followed by four repeats of the α-factor (Thorner, 1981; Kurjan and Herskowitz, 1982). The *S. cerevisiae* α-factor leader has become the standard leader for secretory expression in yeast, and numerous heterologous proteins, including insulin, have been expressed as a fusion protein associated with this 85 amino acid sequence (Thim *et al.*, 1986; Zsebo *et al.*, 1986; Markussen *et al.*, 1987; Botstein and Fink, 1988; Brake, 1989; Romanos *et al.*, 1992).

Secretion of the α-factor requires processing of the prepro-α-factor by proteolytic enzymes (Julius *et al.*, 1984a; Achstetter and Wolf, 1985; Fuller *et al.*, 1988; Brake, 1989). The proteolytic enzymes involved in processing of the prepro-α-factor are also applied for processing of heterologous proteins expressed in *S. cerevisiae* fused to the α-factor leader. The signal peptidase removes the pre-peptide (signal peptide) and the Kex2 endoprotease cleaves the glycosylated α-factor pro-peptide on the C-terminal side of the dibasic sequence KR. In the prepro-α-factor, each 13-residue α-factor is preceded by a spacer peptide of 4–6 amino acid residues $((E/D)A)_{2-3}$ (Thorner, 1981; Kurjan and Herskowitz, 1982), which is removed by the dipeptidyl aminopeptidase A encoded by the *STE13* gene (Julius *et al.*, 1983; Fuller *et al.*, 1988). However, the efficiency of the dipeptidyl aminopeptidase is a limiting step in maturation of several heterologous proteins expressed in yeast (Bitter *et al.*, 1984; Brake *et al.*, 1984; Singh *et al.*, 1984; Thim *et al.*, 1986; Piggott *et al.*, 1987; Brake, 1989). Furthermore, expression of the *MFα1* gene on a high-copy plasmid resulted in secretion of incompletely processed α-factor similar to that produced by *ste13 S. cerevisiae* mutants (Julius *et al.*, 1983). Consequently, the spacer peptide between the leader sequence and the heterologous protein has typically been deleted and the heterologous protein fused directly to the leader.

EXPRESSION OF INSULIN IN THE YEAST *S. CEREVISIAE*

Yeast cells lack the features characterizing almost all animal β-cells, e.g. abundances of rough endoplasmic reticulum, a high zinc concentration and storage vesicles for accumulation of insulin in a quasi-crystalline form, which allows proinsulin synthesis and processing as well as storage and regulated secretion of insulin. Thus, an adaptation of the insulin molecule to efficient synthesis and secretion in *S. cerevisiae* cells was a necessity. Proinsulin, fused to the α-factor leader, is not readily expressed and secreted in *S. cerevisiae* (Thim *et al.*, 1986). However, the insulin molecule can be adapted for successful expression in *S. cerevisiae* as a modified single-chain proinsulin-like molecule. This insulin precursor was characterized by deletion of Thr^{B30} and replacement of human proinsulin's C-peptide with either a mini C-peptide having a C-terminal Lys (e.g. AAK) connecting Lys^{B29} to Gly^{A1} or direct connection of these amino acids and fusion to the α-factor leader (*Figure 5.1, Table 5.1*)

Table 5.1. Small-scale batch-fermentation yield of *S. cerevisiae* transformants expressing the insulin precursor[a] fused to different leaders[b]

S. cerevisiae transformants[c]	Pro-peptide[d]	Pro-peptide amino acid sequence[e]	Spacer peptide	Yield (mg/l)
YAK944	α*	APV**NTT**TEDETAQAEAVIGYSDLEGDFDVAVLPFS**NST**NNGLLFI**NTT**IASIAAKEEGVSLMAKR	–	17.5 ± 0.5
YAK708	α*	APV**NTT**TEDETAQAEAVIGYSDLEGDFDVAVLPFS**NST**NNGLLFI**NTT**IASIAAKEEGVSMAKR	EEAEAEAEPK	41.1 ± 0.9
YAK492	X1	QPIDE**NDT**SSMAKR	–	1.3 ± 0.4
YAK721	LA19	QPIDDTES**NTT**SVNLMADDTESRFAT**NTT**LALDVVNLISMAKR	EEAEAEAEPK	50.3 ± 3.8
YAK744	LA23	QPIDDTES**QTT**SVNLMADDTESRFAT**QTT**LALDVVNLISMAKR	EEAEAEAEPK	46.2 ± 2.8
YAK775	LA34	QPIDDTES**QTT**SVNLMADDTESRFAT**QTT**LALDVVNLISMAΔΔ	EEAEAEAEPK	25.6[f] ± 3.9
YAK761	TA32	QPIDDTES**NTT**SVNLMADDTESRFAT**NTT**LALDVVNLISMAΔΔ	EEAEAEAEPK	35.9[f] ± 3.9
YAK817	TA39	QPIDDTES**NTT**SVNLMADDTESRFAT**NTT**LAGGLDVVNLISMAKR	EEGEPK	59.5 ± 6.7
YAK855	TA57	QPIDDTES**QTT**SVNLMADDTESAFAT**QT**NSGGLDVVGLISMAKR	EEGEPK	79.2 ± 3.7
YAK1097	TA107	QPIDDTES**QTT**SVNLMADDTESAFAT**QT**NSGGLDVVGLISMAΔΔ	EEGEPK	48.4[f] ± 4.2

[a] The insulin precursor comprising the first 29 amino acids of the human insulin B-chain joined to the 21 amino acids of the human insulin A-chain by the mini C-peptide, AAK.
[b] Each leader consists of a pre-peptide (the α-factor leader signal peptide or the Yap3 endoprotease signal peptide) and a pro-peptide; only the pro-peptide and spacer peptides are shown here.
[c] The insulin precursor was expressed in the *S. cerevisiae* strain MT663 (*MATa/MATα pep4-3/pep4-3 HIS4/his4 tpi1::LEU2/tpi1::LEU2* Cir[+]) using the POT expression plasmid (*Figure 5.2*).
[d] α* indicates an α-factor leader in which the C-terminus has been modified from 'SLDKR' to 'SMAKR'.
[e] Where leaders lack the Kex2 endoprotease site, fermentation yield was determined as des(B30) insulin after maturation by *A. lyticus* lysyl specific protease.
Genetically engineered pro-peptide amino acid sequences that are identical to the LA19 pro-peptide sequence are underlined. Consensus N-linked glycosylation sites (NX(T/S)) as well as mutated non-functional glycosylation sites (QTT) are shown in bold.
[f] Fermentation was at 30°C for 72 h in 5 ml YPD medium. Yield was determined by RP-HPLC of the culture supernatant and is expressed as mean ± standard deviation (SD) of values obtained in 4 individual fermentations.

(Markussen *et al.*, 1987). In addition, Jonassen *et al.* (1994) developed a similar *S. cerevisiae* secretory expression system for insulin based on the insulin B-chain residues 1–29 linked to the A-chain by short connecting peptides characterized by having a dibasic processing site prior to Gly^{A1}, e.g. EKR. Single-chain insulin precursors, lacking Thr^{B30}, can be converted into human insulin by transpeptidation by a suitable enzyme, e.g. trypsin (EC 3.4.21.4) (see section describing transpeptidation). Furthermore, expression of a full-length B-chain connected to the A-chain by various mini C-peptides (e.g. RRLQKR) led to secretion of a mixture of these insulin precursor and insulin-related material to the culture supernatant (Thim *et al.*, 1986, 1987, 1989). This type of insulin precursor can be converted into insulin by digestion with trypsin and subsequently with carboxypeptidase B to remove the basic amino acids extending the B-chain. These data indicate that the *S. cerevisiae* secretory pathway does not have the ability to express human proinsulin (efficient expression being impaired by the C-peptide). However, it does have the ability to efficiently fold and export single-chain proinsulin-like precursors either lacking or featuring mini C-peptides (see section describing the structure of the insulin precursor).

The synthetic gene encoding the α-factor-leader–insulin precursor fusion protein was inserted into the *POT* plasmid (*Figure 5.2*), which is an *S. cerevisiae–E. coli* shuttle plasmid based on the *S. cerevisiae* 2 μm plasmid and the pBR322 plasmid (Markussen *et al.*, 1987; MacKay *et al.*, 1990; Kawasaki and Bell, 1999). Selection of the *POT* expression plasmid in *S. cerevisiae* was based on *Schizosaccharomyces pombe*'s triose phosphate isomerase (TPI), encoded by the *POT* gene, a glycolytic enzyme that catalyzes the interconversion of glyceraldehyde-3-phosphate and dihydroxyacetone-3-phosphate. *S. cerevisiae* strains (like MT663 (*MATa/MATα pep4-3/pep4-3 HIS4/his4 tpi1::LEU2/tpi1::LEU2* cir^+)) carrying a deletion in the *TPI1* gene (encoding the corresponding enzyme) grow poorly on glucose as the carbon source (MacKay *et al.*, 1990; Kawasaki and Bell, 1999). Transformation of Δ*tpi1 S. cerevisiae* strains with the *POT* expression plasmid allows autoselection by the ability to grow on glucose (MacKay *et al.*, 1990; Kawasaki and Bell, 1999). Thus, the *TPI1* selection system provides the opportunity for the utilization of complex medium for production and a plasmid loss rate <1% after 30 generations (Kawasaki and Bell, 1999) compared to a generally reduced stability of most 2 μm-based expression plasmids (Armstrong *et al.*, 1989). Plasmid stability is essential for the ability to perform to large-scale fermentation. To limit recombination between plasmids and improve long-term stability, certain regions of the *POT* expression plasmid were modified, e.g. the 2 μm plasmid *FLP* gene and the inverted repeat sequences. The flanking regions of the *TPI1* and *POT* genes show little homology and the *S. pombe POT* gene was weakly expressed in *S. cerevisiae* (Russell, 1985). Consequently, multiple copies of the *POT* plasmid were required to generate sufficient gene product to allow growth on glucose (MacKay *et al.*, 1990; Kawasaki and Bell, 1999). The copy number of the *POT* plasmid without an expression cassette has been estimated to be approximately 50–60 per cell and with an expression cassette encoding the insulin precursor estimated to be approximately 20 per cell (Egel-Mitani *et al.*, 1988; MacKay *et al.*, 1990).

Transcription of the synthetic fusion protein gene was initiated by the strong constitutive *S. cerevisiae TPI1* gene promoter and terminated by the *TPI1* terminator (*Figure 5.2*) (Alber and Kawasaki, 1982; Markussen *et al.*, 1987; MacKay *et al.*, 1990; Kawasaki and Bell, 1999). However, the *TPI1* gene promoter might not be

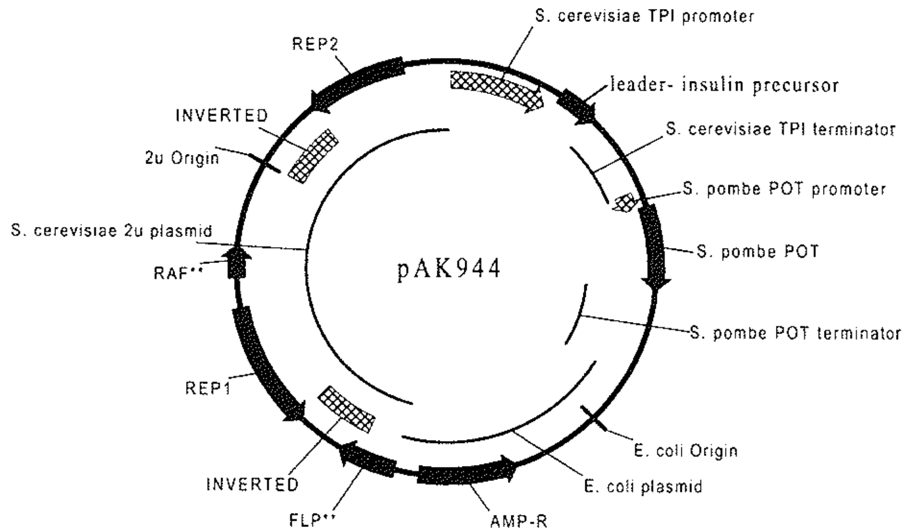

Figure 5.2. *S. cerevisiae* expression plasmid pAK944 used for expression of the α-factor leader–insulin precursor fusion protein. The pAK944 expression plasmid was based on the *S. cerevisiae–E. coli* shuttle *POT* plasmid. 'leader-insulin precursor' indicates the fusion protein gene expression cassette; 'S. cerevisiae TPI promoter' indicates the *S. cerevisiae TPI1* promoter; 'S. cerevisiae TPI terminator' indicates the *S. cerevisiae TPI1* terminator; 'S. pombe POT promoter', 'S. pombe POT' and 'S. pombe POT terminator' indicates the *S. pombe POT* gene used for selection in *S. cerevisiae*; 'E. coli plasmid' indicates DNA derived from the pBR322 plasmid; 'E. coli Origin' indicates an *E. coli* origin of replication; 'AMP-R' indicates the β-lactamase gene; 'S. cerevisiae 2u plasmid' indicates DNA derived from the 2 μm plasmid (various 2 μm plasmid coding regions as well as inverted repeats and the origin of replication are also indicated).

strictly constitutive and may exhibit some regulation in response to the glucose concentration, which may result in a sub-optimal gene expression pattern during fermentation. Furthermore, *S. cerevisiae* has a Crabtree effect and surplus glucose can result in extensive ethanol formation at the expense of heterologous protein production. These aspects of the *S. cerevisiae-POT* expression system represent a liability that needs to be carefully controlled during large-scale fermentations by ethanol monitoring and glucose dosing. Nevertheless, industrial application of the *S. cerevisiae-POT* expression system has shown it to be remarkably stable and easy to scale-up. For more than 15 years the *S. cerevisiae-POT* expression system has been used very successfully for long-term (500 h), large-scale continuous fermentations with high production levels of the insulin precursor (Diers *et al.*, 1991) and is producing a substantial part of the insulin required for treatment of diabetes mellitus.

Optimization of secretory expression of insulin in *S. cerevisiae*

As mentioned above, removal of the α-factor spacer peptide $((E/D)A)_{2-3}$ from heterologous proteins (including insulin) expressed in *S. cerevisiae* as a fusion protein can be incomplete (Bitter *et al.*, 1984; Brake *et al.*, 1984; Singh *et al.*,1984; Thim *et al.*, 1986; Piggott *et al.*, 1987; Brake, 1989). Consequently, spacer peptides were generally avoided and heterologous proteins, including insulin, were typically fused directly to the leader (Markussen *et al.*, 1987; Brake, 1989). However, this might

result in secretion of unprocessed fusion protein, indicating that the spacer peptide may facilitate Kex2 endoprotease activity (Zsebo *et al.*, 1986; Piggott *et al.*, 1987; Brake, 1989). Expression of α-interferon fused to the α-factor leader with a $(EA)_2$ spacer peptide was ≈90% cleaved at the KR site, whereas only ≈50% of the corresponding fusion protein lacking spacer peptide was processed by the Kex2 endoprotease (Zsebo *et al.*, 1986). Interestingly, the spacer peptide was removed from the secreted α-interferon. Piggott *et al.* (1987) expressed a hybrid interferon and reported improved processing of the fusion protein with a spacer peptide $(EA)_2$. It appeared that interferon was secreted with an N-terminal extension showing inefficiency of the dipeptidyl aminopeptidase processing. Furthermore, removal of the spacer peptide by introduction of an additional Kex2 endoprotease processing site (SLDKR↓EAEAKR↓) was also unsuccessful. Expressing an α-factor leader–insulin precursor fusion protein without a spacer peptide in *S. cerevisiae* showed secretion of both the insulin precursor and hyperglycosylated fusion protein to the culture supernatant (Kjeldsen *et al.*, 1996). Thus, the Kex2 endoprotease processing of the pro-peptide–insulin precursor fusion protein was also a rate-limiting step in secretion of the insulin precursor and optimization would improve the fermentation yield. Enzymatic processing of the pro-peptide–insulin precursor fusion protein might be improved either by increasing the quantity of Kex2 endoprotease or by optimizing the fusion protein as substrate for the Kex2 endoprotease. As mentioned above, the *S. cerevisiae* prepro-α-factor spacer peptide $((E/D)A)_{2-3}$ following the dibasic processing site can improve the Kex2 endoprotease processing. Therefore, a spacer peptide was re-introduced into the fusion protein between the pro-peptide and the insulin precursor in an attempt to optimize Kex2 endoprotease processing (*Figure 5.1*) (Kjeldsen *et al.*, 1996). However, this results in secretion of an insulin precursor molecule that is N-terminally extended. Because removal of the N-terminal extensions so far had been unsuccessful, an alternative procedure for removal was required. To allow *in vitro* enzymatic removal of the spacer peptide from the insulin precursor a Lys (K) residue was introduced in the spacer peptide C-terminus before Phe^{B1}.

DEVELOPMENT OF REMOVABLE SPACER PEPTIDES OPTIMIZING KEX2 ENDOPROTEASE PROCESSING OF THE LEADER–INSULIN PRECURSOR FUSION PROTEIN AND CONCOMITANTLY THE INSULIN PRECURSOR FERMENTATION YIELD

Introduction of a spacer peptide (EAEAEAK) after the dibasic Kex2 endoprotease site and before the Phe^{B1} of the insulin precursor significantly increased the fermentation yield of the insulin precursor expressed using the *S. cerevisiae-POT* expression system (Kjeldsen *et al.*, 1996). Concomitantly, the quantity of secreted hyperglycosylated pro-peptide–insulin precursor fusion protein decreased. As expected, incomplete removal of the EA dipeptides from the spacer peptide/N-terminal extension by the dipeptidyl aminopeptidase A resulted in a heterogeneous population of insulin precursors in the culture supernatant. To prevent *in vivo* modification of the spacer peptide, it was modified by introduction of additional amino acid residues. Addition of a Glu residue to the spacer peptide N-terminus (to give EEAEAEAK) prevented processing by the dipeptidyl aminopeptidase A. Surprisingly, this spacer peptide was also removed from part of the secreted insulin precursor, implying that *S. cerevisiae* had a protease which removed the spacer

peptide by cleavage at the Lys residue before Phe^{B1} of the insulin precursor (Kjeldsen *et al.*, 1996). Cleavages were not occurring at other basic amino acid residues in the insulin precursor (Arg^{B22}, Lys^{B29} or the Lys in the mini C-peptide). Characterization of this proteolytic activity indicated that the Yap3 endoprotease (Yapsin1) was responsible for removing the spacer peptide from the insulin precursor. However, *in vitro* removal of this spacer peptide by the Yap3 endoprotease was incomplete. Consequently, the spacer peptide was further modified by insertion of different amino acid residues before the Lys residue to prevent proteolytic cleavage by the Yap3 endoprotease. For instance, removal of the spacer peptides EEAEAEAEPK or EEAEAEAPK from the insulin precursor by the Yap3 endoprotease was prevented (*Table 5.1*) (Kjeldsen *et al.*, 1996). Different modifications of the spacer peptide were subsequently generated (e.g. EEGEPK) which also increased the insulin precursor fermentation yield (*Table 5.1*). A spacer/N-terminal extension developed to be resistant toward proteolytical processing *in vivo* but for efficient enzymatic removal *in vitro*, solves sub-optimal yield due to partial processing by the Kex2 endoprotease and/or the dipeptidyl aminopeptidase A. Transpeptidation of the N-terminally extended insulin precursor, either by trypsin (EC 3.4.21.4) or by *Achromobacter lyticus* lysine specific protease (EC 3.4.21.50), will remove both the spacer peptide and the mini C-peptide as well as adding a Thr residue to Lys^{B29} to generate human insulin (Markussen, 1987; Markussen *et al.*, 1987) (see section on conversion of the insulin precursor to human insulin). This concept has been extended to optimize the fermentation yield of other heterologous proteins expressed in *S. cerevisiae* and can be advantageously combined with either *in vivo* or *in vitro* removal of the spacer peptide/N-terminal extension by a suitable enzyme.

Interestingly, the fusion protein expressed in yeast has remarkable similarities with preproinsulin (*Figure 5.1*). The overall requirements for successful biosynthesis appear to be the same: a signal peptide and a C-peptide/pro-peptide. Processing is performed by the Kex2 endoprotease that is a yeast equivalent of the mammalian prohormone convertases that process proinsulin. In both *S. cerevisiae* cells and β-cells enzymatic conversion occurs in the late secretory pathway. Moreover, most proinsulin C-peptides are conserved at the B-chain/C-peptide junction featuring a variation of the sequence $(R/K)R((E/D)X)_2$, where X is alanine, valine, or leucine, which is important for maturation of proinsulin (Gross *et al.*, 1989). The spacer peptide was modelled on the prepro-α-factor spacer but shows strong similarities to proinsulin's B-chain/C-peptide junction and both are important for efficient conversion. The main difference is the localization of the pro-peptide and the C-peptide. The distinction is that yeast secretes a single-chain precursor that requires *in vitro* conversion to obtain human insulin. However, the similarities are so remarkable that one might wonder whether it is possible to adapt the insulin molecule in such a way that the yeast secretory pathway can produce mature processed human insulin.

DEVELOPMENT AND CHARACTERIZATION OF LEADERS FOR SECRETORY EXPRESSION OF THE INSULIN PRECURSOR

Expression of a signal peptide–insulin precursor molecule (lacking pro-peptide) in *S. cerevisiae* did not result in secretion of the insulin precursor to the culture supernatant. Thus, the pro-peptide was necessary for secretion of the insulin precursor. Conse-

quently, development of leaders and especially pro-peptides was a key target for optimizing yeast secretory expression of the insulin precursor.

A genetically engineered (constructed/synthetic) pro-peptide, X1 (*Table 5.1*), featuring a hydrophilic polypeptide sequence and a consensus site for attachment of N-linked glycosylation, combined with potentially efficient processing sites for both the signal peptidase and the Kex2 endoprotease, was designed (*Table 5.1*) (Kjeldsen *et al.*, 1997). The Yap3 endoprotease signal peptide (MKLKTVRSAVLSSLFASQVLG), which has been shown to provide efficient secretion of heterologous proteins in yeast in combination with a pro-peptide (Christiansen and Petersen, 1994), was chosen as pre-peptide for the X1 leader. The X1 leader facilitated secretion of the insulin precursor expressed in *S. cerevisiae*, although not very efficiently (*Table 5.1*). Subsequently, the X1 pro-peptide was optimized and extended by insertion of additional amino acid residues by semi-random mutation (Kjeldsen *et al.*, 1997, 1998b). Semi-random mutations were performed by construction of DNA fragments encoding the pro-peptides using doped oligonucleotides to introduce codon variations at one or more positions. The DNA fragment encoding the fusion protein was expressed using the *S. cerevisiae-POT* expression system. The quantity of insulin precursor secreted to the culture supernatant by individually transformed *S. cerevisiae* clones was determined. Subsequently, the expression plasmid was isolated from the clones secreting the highest quantity of insulin precursor and the DNA sequence encoding the pro-peptide was determined. The identified pro-peptide sequence was subjected to additional sequential optimization. This procedure generated a number of genetically engineered leaders highly efficient in facilitation secretion of the insulin precursor, and some of these are listed in *Table 5.1*. Thus, systematic adaptation of the fusion protein expressed in *S. cerevisiae*, by optimizing secretion efficiency (by genetically engineered leaders, e.g. LA19, TA39 and TA57) and by improving conversion, increased the yield of the insulin precursor up to five times (*Table 5.1*).

Characterization of the insulin precursor biosynthesis in *S. cerevisiae*

SIGNIFICANCE OF N-LINKED GLYCOSYLATION FOR THE ABILITY OF LEADERS TO CONFER SECRETORY COMPETENCE TO THE INSULIN PRECURSOR

Insulin and the insulin precursor are distinct hydrophobic molecules that readily aggregate at neutral pH. This is even more pronounced for the unfolded insulin precursor with exposed hydrophobic regions. Consequently, it can be anticipated that solubility and limiting aggregation of the insulin precursor are important for secretion efficiency. Glycosylation of a protein can aid both generation of the correct conformation and passage of the molecule through the secretory pathway (Romanos *et al.*, 1992). It was therefore investigated whether the pro-peptide hydrophilic N-linked carbohydrate chains might contribute to secretion of the insulin precursor. The *S. cerevisiae* α-factor leader has three consensus sites for attachment of N-linked carbohydrate chains. N-linked glycosylation is important, but not essential, for the ability of the α-factor leader to secrete α-factor (Julius *et al.*, 1984b; Caplan *et al.*, 1991). The importance of the α-factor pro-peptide's three N-linked carbohydrate chains for secretion of the insulin precursor was determined by mutation of the

consensus attachment sites. Mutation of all three consensus sites for N-linked glycosylation decreased the quantity of secreted insulin precursor to ≈10%. However, there was a clear difference in the relative importance of the three N-linked carbohydrate chains with regard to influence on insulin precursor secretion. The two N-linked carbohydrate chains closest to the insulin precursor were significantly more important than the third for the ability to facilitate secretion of the insulin precursor and there was no synergistic effect of the N-linked oligosaccharide chains with respect to secretion (Kjeldsen *et al.*, 1998a). The majority of the α-factor pro-peptide–insulin precursor fusion protein lacking N-linked glycosylation was retained intracellularly in an unprocessed form, presumably in an early part of the secretory pathway (Kjeldsen *et al.*, 1998b). This indicates that the α-factor pro-peptide's N-linked glycosylation assists secretion of the insulin precursor, most probably by improving the hydrophilicity and limiting aggregation, and that this contributes to conferring secretory competence to the insulin precursor. Furthermore, glycosylation has been shown to play an important role in folding and degradation of glycoproteins in mammalian cells and in the yeast *S. pombe* (for a review see e.g. Parodi, 2000). Endoplasmic reticulum-resident lectin-like chaperones like calnexin are central in this quality control mechanism. It has not yet been clearly established whether *S. cerevisiae* features a similar quality control mechanism and *S. cerevisiae's* calnexin has significant structural variations when compared with mammalian and *S. pombe* calnexin. Nevertheless, it has been suggested that N-linked carbohydrate might be involved in folding and degradation of glycoproteins in *S. cerevisiae* (Jakob *et al.*, 1998; Parodi, 2000) and the α-factor leader N-linked carbohydrate chains may mediate folding and degradation of the insulin precursor moiety of the fusion protein.

Surprisingly, elimination of the genetically engineered LA19 pro-peptides' two consensus sites for N-linked glycosylation (the LA23 leader) did not impair the ability to facilitate secretion of the insulin precursor (*Table 5.1*) (Kjeldsen *et al.*, 1998b). The secretory capacity of the genetically engineered leaders lacking N-linked carbohydrate chains is illustrated by the TA57 leader, which increased the fermentation yield of insulin precursor to 480% relative to an *S. cerevisiae* transformant expressing an α-leader–insulin precursor fusion protein (yAK944) (*Table 5.1*). Thus, these genetically engineered leaders, without N-linked glycosylation, provide an efficient alternative for secretory expression of the insulin precursor in *S. cerevisiae*. These fusion proteins were not retained intracellularly and were efficiently processed by the Kex2 endoprotease, showing that the leader amino acid sequence alone had the ability to confer secretory competence to the insulin precursor (Kjeldsen *et al.*, 1998b).

However, this raises the question as to why the genetically engineered leaders do not require N-linked glycosylation to facilitate secretion of the insulin precursor. Genetically engineered pro-peptides, e.g. TA57, are markedly hydrophilic with a high frequency of polar amino acids (36.4% versus 28.8%) and a concomitantly lower frequency of hydrophobic amino acids (29.5% versus 39.4%) compared to the α-factor pro-peptide, and the insulin precursor relative to a pro-peptide–insulin precursor fusion protein has a difference in calculated net charge at neutral pH of –7 and these properties might adequately increase the solubility of the fusion protein. Additionally, the genetically engineered pro-peptides have amino acid sequences that may interact with chaperones of the BiP/Kar2 type (Flynn *et al.*, 1991; Blond-Elguindi *et al.*, 1993). These leader properties might provide interaction of the translocating

fusion protein with the BiP/Kar2 protein chaperone, which may assist entrance into and limit aggregation in the endoplasmic reticulum. Intriguingly, the C-peptide of proinsulin is also distinctly hydrophilic and plays a central role in biosynthesis (Steiner *et al.*, 1967; Dodson and Steiner, 1998). The similar requirement of proinsulin and the insulin precursor for hydrophilic pro-peptides for biosynthesis is remarkable and suggests that the C-peptide and the pro-peptide fulfil a comparable function, albeit in different positions. It would be interesting to establish whether the C-peptide could function as a pro-peptide and mediate folding and secretion of the insulin precursor also in yeast if localized between the signal peptide and Phe^{B1}.

SECRETION KINETICS OF THE INSULIN PRECURSOR EXPRESSED IN *S. CEREVISIAE*

Secretion of the insulin precursor fused to different genetically engineered leaders (TA57, TA39) and the α-factor leader was fast and the first insulin precursor reached the extracellular space within 2 to 4 min of chase (Kjeldsen *et al.*, 1999b). These leaders facilitated secretion of the insulin precursor with comparable kinetic profiles, indicating a similar intracellular transport and processing of these fusion proteins in the *S. cerevisiae* secretory pathway. The greater part of the exported insulin precursor was secreted to the culture supernatant within the first 15 min and $t_{½}$ was 5–10 min (Kjeldsen *et al.*, 1999b). The rapid secretion of the insulin precursor indicates both a rapid formation of the three disulphide bonds and folding in the endoplasmic reticulum and that the secreted insulin precursor followed a constitutive route to the plasma membrane.

TRANSLOCATION AND STABILIZATION OF THE INSULIN PRECURSOR

Pulse–chase analysis of a signal peptide-insulin precursor molecule (lacking pro-peptide) expressed in *S. cerevisiae* showed little synthesized insulin precursor present after 2.5 min pulse labelling with [^{35}S]cysteine. Fusion of the insulin precursor to the α-factor, TA39 or TA57 leader considerably augmented the quantity of intracellular fusion protein present after a 2.5 min [^{35}S]cysteine pulse (Kjeldsen *et al.*, 1999b). The pro-peptide was rapidly and efficiently core-glycosylated, demonstrating that the fusion protein was present either in the endoplasmic reticulum or in the Golgi apparatus. These data show that the leader is essential for translocation into the endoplasmic reticulum and that the pro-peptide is necessary for this process. As mentioned above, the genetically engineered pro-peptides have amino acid sequences that may mediate interactions with chaperones of the BiP/Kar2 protein type and this may facilitate translocation. Furthermore, efficient translocation of the insulin-like growth factor I expressed in *S. cerevisiae* also required a pro-peptide, emphasizing that the pro-peptide can be necessary for entrance of heterologous protein into the lumen of the endoplasmic reticulum (Chaudhuri *et al.*, 1992). A minor fraction of the pro-peptide already featured extended carbohydrate chains after 2.5 min and enzymatic processing of the fusion protein by the Kex2 endoprotease localized in a late Golgi compartment (Cunningham and Wickner, 1989; Wilcox *et al.*, 1992; Brickner and Fuller, 1997) had already occurred to a minor degree. These data indicate that after 2.5 min the

majority of the synthesized fusion protein was localized in the endoplasmic reticulum or in the Golgi apparatus and that a minor fraction of the fusion protein already had reached a late Golgi apparatus compartment.

INTRACELLULAR RETENTION OF THE INSULIN PRECURSOR EXPRESSED IN *S. CEREVISIAE*

After metabolic labelling of the TA39–insulin precursor and the TA57–insulin precursor fusion proteins for 2.5 min followed by 35 min of chase, approximately 30% of the metabolically labelled insulin precursor was still present as processed intracellular insulin precursor (Kjeldsen *et al.*, 1999b). Intracellular retention of a substantial quantity of the synthesized insulin precursor indicated that the insulin precursor followed two different intracellular routes in the late secretory pathway. Secretion to the culture supernatant may reflect either saturation of a sorting mechanism due to over-expression or secretion occurring in competition with intracellular retention. The Kex2 endoprotease is localized in a late Golgi compartment, showing that the intracellular retained insulin precursor was located in a Golgi or post-Golgi compartment. Proteins can be routed to the vacuole from the Golgi apparatus and apparently the insulin precursor retained intracellularly had been sorted to the vacuole (*Figure 5.3*) (the yeast strain used for expression was *pep4*Δ, impairing vacuole protein degradation). Routing of the insulin precursor to the vacuole may reflect either interactions with various sorting proteins or that it features a sorting signal. Furthermore, routing of misfolded protein to the vacuole in *S. cerevisiae* constitutes a quality control system, whereas the endoplasmic reticulum quality control system appears less stringent than in mammalian cells (Ellgaard *et al.*, 1999). Intracellular retention may reflect sorting of an aberrant insulin precursor by a quality control mechanism associated with routing to the vacuole. It is unclear if the insulin precursor expressed in *S. cerevisiae* might self-associate in the secretory pathway. Formation of dimers would actually increase the solubility, but nevertheless intracellular retention could reflect the ability of the insulin precursor to self-assemble. This property might be concealed in the fusion protein by the pro-peptide (e.g. stabilizing the monomeric form of the molecule) since only the insulin precursor was found to be retained intracellularly.

Intracellular retention of the insulin precursor (fused to a genetically engineered leader with consensus sites for N-linked glycosylation) was further analysed by prevention of folding either by mutation (Pro^{B16}) or by prevention of disulphide bond formation (using dithiothreitol). This resulted in lack of secretion, intracellular retention possibly in the vacuole, degradation and slow maturation of the pro-peptide–insulin precursor fusion protein (Kjeldsen *et al.*, 1999b). However, the pro-peptide was N-linked glycosylated. This intracellular retention of misfolded insulin precursor resembled routing of misfolded protein to the vacuole (Hong *et al.*, 1996; Ellgaard *et al.*, 1999). Routing of misfolded insulin precursor to the vacuole may be a consequence of exposure of the molecule's hydrophobic core, providing a signal for a quality control system and subsequent intracellular retention. Moreover, prevention of folding of the insulin precursor fused to a genetically engineered leader without glycosylation also resulted in prevention of secretion and degradation, but maturation of this fusion protein was impaired, indicating

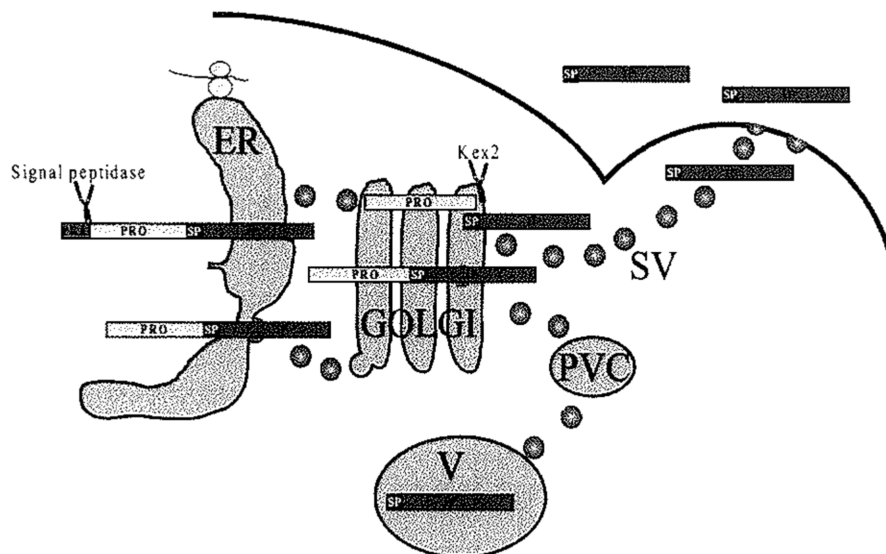

Figure 5.3. Schematic presentation of the secretory pathway, showing transport, processing and export of insulin expressed in *S. cerevisiae*. The insulin precursor was fused to a spacer peptide (SP) and a genetically engineered leader lacking N-linked glycosylation. Processing of the fusion protein by the signal peptidase in the endoplasmic reticulum (ER) and transported to and subsequently processed by the Kex2 endoprotease in the Golgi apparatus (GOLGI), are shown. Vacuole (V), pre-vacuolar compartment (PVC), secretory vesicles (SV) are indicated. Also the Yap3 endoprotease signal peptide (pre), pro-peptide (pro) and insulin precursor (IP) are indicated. After maturation of the fusion protein by the Kex2 endoprotease, the spacer peptide forms an N-terminal extension on the insulin precursor. After purification of the N-terminally extended insulin precursor, the N-terminal extension is removed, together with the mini C-peptide, during enzymatic conversion of the insulin precursor to human insulin.

retention in an earlier part of the secretory pathway, possibly the endoplasmic reticulum.

The well-characterized biosynthesis, structure and biochemical properties of insulin make it attractive for analysing yeast secretory expression, exemplifying production of heterologous proteins of scientific and commercial interest. The efficient translocation, folding, intracellular transport, processing and secretion of a modified insulin molecule as a fusion protein, emphasize the importance of molecular adaptation for biosynthesis in yeast. Fusion of the insulin precursor to a leader was a necessary adaptation for secretion efficiency. The leader facilitates translocation into the endoplasmic reticulum lumen and the pro-peptide may provide interaction with chaperones and have properties that enhance solubility of the insulin precursor. It is suggested that these pro-peptide properties assist entrance into the endoplasmic reticulum and facilitate folding of the insulin precursor moiety by limiting aggregation.

Secretory expression of the insulin precursor fused to genetically engineered leaders lacking both N-linked glycosylation and a dibasic site for Kex2 endoprotease processing

Secretory expression of heterologous protein in yeast is generally performed as a fusion protein and relies on *in vivo* intracellular maturation by the Kex2 endoprotease. Consequently, the Kex2 endoprotease catalytic efficiency toward the fusion protein can be a limiting factor in yeast secretory expression. As described in the previous sections, this problem can be solved by a spacer peptide that is resistant toward *in vivo* processing but readily removed *in vitro*. This problem might also be circumvented by expression of a fusion protein not dependent on Kex2 endoprotease for maturation. *In vitro* maturation of secreted purified fusion protein by an alternative enzyme would eliminate Kex2 endoprotease limitation on the fermentation yield and would also prevent potential N-terminal processing of heterologous proteins by dipeptidyl aminopeptidase A. Hyperglycosylation or extensive O-linked glycosylation of pro-peptides may hamper purification of secreted fusion protein. This would be a problem especially in large-scale purification of proteins for pharmaceutical applications. Genetically engineered leaders lacking both the dibasic Kex2 endoprotease processing site and consensus sites for attachment of N-linked carbohydrate chains efficiently facilitated secretion of an unprocessed pro-peptide–insulin precursor fusion protein (LA34, TA32 and TA107) (*Table 5.1*). The fusion protein was readily purified and efficiently processed into des(B30) insulin by *A. lyticus* lysine specific protease (Kjeldsen *et al.*, 1998b). Thus, these genetically engineered leaders were shown to provide an alternative expression system for the insulin precursor. This concept may have general application in yeast secretory expression, providing an opportunity for convenient purification of secreted fusion protein and subsequent *in vitro* maturation with a suitable enzyme. Moreover, secretory expression in yeast of heterologous proteins with internal dibasic sites may lead to unwanted Kex2 endoprotease processing and concomitant decrease in fermentation yield. This type of protein may advantageously be expressed in a Kex2 endoprotease negative *S. cerevisiae* strain if fused to a leader lacking N-linked glycosylation and using an alternative-processing site. Furthermore, secretion of heterologous proteins associated with the pro-peptide could be advantageous by enhancing the solubility and stability until purification and maturation of the fusion protein.

Characterization of the insulin precursor expressed in *S. cerevisiae*

The secreted insulin precursor has been isolated from the culture supernatant by RP-HPLC and characterized regarding post-translational modification features. The molecular weights of the different purified insulin precursors were subsequently determined by mass spectrometry (MALDI–MS or ES–MS with on-line desalting) before further characterization. RP-HPLC fractions containing the insulin precursor were dried in vacuum and subsequently reconstituted in a buffer suitable for proto-typical digestion. Digestion of the insulin precursor was performed by *Staphylococus aureus* V8 protease for 6 h at 37°C. The reaction was terminated by the addition of acetic acid. Finally, the resulting peptides were analysed by on-line RP-HPLC/MS analysis. Digestion of an insulin precursor with correct disulphide bonds by *S. aureus* V8 protease will result in three major and one or two minor peptides, depending on the specific insulin precursor, see *Table 5.2*. The three disulphide bonds in the insulin

Table 5.2. Characterization of secreted insulin precursors by *S. aureus* V8 peptide mapping and on-line RP-HPLC–MS analysis (Kjeldsen *et al.*, 1998b)

	Fragment no.	Theoretical mass	Experimental observed mass
Residues in correctly folded insulin precursor yAK708: E(EA)$_3$EPK-B1-B29-AAK-A1-A21			
EA		218.21 Da	ND
EE		276.24 Da	ND
B14-B21/A18-A21	I	1377.56 Da	1377.50 Da
B22-B29-AAK-A1-A4	II	1683.97 Da	1684.07 Da
AEPK-B1-B13/A5-A17	III	3394.88 Da	3394.67 Da
Residues in correctly folded insulin precursor yAK721:E(EA)$_3$EPK-B1-B29-AAK-A1-A21			
EA		218.21 Da	ND
EE		276.24 Da	ND
B14-B21/A18-A21	I	1377.56 Da	1377.47 Da
B22-B29-AAK-A1-A4	II	1683.97 Da	1684.10 Da
AEPK-B1-B13/A5-A17	III	3394.88 Da	3395.13 Da
Residues in correctly folded insulin precursor #yAK855: E(EA)3EPK-B1-B29-AAK-A1-A21			
EE		276.24 Da	ND
B14-B21/A18-A21	I	1377.56 Da	1377.50 Da
B22-B29-AAK-A1-A4	II	1683.97 Da	1684.07 Da
GEPK-B1-B13/A5-A17	III	3380.85 Da	3381.31/3669.31[a] Da

[a] Partial digestion of fragment III with additional N-terminally EE (theoretical mass 3639.08 Da), ND not determined.

precursor molecule are located in two of the major peptides (fragments I and III, see *Table 5.2*). Fragment III contains two of the three disulphide bonds, the inter-chain disulphide bond Cys^{A7} to Cys^{B7} and the intra-chain disulphide bond, Cys^{A6} to Cys^{A10}. The correct position of these disulphide bonds cannot be assigned using this method. However, comparison by RP-HPLC of the *S. aureus* V8 peptide map derived from the insulin precursors with the corresponding peptide map from human insulin showed similar retention times of these fragments, indicating identical disulphide bonds.

Insulin precursors isolated from the yAK708, yAK721 and yAK855 *S. cerevisiae* transformants (see *Table 5.1*) were characterized using the protocol outlined above. On-line RP-HPLC analysis showed only secreted insulin precursor with retention times identical to that of correctly folded insulin precursor, indicating that *S. cerevisiae's* secretory pathway has comprehensive quality control of the insulin precursor before export to the extracellular space. Post-translational modification features of the insulin precursor appeared to be independent of the leader used to facilitate secretion. Comparison of the insulin precursor accumulated after 72 h fermentation with insulin precursor after a 2.5 min metabolic pulse by RP-HPLC showed identical retention time, indicating that a similar species of insulin precursor was secreted (*Figure 5.4*). SDS-PAGE analysis (insert C in *Figure 5.4*) showed that the RP-HPLC fraction with the insulin precursor contained a labelled protein that migrated with a molecular weight corresponding to 6000 Da (lane 4), equivalent to the insulin precursor. Moreover, a more hydrophobic molecule with low intensity radioactivity and a molecular weight similar to the insulin precursor was present in the gradient fraction (lane 11). This may either be a more hydrophobic isomer of the insulin precursor or an unrelated molecule.

Insulin structure

The processing, transport and storage of insulin are based on the structural, biochemical and self-associating properties of the molecule and these are selectively adapted for biosynthesis in the β-cells. The structural basis in this system has recently been reviewed (Dodson and Steiner, 1998). The insulin molecule has structural properties that can allow heterologous expression in yeast cells, which have radically different properties compared to the highly specialized β-cell. Here we describe the structural properties of the insulin molecule and the insulin precursor adapted for secretory expression in yeast.

DESCRIPTION OF INSULIN STRUCTURE

The structure of human insulin and mutants thereof are well described under a number of different conditions, both in the crystal state (Adams *et al.*, 1969; Baker *et al.*, 1988) and in aqueous solution (Ludvigsen *et al.*, 1994; Olsen *et al.*, 1996). The A-chain (21 amino acids) of human insulin is covalently linked to the B-chain (30 amino acids) by two disulphide bonds from Cys^{A7} to Cys^{B7} and from Cys^{A20} to Cys^{B19} (*Figure 5.5a*). An additional, intra-chain disulphide bond from Cys^{A6} to Cys^{A11} assists the A-chain to maintain its conformation, consisting of two almost anti-parallel helices A1–A8 and A12–A20. The B-chain can exist in one of two main conformations designated 'T' and 'R' (Kaarsholm *et al.*, 1989). In the R-state, insulin is characterized by a long

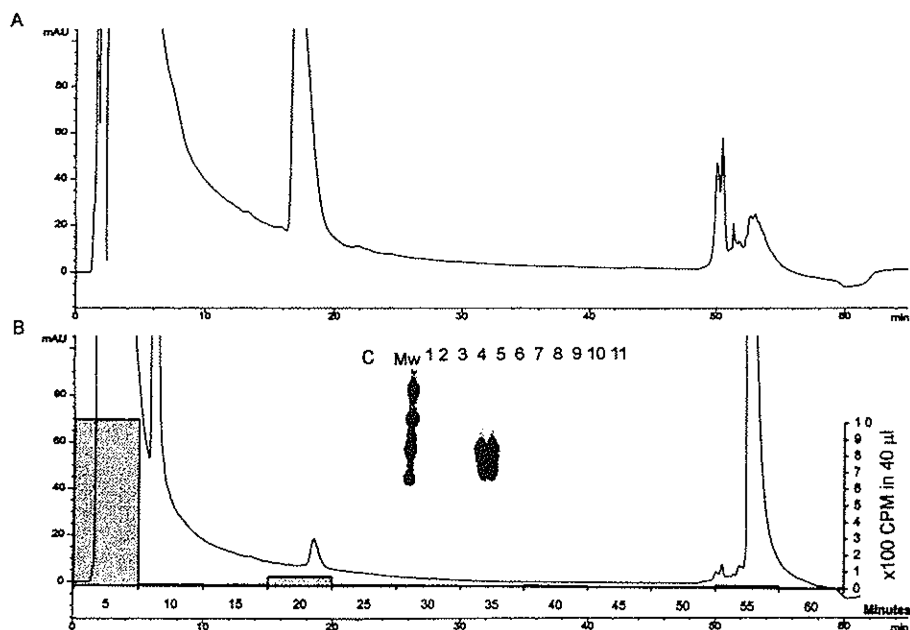

Figure 5.4. RP-HPLC chromatograms of cell-free culture supernatants from the *S. cerevisiae* transformant yAK817 either after 72 h fermentation (A) or after a short metabolical labelling (B). A: RP-HPLC chromatogram of culture supernatant after 72 h fermentation at 30°C in YPD medium. The N-terminally extended insulin precursor has a retention time of approximately 17 min. B: RP-HPLC chromatogram of culture supernatant after 2.5 min metabolic labelling with [^{35}S]cystine. Fractions were collected at 5 min intervals and the radioactive content in 40 µl was determined. Radioactivity, presumably non-incorporated ^{35}S-cystine, was also found in a fraction typically containing small molecules. C: The remaining of the RP-HPLC fraction was lyophilized and re-suspended in SDS-PAGE loading buffer and subjected to SDS-PAGE. The gel was fixed, dried and exposed to a Molecular Dynamics phosphor screen. MW, ^{14}C rainbow molecular marker: 21.5, 14.3, 6.5 and 3.4 kDa. Reprinted from *Journal of Biotechnology*, Vol 75, Kjeldsen, T. *et al.* The role of leaders in intracellular transport and secretion of the insulin precursor in the yeast *Saccharomyces cerevisiae*, pp 195–208, Copyright 1999, with permission from Elsevier Science.

helix, B1–B19, essentially perpendicular to the A-chain helices. In the T-state the B-chain's α-helix only extended from B9–B19. B20–B23, in both states, comprises a turn followed by an extended structure throughout the remaining B-chain, forming one half of a β-sheet structure. So far, the R-state has not been observed in monomeric insulin or mutant insulin molecules in solution, but exclusively in the self-associated hexameric state (*Figure 5.6*) under influence of zinc ions and, in particular, sufficient amounts of additives like *m*-cresol or phenol.

In solution insulin exhibits a complex pattern of aggregation and precipitation depending on protein concentration, pH, temperature, metal ions, ionic strength, and solvent composition. These properties have complicated structural studies of monomeric insulin in solution by NMR, which inherently is limited in sensitivity to the 0.5 mM range concentration for structural study purposes. For NMR it is required that the majority of molecules are in the monomeric state, and insulin's self-

association and aggregation over a wide pH range can be an impediment for NMR-based structural studies. However, this problem has been circumvented by design of insulin analogues with full activity and their structure solved by NMR in a soluble monomeric state at both pH 2.4 (Ludvigsen *et al.*, 1994) and pH 6.5 (Olsen *et al.*, 1996) in aqueous solution at each side of the aggregation zone. These structures have overall similarity to the insulin T-state observed in the crystal state. In the blood-stream insulin circulates in sub-nanomolar concentrations in a monomeric form, which binds to the insulin receptor and initiates a signal cascade that finally results in cellular glucose uptake and lowering of the blood glucose concentration. Thus, the majority of the structural studies on monomeric insulin molecules have naturally correlated the structure to the potency of the specific mutant. The dimer of insulin (*Figure 5.5a* and *Figure 5.5b*) is formed by an anti-parallel β-sheet between two monomers as well as additional contacts, especially the residues Ser^{B9}, Val^{B12}, Glu^{B13} and Tyr^{B16}, from the α-helix in the B-chain. Dimers associate to hexamers in the presence of zinc ions, using another hydrophobic patch in the insulin molecule. This patch is, in the R_6 state, composed by the residues Phe^{B1}, Val^{B2}, His^{B5}, Leu^{B6}, Ser^{B9}, His^{B10} from one monomer and the residues Cys^{A6}, Cys^{A7}, Thr^{A8}, Ser^{A9}, Ile^{A10}, Leu^{A13}, His^{B10}, Leu^{B17} and Val^{B18} from the second monomer in the dimer, to the corresponding surface of a second dimer in combination with phenol or alike molecules and water in the interface. In the T_6 hexamer the dimer interface is composed by His^{B10} residue from one monomer and Leu^{A13}, Glu^{A17}, Phe^{B1}, Val^{B2}, Gln^{B4}, Leu^{B6}, His^{B10}, Ala^{B14}, Leu^{B17} and Gly^{B20} residues from a second monomer in contact with the corresponding surface of another dimer. It is clear from *Figure 5.5a* and *5.5b* that a C-peptide as well as a leader attached at the N-terminal of the B-chain would be located on the surface of either the R_6 or the T_6 hexamer of insulin.

INSULIN PRECURSOR STRUCTURE

A substantial number of insulin and insulin mutant structures have been published but structures of proinsulin or single-chain insulin precursors are less common. Structurally, the proinsulin molecule has not thus far been well described in the literature. One study (Weiss *et al.*, 1990) compared the NMR spectrum of the two-chain insulin with that of proinsulin, indicating that the natural C-peptide (35 amino acids) did not fold with any well-defined structure and that the A-chain and the B-chain were structurally similar to what is known for human insulin in the T-state. In support of this, Oh *et al.* (1998) also reported similar structural arrangements for the A- and B-chain proteins of the proinsulin, leaving the C-peptide and the signal peptide with random structures. These studies were both performed using 20% acetic acid as co-solvent, a medium that is a weak denaturant (Pittman *et al.*, 1997; Ludvigsen *et al.*, 1998). However, these observations of the proinsulin structure have been further confirmed by Fourier Transform infrared spectroscopic studies (Xie and Tsou, 1993). The structure of single-chain insulin precursors wherein the C-peptide is either substituted with a mini C-peptide (AAK) or B29 is linked directly to A1 by a peptide bond have been determined by X-ray crystallography. The best resolved single-chain insulin molecule structure is the insulin precursor where Lys^{B29} is directly linked to Gly^{A1} by a peptide bond (Markussen *et al.*, 1985; Derwenda *et al.*, 1991). This

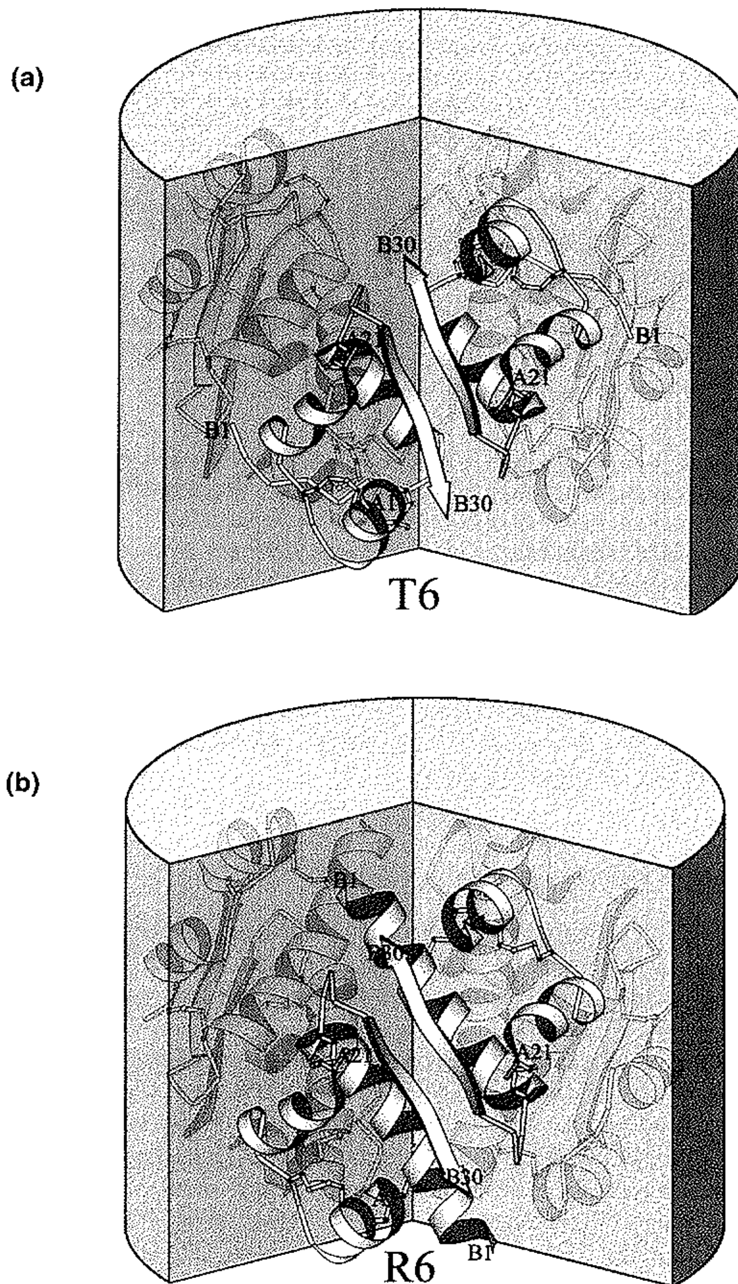

Figure 5.5. Insulin hexamer in the T-state (a) and R-state (b). Three dimers are arranged to form the hexameric unit. To illustrate the near rotational symmetry of the hexamer, the dimers have been moved a little away from the centre of the hexamer. The molecule in cartoon presentation (MOLSCRIPT, Kraulis, 1991) is composed of two helices in the A-chain peptide and one helix of the B-chain. The anti-parallel β-sheet consists of two extended strands, one from each B-chain in the dimer molecule. The C-terminal of the B-chain and the N-terminal are located on the surface, both in the R6 and in the T6 hexamer.

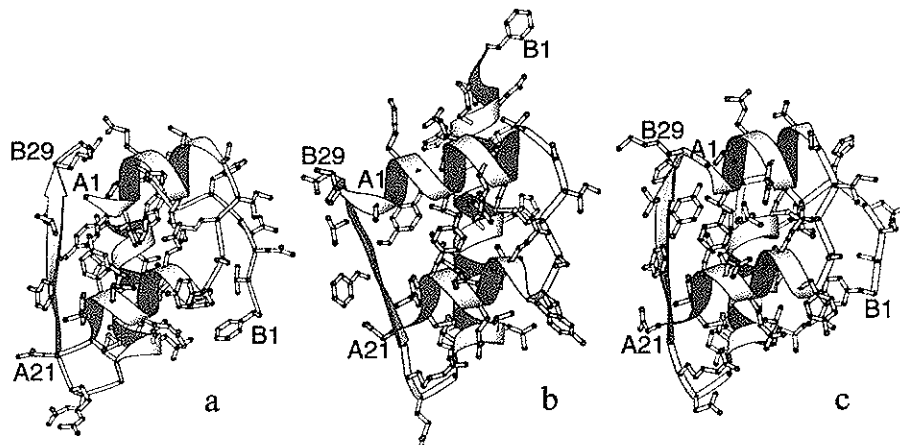

Figure 5.6. The conventional insulin structure, 2Zn insulin or insulin in the T-state (a), here represented by the His^{B16} insulin mutant as determined in solution (Ludvigsen *et al.*, 1994), is characterized besides the normal arrangement of helices by the tight packing of the C-terminal part of the B-chain to particular hydrophobic residues of the B-chain helix and Ile^{A2} and Tyr^{A19}. The crystal structure of single-chain insulin precursor featuring a peptide bond between B29 and A1 (c) and the structure of insulin in the T-state are remarkably similar. The insulin analogue precursor with an amino acid substitution in position B28 from Pro to Asp (b) has a modified structural arrangement. The insulin analogue precursor is forced into the R-state by phenol and *m*-cresol in the crystal structure (Whittingham *et al.*, 1998) and furthermore the C-terminal of the B-chain is pushed away from the core of the molecule partly by the elongation of the N-terminal A-chain helix extended by the connecting peptide. Structures were produced using MOLSCRIPT (Kraulis, 1991).

structure (*Figure 5.6c*) is very similar to both the T-state insulin crystal structure and to the high-resolution NMR structure (Ludvigsen *et al.*, 1994) (*Figure 5.6a*) in aqueous solution, except of course for the connecting peptide bond. Interestingly, the insulin precursor has a potency <0.1% of human insulin, leading to the conclusion that the insulin molecule must undergo structural changes upon receptor binding. This idea was conceived even earlier by Dodson *et al.* (1983) and substantiated by study of several modified insulin molecules (Nakagawa and Tager, 1987, 1993) and spectroscopically (Pittman *et al.*, 1997) and structurally (Ludvigsen *et al.*, 1998), as well as from correlations of the structure of insulin analogues with stability and biological activity (Olsen *et al.*, 1998). A second single-chain analogue with the amino acid substitution Pro^{B28} to Asp^{B28} (*Figure 5.6b*) and the connecting tri-peptide (AAK) has been published (Whittingham *et al.*, 1998), however the C-peptide is generally poorly defined. Comparison of the R_6 hexameric structure of the single-chain insulin analogue precursor with that of the human insulin precursor shows complex differences within the hexamer. Among the six insulin molecule structures in the hexamer, two in the third dimer form a structurally well-defined connecting peptide, emphasizing that the tri-peptide essentially extends the first helix in the A-chain with one more helix round. Generally, the extension of the N-terminal A-chain helix results in a movement of the C-terminal part of the B-chain compared to the conventional insulin structure (*Figure 5.6a*).

Relationship between the fermentation yield of insulin analogue precursors expressed in *S. cerevisiae* and the *in vitro* folding stability of the corresponding insulin analogues

To test whether there is a relationship between the mature insulin's *in vitro* folding stability and the secretion efficiency of the corresponding precursor expressed in *S. cerevisiae*, the folding stability of a number of insulin analogues was determined. Different insulin analogue precursors fused to the α-factor leader and with amino acid Lys^{B29} connected to Gly^{A1} by the mini C-peptide AAK, were expressed in yeast using the *S. cerevisiae-POT* expression system described in the previous sections. Yeast transformants expressing the different insulin analogue precursors were grown in YPD medium for 3 days at 30°C and the fermentation yield determined by RP-HPLC. Yeast cells were removed and the insulin analogue precursors were purified and subsequently converted into the mature two-chain molecule. The *in vitro* folding–unfolding transition of native insulin and insulin analogues may be evaluated by changes in the far-UV circular dichroism (CD) spectrum following the addition of increasing amounts of the protein denaturant guanidine hydrochloride (GuHCl). Upon denaturation, the negative CD in the 240–218 nm region gradually diminishes, consistent with the loss of ordered structure that accompanies protein unfolding. The GuHCl-induced reversible unfolding curve of human insulin is reasonably well described by a two-state denaturation scheme. The unfolding curve is characterized by a low degree of co-operativity and the resulting free energy of unfolding, $\Delta G_{H_2O} \sim$ 3.8 kcal/mol, is on the low side for a globular protein (Pace, 1975). The stability of the insulin fold may be enhanced by various amino acid substitutions near the N- and C-caps of α-helices, the removal of hydrophobic side-chains at the protein surface or creation of hydrogen bonding interactions (Kaarsholm *et al.*, 1993). Interestingly, there is a positive correlation between *in vitro* folding stability of insulin analogues and the fermentation yield of the corresponding insulin analogue precursor. As shown in *Figure 5.7*, the more stable insulin analogues are expressed more efficiently as the corresponding single-chain insulin precursors. This demonstrates an independent contribution of the folding of the A- and B-chain segments of the insulin molecule to the overall fermentation yield of insulin precursor. Indeed, the two insulin analogues that fall significantly outside this correlation (open symbols in *Figure 5.7*) both feature the amino acid substitution threonine to histidine in position A8. This particular modification enhances the folding stability of the insulin molecule by a C-terminal capping effect of the first α-helix of the insulin A-chain. Because this stabilizing effect is partially counteracted by the AAK mini C-peptide, these two modifications are not independent and the positive effect on folding stability of the two-chain insulin analogue is not expected to have full effect on the fermentation yield of the corresponding insulin precursor. These data show that the secretion efficiency of the insulin precursor can be related to the folding stability. However, this correlation is only valid as long as the secretory pathway does not otherwise limit secretion efficiency. It is suggested that the correlation between *in vitro* folding stability and fermentation yield reflects a selective adaptation of the plasmid copy number resulting partially from the constitutive expressed insulin analogue precursor's folding properties and from the need for complementation of the host marker (*tpi1*Δ) by the plasmid-encoded selection gene (*POT*). Thus, the folding stability of

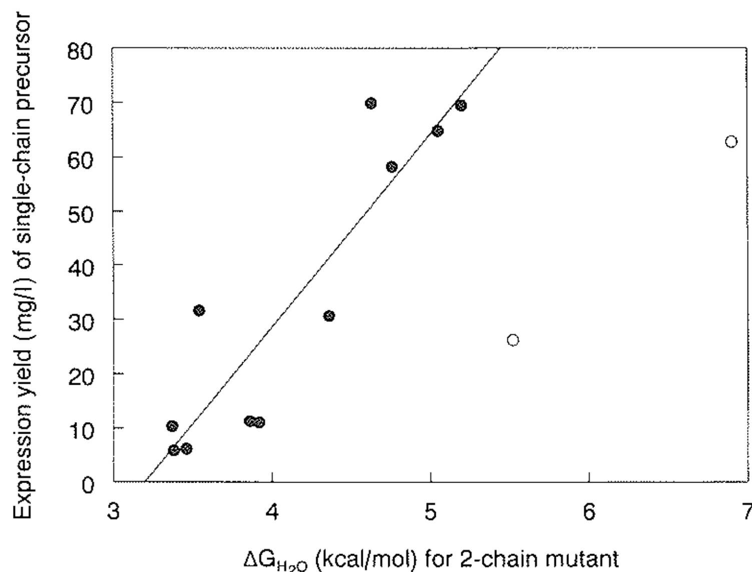

Figure 5.7. Relationship between the fermentation yield of insulin analogue precursors expressed in *S. cerevisiae* and the *in vitro* folding stability of the corresponding insulin analogues. Yeast transformants expressing different insulin analogue precursors with amino acid Lys^{B29} connected to Gly^{A1} by the mini C-peptide AAK and fused to the α-factor leader were grown in YPD medium for 3 days at 30°C and the fermentation yield determined by RP-HPLC. Yeast cells were removed and the insulin analogue precursors were purified and subsequently converted into the mature two-chain molecule. The *in vitro* folding stability of the insulin analogues was determined by evaluating the changes in the far-UV CD spectrum by increasing the concentration of the protein denaturant guanidine hydrocloride (GuHCl) (see section describing relationship between expression yield and *in vitro* folding stability).

the insulin precursor has feedback regulation on the gene copy number (i.e. plasmid copy number) by selective pressure and consequently influences its own fermentation yield.

Conversion of the insulin precursor to human insulin

The maturation of the insulin precursor molecules with mini C-peptide and lacking Thr^{B30} to human insulin can be performed by two related enzymatic procedures (reviewed by Markussen, 1987). The insulin precursor molecule can be converted to human insulin in a two-step procedure by first cleaving at the Lys residues (B29 and in the mini C-peptide) by a suitable enzyme (e.g. trypsin or *A. lyticus* protease) and subsequent isolation of the resulting desB30-insulin molecule. This step is then followed by an enzymatic catalyzed coupling of threonine ester to the terminal Lys^{B29} residue of the desB30-insulin molecule to generate human insulin. The coupling is performed at minimal water conditions, which reverse the enzymatic activity from hydrolysis toward synthesis to generated human insulin ester. The cleavage–coupling two-step conversion procedure is illustrated in *Figure 5.8*: the enzymatic cleavages of the insulin precursor (→) is followed by the coupling of the threonine ester to the desB30-insulin molecule (↓). The transpeptidation one-step conversion is also shown in *Figure 5.8* (↓). Finally, the isolated human insulin ester from both conversion procedures is transformed to human insulin by chemical hydrolysis. The kinetics

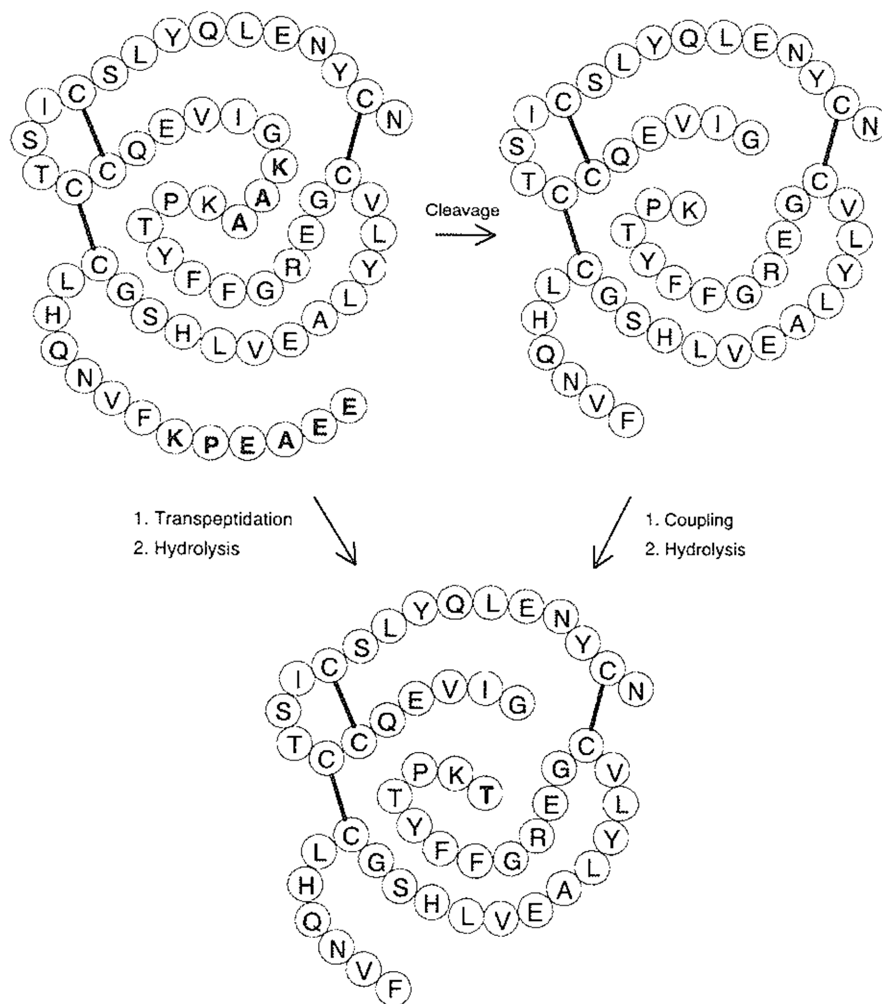

Figure 5.8. Schematic presentation of the two enzymatic procedures for conversion of the insulin precursor molecule to human insulin. The cleavage–coupling two-step method is shown: first the insulin precursor is cleaved by *A. lyticus* protease at Lys residues (→) in aqueous condition at pH 9, followed by coupling of the (right ↓) threonine ester to the B29 residue in desB30-insulin in an aqueous solution containing *N*-methyl-pyrrolidone (NMP) at pH 6.5. The one-step transpeptidation conversion of the insulin precursor to human insulin is indicated by (left ↓). The amino acid residues of the N-terminal extension and of the mini C-peptide removed from the insulin precursor molecule by the *A. lyticus* protease cleavage, as well as the threonine residue added in the coupling/transpeptidation, are shown in bold characters.

of a typical cleavage–coupling two-step conversion of the insulin precursor to human insulin is shown in *Figure 5.9* (panel A (cleavage) and panel B (coupling)). Likewise, the kinetics of a typical transpeptidation one-step conversion of the insulin precursor to human insulin is shown in *Figure 5.9* (panel C).

Reversing the enzymatic activity from hydrolysis to synthesis requires a high substrate concentration as well as a high threonine ester to substrate ratio. The water content of the conversion reaction mixture is lowered by addition of an aprotic

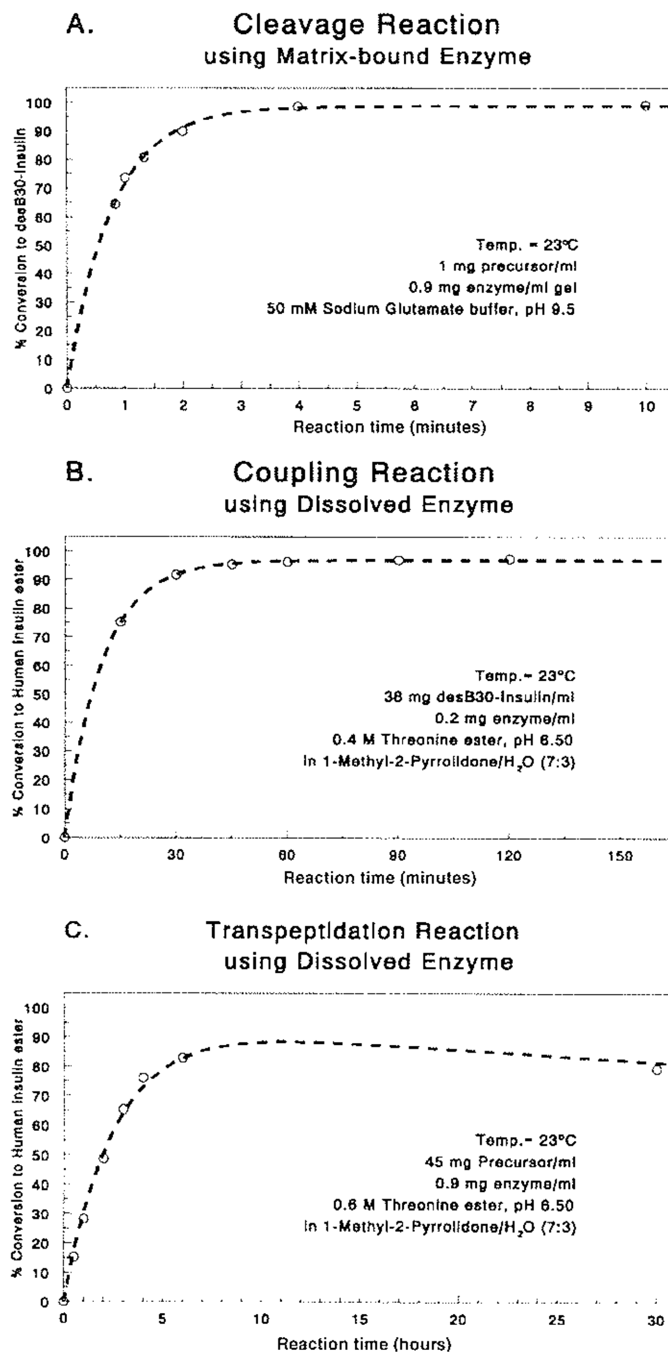

Figure 5.9. Typical conversion kinetics of the insulin precursor catalyzed by *A. lyticus* protease. A: the panel shows cleavage of the insulin precursor by matrix-bound enzyme. B: the panel shows coupling of threonine ester to desB30 insulin. C: the panel shows a transpeptidation conversion of the insulin precursor to human insulin ester. The specific enzymatic conditions are indicated in each of the panels and the conversion of the insulin precursor was determined by RP-HPLC analysis.

organic solvent such as (*N*,*N*)-dimethyl acetamide or *N*-methyl-2-pyrrolidone and the pH value should be close to neutral. The conditions for the coupling reaction and for the transpeptidation reaction are similar. However, the water content of the conversion reaction mixture is very critical for achieving an optimal transpeptidation reaction (Markussen, 1987). The described methods for conversion of the insulin precursor to human insulin are suitable for up-scaling and have been used for pharmaceutical production of human insulin.

Secretory expression and characterization of insulin in alternative yeast hosts

S. cerevisiae has certain limitations as the expression host for heterologous proteins. In general, fermentation yields tend to be low, with a maximum between 1–5% of total protein synthesis. Furthermore, plasmid instability and hyperglycosylation may also present problems (Buckholz and Gleeson, 1991). Hyperglycosylation of secreted heterologous protein expressed in *S. cerevisiae* may cause undesired immunogenic effects (Moir and Dumais, 1987; Van Arsdell *et al.*, 1987) although this is not an issue for insulin production. Furthermore, production of heterologous proteins by *S. cerevisiae* is hampered by the Crabtree effect, which can result in extensive ethanol formation at the expense of biomass and production of heterologous protein. However, a number of other yeast species are now becoming important alternative expression organisms for the production of heterologous proteins. These include methylotrophic species such as *Pichia pastoris*, *Hansenula polymorpha* and *Pichia methanolica,* as well as other yeast species like *Saccharomyces kluyveri, S. pombe, Kluyveromyces lactis* and *Yarrowia lipolytica* (for comparison of yeast expression systems see e.g. Buckholz and Gleeson, 1991; Romanos *et al.*, 1992; Romanos, 1995; Gellissen and Hollenberg, 1997; Müller *et al.*, 1998). This opens new possibilities for secretory expression of insulin in yeast. To evaluate the potential of new yeast expression hosts for expression of insulin, the methylotrophic yeast *P. pastoris* has been chosen based on efficient secretion of a number of different proteins, a strong and tightly regulated promoter (*AOX1*), integrating expression plasmids, a Crabtree negative status and its ability to grow to a high cell density. In the previous sections we have described adaptation of the insulin molecule to efficient secretory expression in *S. cerevisiae*, creating an efficient large-scale, high-yield and long-term continuous fermentation production system. Here we describe secretory expression of insulin in the methylotrophic yeast *P. pastoris.*

SECRETORY EXPRESSION OF INSULIN IN THE YEAST *P. PASTORIS*

Proinsulin and the insulin precursor were expressed in *P. pastoris* using different leaders to confer secretory competence (Kjeldsen *et al.*, 1999a). Expression of an α-factor leader–proinsulin fusion protein in *P. pastoris* did not result in secretion of proinsulin. However, shortening of the C-peptide, deletion of Thr^{B30} and association with the *S. cerevisiae* α-factor leader, as described for secretory expression in *S. cerevisiae,* resulted in efficient secretory expression of the insulin precursor also in *P. pastoris*. A *P. pastoris* endopeptidase functionally equivalent to the *S. cerevisiae* Kex2 endoprotease cleaved the fusion protein. Genetically engineered leaders, developed for secretory expression in *S. cerevisiae*, also

facilitated secretion of the insulin precursor expressed in *P. pastoris*. As in *S. cerevisiae*, genetically engineered leaders without N-linked glycosylation sites also facilitated secretion of the insulin precursor expressed in *P. pastoris*. In contrast, an α-factor leader without N-linked glycosylation impaired the ability to facilitate secretion of the insulin precursor, showing the importance of N-linked carbohydrates for the α-factor leader's ability to facilitate secretion also in *P. pastoris*. These data indicate that the yeast species *P. pastoris* and *S. cerevisiae* have surprisingly similar overall characteristics regarding expression and secretion of insulin. This similarity indicates that the structural and biochemical properties of insulin are major determinants of biosynthesis in yeast and that species differences in the secretory pathway play a minor role. Comparable small-scale batch-fermentations of *S. cerevisiae* and *P. pastoris* had similar insulin precursor fermentation yield (Kjeldsen *et al.*, 1999a). A more relevant comparison of yeast species like *P. pastoris* for insulin precursor productivity as a function of time (efficiency use of fermenter capacity) requires comparison of insulin productivity in a high-cell-density batch-fermentation with that of *S. cerevisiae* in long-term continuous fermentation. However, a disadvantage of methylotroph expression hosts is the methanol requirement for induction of gene expression. This is an impediment to the production of the insulin precursor in fermentors of relevant volumes, especially compared to the stable long-term continuous *S. cerevisiae* fermentation based on the *POT* expression plasmid expressing a high-yield selective adapted insulin molecule as described in the previous sections. However, recent developments indicate that efficient gene expression in methylotroph expression hosts might be possible using promoters that do not require methanol for induction.

CHARACTERIZATION OF THE INSULIN PRECURSOR SECRETED FROM *P. PASTORIS*

In contrast to *S. cerevisiae*, the α-factor leader/insulin precursor fusion protein expressed in *P. pastoris* has been completely processed and only the insulin precursor and no hyperglycosylated fusion protein has been secreted to the culture supernatant. Thus, an advantage of *P. pastoris* is a more efficient maturation of the fusion protein. In *S. cerevisiae,* the Kex2 endoprotease processing of the fusion protein could be optimized and fermentation yield increased by insertion of a suitable spacer peptide in the fusion protein. Surprisingly, expression of a similar fusion protein with a spacer peptide in *P. pastoris* also increased the fermentation yield of the insulin precursor. Mass spectrometry of culture supernatant from a *P. pastoris* transformant expressing the α-factor leader/insulin precursor fusion protein (without a spacer peptide), showed that the secreted insulin precursor had the expected molecular weight (Kjeldsen *et al.*, 1999a). Thus, *P. pastoris* did not proteolytically degrade or modify the secreted insulin precursor. However, expression of a fusion protein with a spacer peptide (EEAEAEAEPK) resulted in secretion of insulin precursor molecules with different molecular weights. These included a molecule with molecular weight equivalent to that of the insulin precursor (indicating a cleavage after the spacer C-terminus and before Phe^{B1}), another molecule with a molecular weight equivalent to the N-terminally extended insulin precursor and other molecules with molecular weights that indicated

additional processing of the N-terminally extended insulin precursor, presumably within the spacer peptide. These results indicate that only the spacer peptide and not the insulin precursor was cleaved in *P. pastoris*. Consequently, other spacer peptides were tested and the spacer peptide EEGEPK was found to be resistant toward enzymatic processing in *P. pastoris*. Thus, the concept of increasing the fermentation yield by a spacer peptide, that is stable *in vivo* but easily removed *in vitro*, was also applicable for secretory expression in *P. pastoris*, although it was unclear how the spacer increased the fermentation yield.

Interestingly, insulin has also been expressed in another Crabtree negative yeast, *S. kluyveri*, with fermentation yield of approximately 12 mg/l (specific yield ≈ 1.2 mg/g) (Srivastava *et al.*, 2000), which is comparable to that obtained with *S. cerevisiae* and *P. pastoris* in small-scale batch fermentations. The Crabtree effect in *S. cerevisiae* complicates fermentations and this problem is further augmented by the promoters of glycolytic genes (often used for gene expression) which may not be as absolutely constitutive as first anticipated. A Crabtree negative *Saccharomyces* species such as *S. kluyveri* has potential for large-scale expression of the insulin precursor without requirement of toxic compounds like methanol for gene expression. Comparison of different yeast species and expression systems with respect to secretory expression of the insulin precursor is obviously complex, given the vast number of differences between plasmids, promoters, metabolism, glycosylation pattern, secretory capacity, large-scale fermentation characteristics, etc. Moreover, only small-scale batch fermentation of the insulin precursor has been described and it might be argued that the advantages of yeasts like *P. pastoris* and *S. kluyveri* would be more profound in large-scale fermentations.

A yeast expression host for a second-generation production system for insulin production should preferentially be Crabtree negative, have a high specific growth rate, growth characteristics suitable for continuous fermentation, grow to a high biomass in fermentors at commercially relevant volumes, have a robust secretory capacity and not modify the insulin precursor, e.g. by *O*-linked glycosylation. Furthermore, stable episomal plasmids or preferential integrating plasmids and strong promoters not requiring induction by toxic compounds should be available for gene expression. However, compared to the β-cell's elegant and efficient production of insulin, the secretory expression of insulin in yeast appears crude. The real challenge in the future will be development of a yeast production system for mature human insulin. This requires further understanding of insulin biosynthesis in yeast and additional adaptation of the insulin molecule for expression, processing and secretion in yeast. An important aspect of a second-generation insulin production system is likely to be yeast strains that co-express insulin and appropriate maturation enzyme, e.g. prohormone convertases, facilitating *in vivo* conversion to human insulin.

Acknowledgements

Excellent technical assistance by Annette F. Pettersson and critical reading of the manuscript by Carsten Hjort and Klavs Jørgensen are highly appreciated.

References

ACHSTETTER, T. AND WOLF, D.H. (1985). Hormone processing and membrane-bound proteinases in yeast. *EMBO Journal* **4**, 173–177.

ADAMS, M.J., BLUNDELL, T.L., DODSON, E.J., DODSON, G.G., VIJAYAN, M., BAKER, E.N., HARDING, M.M., HODGKIN, D.C., RIMMER, B. AND SHEAT, S. (1969). Structure of rhombohedral 2 Zinc insulin crystals. *Nature* **224**, 491–495.

ALBER, T. AND KAWASAKI, G. (1982). Nucleotide sequence of the triose phosphate isomerase gene of *Saccharomyces cerevisiae*. *Journal of Molecular and Applied Genetics* **1**, 419–434.

ARMSTRONG, K.A., SOM, T., VOLKERT, F.C., ROSE, A. AND BROACH, J.R. (1989). Propagation and expression of genes in yeast using 2-micron circle vectors. In: *Yeast genetic engineering*. Eds. P.J. Barr, A.J. Brake and P. Valenzuela, pp 165–192. Boston, MA: Butterworth.

BAKER, E.N., BLUNDELL, T.L., CUTFIELD, J.F., CUTFIELD, S.M., DODSON, E.J., DODSON, G.G., HODGKIN, D.M.C., HUBBARD, R.E., ISAACS, N.W., REYNOLDS, C.D., SAKABE, K., SAKABE, N. AND VIJAYAN, N.M. (1988). The structure of 2Zn pig insulin crystals at 1.5 Å resolution. *Royal Society of London. Philosophical Transactions. Series B. Biological Sciences* **319**, 369–456.

BITTER, G.A., CHEN, K.K., BANKS, A.R. AND LAI, P-H. (1984). Secretion of foreign proteins from *Saccharomyces cerevisiae* directed by α-factor gene fusions. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 5330–5334.

BLOND-ELGUINDI, S., CWIRLA, S.E., DOWER, W.J., LIPSHUTZ, R.J., SPRANG, S.R., SAMBROOK, J.F. AND GETHING, M.J. (1993). Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. *Cell* **75**, 717–728.

BLUNDELL, T., DODSON, G., HODGKIN, D. AND MERCOLA, D. (1972). Insulin: the structure in the crystal and its reflection in chemistry and biology. *Advances in Protein Chemistry* **26**, 279–402.

BOTSTEIN, D. AND FINK, G.R. (1988). Yeast: an experimental organism for modern biology. *Science* **240**, 1439–1443.

BRAKE, A.J. (1989). Secretion of heterologous proteins directed by the yeast α-factor leader. In: *Yeast genetic engineering*. Eds. P.J. Barr, A.J. Brake and P. Valenzuela, pp 269–280. Boston: Butterworth.

BRAKE, A.J., MERRYWEATHER, J.P., COIT, D.G., HEBERLEIN, U.A., MASIARZ, F.R., MULLENBACH, G.T., URDEA, M.S., VALENZUELA, P. AND BARR, P.J. (1984). α-Factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 4642–4646.

BRANGE, J., SKELBAEK-PEDERSEN, B., LANKJAER, L., DAMGAARD, U., EGE, H., HAVELUND, S., HEDING, L.G., JØRGENSEN, K.H., LYKKEBERG, J., MARKUSSEN, J., PINGEL, M. AND RASMUSSEN, E. (1987). *Galenics of insulin*. Berlin: Springer-Verlag.

BRICKNER, J.H. AND FULLER, R.S. (1997). SOI1 encodes a novel, conserved protein that promotes TGN-endosomal cycling of Kex2p and other membrane proteins by modulating the function of two TGN-localization signals. *Journal of Cell Biology* **139**, 23–36.

BUCKHOLZ, R.G. AND GLEESON, M.A.G. (1991). Yeast systems for the commercial production of heterologous proteins. *Bio/Technology* **9**, 1067–1072.

CAPLAN, S., GREEN, R., ROCCO, J. AND KURJAN, J. (1991). Glycosylation and structure of the yeast MFα1 α-factor precursor is important for efficient transport through the secretory pathway. *Journal of Bacteriology* **173**, 627–635.

CHAUDHURI, B., STEUBE, K. AND STEPHAN, C. (1992). The pro-region of the yeast prepro-α-factor is essential for membrane translocation of the human insulin-like growth factor-I *in vivo*. *European Journal of Biochemistry* **206**, 793–800.

CHRISTIANSEN, L. AND PETERSEN, J.G.L. (1994). A DNA construct encoding the YAP3 signal peptide. *Patent application* WO95/02059.

CUNNINGHAM, K.W. AND WICKNER, W.T. (1989). Yeast KEX2 protease and mannosyltransferase I are localized to distinct compartments of the secretory pathway. *Yeast* **5**, 25–33.

DEREWENDA, U., DEREWENDA, Z., DODSON, E.J., DODSON, G.G., BING, X. AND MARKUSSEN,

J. (1991). X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule. A completely inactive analogue. *Journal of Molecular Biology* **220**, 425–433.
DIERS, I.V., RASMUSSEN, E., LARSEN, P.H. AND KJÆRSIG, I-L. (1991). Yeast fermentation processes for insulin production. In: *Drug biotechnology regulation, scientific basis and practices*. Eds. Y-Y.H. Chiu and J.L. Gueriguian, pp 166–176. New York, NY: Marcel Dekker, Inc.
DODSON, E.J., DODSON, G.G., HUBBARD, R.E. AND REYNOLDS, C.D. (1983). Insulin's structural behavior and its relation to activity. *Biopolymers* **22**, 281–291.
DODSON, G. AND STEINER, D.F. (1998). The role of assembly in insulin's biosynthesis. *Current Opinion in Structural Biology* **8**, 189–194.
EGEL-MITANI, M., HANSEN, M.T., NORRIS, K., SNEL, L. AND FIIL, N.P. (1988). Competitive expression of two heterologous genes inserted into one plasmid in *Saccharomyces cerevisiae*. *Gene* **73**, 113–120.
ELLGAARD, L., MOLINARI, M. AND HELENIUS, A. (1999). Setting the standards: quality control in the secretory pathway. *Science* **286**, 1882–1888.
FLYNN, G.C., POHL, J., FLOCCO, M.T. AND ROTHMAN, J.E. (1991). Peptide-binding specificity of the molecular chaperone BiP. *Nature* **353**, 726–730.
FULLER, R.S., STERNE, R.E. AND THORNER, J. (1988). Enzymes required for yeast prohormone processing. *Annual Review of Physiology* **50**, 345–362.
GELLISSEN, G. AND HOLLENBERG, C.P. (1997). Application of yeasts in gene expression studies: a comparison of *Saccharomyces cerevisiae, Hansenula polymorpha* and *Kluyveromyces lactis* – a review. *Gene* **190**, 87–97.
GLICK, B.R. AND PASTERNAK, J.J. (1998). *Molecular biotechnology: principles & applications of recombinant DNA* 2nd Edition, Chap. 7, pp 146–154. Washington D.C.: American Society for Microbiology.
GROSS, D.J., VILLA-KOMAROFF, L., KAHN, C.R., WEIR, G.C. AND HALBAN, P.A. (1989). Deletion of a highly conserved tetrapeptide sequence of the proinsulin connecting peptide (C-peptide) inhibits proinsulin to insulin conversion by transfected pituitary corticotroph (AtT20) cells. *Journal of Biological Chemistry* **264**, 21486–21490.
HITZEMAN, R.A., LEUNG, D.W., PERRY, L.J., KOHR, W.J., LEVINE, H.L. AND GOEDDEL, D.V. (1983). Secretion of human interferons by yeast. *Science* **219**, 620–625.
HONG, E., DAVIDSON, A.R. AND KAISER, C.A. (1996). A pathway for targeting soluble misfolded proteins to the yeast vacuole. *Journal of Cell Biology* **135**, 623–633.
JAKOB, C.A., BURDA, P., TE HEESEN, S., AEBI, M. AND ROTH, J. (1998). Genetic tailoring of N-linked oligosaccharides: the role of glucose residues in glycoproteins processing of *Saccharomyces cerevisiae in vivo*. *Glycobiology* **8**, 155–164.
JONASSEN, I., CLAUSEN, I.G., JENSEN, E.B. AND SVENDSEN, A. (1994). DNA sequences encoding insulin precursors and methods of production. *US Patent* **5**,324,641.
JULIUS, D., BLAIR, L., BRAKE, A., SPRAGUE, G. AND THORNER, J. (1983). Yeast α-factor is processed from a larger precursor polypeptide: the essential role of a membrane-bound dipeptidyl aminopeptidase. *Cell* **32**, 839–852.
JULIUS, D., BRAKE, A., BLAIR, L., KUNISAWA, R. AND THORNER, J. (1984a). Isolation of the putative structural gene for the lysine-arginine-cleaving endoprotease required for processing of yeast prepro-alpha-factor. *Cell* **37**, 1075–1089.
JULIUS, D., SCHEKMAN, R. AND THORNER, J. (1984b) Glycosylation and processing of prepro-α-factor through the yeast secretory pathway. *Cell* **36**, 309–318.
KAARSHOLM, N.C., KO, H-C. AND DUNN, M.F. (1989). Comparison of solution structural flexibility and zinc binding domains for insulin, proinsulin, and miniproinsulin. *Biochemistry* **28**, 4427–4435.
KAARSHOLM, N.C., NORRIS, K., JØRGENSEN, R.J., MIKKELSEN, J., LUDVIGSEN, S., OLSEN, O.H., SØRENSEN, A.R. AND HAVELUND, S. (1993). Engineering stability of the insulin monomer fold with application to structure-activity relationships. *Biochemistry* **32**, 10773–10778.
KAWASAKI, G.H. AND BELL, L. (1999). Stable DNA constructs. *US Patent* **5**,871,957.
KJELDSEN, T., BRANDT, J., ANDERSEN, A.S., EGEL-MITANI, M., HACH, M., PETTERSSON, A.F. AND VAD, K. (1996). A removable spacer peptide in an α-factor-leader/insulin precursor

fusion protein improves processing and concomitant yield of the insulin precursor in *Saccharomyces cerevisiae*. *Gene* **170**, 107–112.

KJELDSEN, T., PETTERSSON, A.F., HACH, M., DIERS, I., HAVELUND, S., HANSEN, P.H. AND ANDERSEN, A.S. (1997). Synthetic leaders with potential BiP binding mediate high-yield secretion of correctly folded insulin precursors from *Saccharomyces cerevisiae*. *Protein Expression and Purification* **9**, 331–336.

KJELDSEN, T., ANDERSEN, A.S., HACH, M., DIERS, I., NIKOLAJSEN, J. AND MARKUSSEN, J. (1998a). α-Factor pro-peptide N-linked oligosaccharides facilitate secretion of the insulin precursor in *Saccharomyces cerevisiae*. *Biotechnology and Applied Biochemistry* **27**, 109–115.

KJELDSEN, T., HACH, M., BALSCHMIDT, P., HAVELUND, S., PETTERSSON, A.F. AND MARKUSSEN, J. (1998b). Prepro-leaders lacking N-linked glycosylation for secretory expression in the yeast *Saccharomyces cerevisiae*. *Protein Expression and Purification* **14**, 309–316.

KJELDSEN, T., PETTERSSON, A.F. AND HACH, M. (1999a). Secretory expression and characterization of insulin in *Pichia pastoris*. *Biotechnology and Applied Biochemistry* **29**, 79–89.

KJELDSEN, T., PETTERSSON, A.F. AND HACH, M. (1999b). The role of leaders in intracellular transport and secretion of the insulin precursor in the yeast *Saccharomyces cerevisiae*. *Journal of Biotechnology* **75**, 195–208.

KRAULIS, P.J. (1991). *MOLSCRIPT*: a program to produce both detailed and schematic plots of protein structures. *Journal of Applied Crystallography* **24**, 946–950.

KURJAN, J. AND HERSKOWITZ, I. (1982). Structure of a yeast pheromone gene (MFα): A putative α-factor precursor contains four tandem copies of mature α-factor. *Cell* **30**, 933–943.

LUDVIGSEN, S., ROY, M., THØGERSEN, H. AND KAARSHOLM, N.C. (1994). High-resolution structure of an engineered biologically potent insulin monomer, B16 Tyr → His, as determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **33**, 7998–8006.

LUDVIGSEN, S., OLSEN, H.B. AND KAARSHOLM, N.C. (1998). A structural switch in a mutant insulin exposes key residues for receptor binding. *Journal of Molecular Biology* **279**, 1–7.

MACKAY, V.L., YIP, C., WELCH, S., GILBERT, T., SEIDEL, P., GRANT, F. AND O'HARA, P. (1990). Glycosylation and export of heterologous proteins expressed in yeast. In: *Recombinant systems in protein expression*. Eds. K.K. Alitalo, M-L. Huhtala, J. Knowles and A. Vaheri, pp 25–36. Amsterdam: Elsevier Science Publishers B.V.

MARKUSSEN, J. (1987). Human insulin by tryptic transpeptidation of porcine insulin and biosynthetic precursors. *Doctoral thesis*, Chap. 5. England: MTP Press Limited.

MARKUSSEN, J., JØRGENSEN, K.H., SØRENSEN, A.R. AND THIM, L. (1985). Single chain des-(B30) insulin. *International Journal of Peptide Protein Research* **26**, 70–77.

MARKUSSEN, J., DAMGAARD, U., DIERS, I., FIIL, N., HANSEN, M.T., LARSEN, P., NORRIS, F., NORRIS, K., SCHOU, O., SNEL, L., THIM, L. AND VOIGT, H.O. (1987). Biosynthesis of human insulin in yeast via single-chain precursors. In: *Peptides 1986*. Ed. D. Theodoropoulos, pp 189–194. Berlin: Walter de Gruyter & Co.

MOIR, D.T. AND DUMAIS, D.R. (1987). Glycosylation and secretion of human alpha-1-antitrypsin by yeast. *Gene* **56**, 209–217.

MÜLLER, S., SANDAL, T., KAMP-HANSEN, P. AND DALBØGE, H. (1998). Comparison of expression systems in the yeasts *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Kluyveromyces lactis*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. *Yeast* **14**, 1267–1283.

NAGAKAWA, S.H. AND TAGER, H.S. (1987). Role of the COOH-terminal B-chain domain in insulin–receptor interactions. *Journal of Biological Chemistry* **262**, 12054–12058.

NAGAKAWA, S.H. AND TAGER, H.S. (1993). Importance of main-chain flexibility and the insulin fold in insulin–receptor interactions. *Biochemistry* **32**, 7237–7243.

OH, M., MOK, K.H. AND LIM, Y. (1998). A comparison of three dimensional structures of biosynthesized preproinsulin and insulin using NMR. *Agricultural Chemistry and Biotechnology* **41**, 572–577.

OLSEN, H.B., LUDVIGSEN, S. AND KAARSHOLM, N.C. (1996). Solution structure of an engineered insulin monomer at neutral pH. *Biochemistry* **35**, 8836–8845.

OLSEN, H.B., LUDVIGSEN, S. AND KAARSHOLM, N.C. (1998). The relationship between insulin bioactivity and structure in the NH_2-terminal A-chain helix. *Journal of Molecular Biology* **284**, 477–488.

PACE, C.N. (1975). The stability of globular proteins. *CRC Critical Reviews in Biochemistry* **3**, 1–43.

PARODI, A.J. (2000). Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoproteins folding and degradation. *Biochemical Journal* **348**, 1–13.

PIGGOTT, J.R., WATSON, M.E.E., DOEL, S.M., GOODEY, A.R. AND CARTER, B.L.A. (1987). The secretion and post translational modification of interferons from *Saccharomyces cerevisiae*. *Current Genetics* **12**, 561–567.

PITTMAN IV, I., NAGAKAWA, S.H., TAGER, H.S. AND STEINER, D.F. (1997). Maintenance of the B-chain β-turn in [GlyB24]insulin mutants: A steady-state fluorescence anisotropy study. *Biochemistry* **36**, 3430–3437.

ROMANOS, M. (1995). Advances in the use of *Pichia pastoris* for high-level expression. *Current Opinion in Biotechnology* **6**, 527–533.

ROMANOS, M.A., SCORER, C.A. AND CLARE, J.J. (1992). Foreign gene expression in yeast: a review. *Yeast* **8**, 423–488.

RUSSELL, P.R. (1985). Transcription of the triose-phosphate-isomerase gene of *Schizosaccharomyces pombe* initiates from a start point different from that on *Saccharomyces cerevisiae*. *Gene* **40**, 125–130.

SCHEKMAN, R. AND NOVICK, P. (1982). The secretory process and yeast cell-surface assembly. In: *The molecular biology of the yeast Saccharomyces*. Eds. J.N. Strathen, E.W. Jones and J.R. Broach, pp 361–393. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

SINGH, A., LUGOVOY, J.M., KOHR, W.J. AND PERRY, L.J. (1984). Synthesis, secretion and processing of α-factor-interferon fusion proteins in yeast. *Nucleic Acids Research* **12**, 8927–8938.

SRIVASTAVA, A., PISKUR, J., NIELSEN, J. AND EGEL-MITANI, M. (2000). Method for the production of heterologous polypeptides in transformed yeast cells. *PCT patent application* WO 00/14258.

STEINER, D.F., CUNNINGHAM, D., SPIGELMAN, L. AND ATEN, B. (1967). Insulin biosynthesis: evidence for a precursor. *Science* **157**, 697–700.

STEINER, D.F., CHAN, S.J., WELSH, J.M., NIELSEN, D., MICHAEL, J., TAGER, H.S. AND RUBENSTEIN, A.H. (1986). Models of peptide biosynthesis: the molecular and cellular basis of insulin production. *Clinical and Investigative Medicine* **9**, 328–336.

THIM, L., HANSEN, M.T., NORRIS, K., HOEGH, I., BOEL, E., FORSTROM, J., AMMERER, G. AND FIIL, N.P. (1986). Secretion and processing of insulin precursors in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **83,** 6766–6770.

THIM, L., HANSEN, M.T. AND SØRENSEN, A.R. (1987). Secretion of human insulin by a transformed yeast cell. *FEBS Letters* **212**, 307–312.

THIM, L., SNEL, L., NORRIS, K. AND HANSEN, M.T. (1989). Insulin and insulin analogs from *Saccharomyces cerevisiae*. In: *Genetics and molecular biology of industrial microorganisms*. Eds. C.L. Hersberger, S.W. Queener and G. Hegeman, pp 322–328. Washington, DC: American Society for Microbiology.

THORNER, J. (1981). Pheromonal regulation of development in *Saccharomyces cerevisiae*. In: *The molecular biology of the yeast Saccharomyces cerevisiae: Life cycle and inheritance*. Eds. J.N. Strathern, E.W. Jones and J.R. Broach, pp 143–180. New York, NY: Cold Spring Harbor Laboratory.

VALENZUELA, P., MEDINA, A., RUTTER, W.J., AMMERER, G. AND HALL, B.D. (1982). Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. *Nature* **298**, 347–350.

VAN ARSDELL, J.N., KWOK, S., SCHWEICKART, V.L., LADNER, M.B., GELFAND, D.H. AND INNIS, M.A. (1987). Cloning, characterization, and expression in *Saccharomyces cerevisiae* of endoglucanase I from *Trichoderma reesei*. *Bio/Technology* **5**, 60–64.

WEISS, M.A., FRANK, B.H., KHAIT, I., PEKAR, A., HEINEY, R., SHOELSON, S.E. AND NEURINGER, L.J. (1990). NMR and photo-CIDNP studies of human proinsulin and prohormone processing intermediates with application to endopeptidase recognition. *Biochemistry* **29**, 8389–8401.

WHITTINGHAM, J.L., EDWARDS, D.J., ANTSON, A.A., CLARKSON, J.M. AND DODSON, G.G. (1998). Interactions of phenol and *m*-cresol in the insulin hexamer, and their effect on the association properties of B28 Pro → Asp insulin analogues. *Biochemistry* **37**, 11516–11523.

WILCOX, C.A., REDDING, K., WRIGHT, R. AND FULLER, R.S. (1992). Mutation of a tyrosine localization signal in the cytosolic tail of yeast Kex2 protease disrupts Golgi retention and results in default transport to the vacuole. *Molecular Biology of the Cell* **3**, 1353–1371.

XIE, L. AND TSOU, C-L. (1993). Comparison of secondary structures of insulin and proinsulin by FTIR. *Journal of Protein Chemistry* **12**, 483–487.

ZSEBO, K.M., LU, H.-S., FIESCHKO, J.C., GOLDSTEIN, L., DAVIS, J., DUKER, K., SUGGS, S.V., LAI, P.-H. AND BITTER, G.A. (1986). Protein secretion from *Saccharomyces cerevisiae* directed by the prepro-α-factor leader region. *Journal of Biological Chemistry* **261**, 5858–5865.