PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [6] :
C12N 15/62, 15/81, 15/11

A1

(11) International Publication Number: WO 95/35384

(43) International Publication Date: 28 December 1995 (28.12.95)

(21) International Application Number: PCT/DK95/00250

(22) International Filing Date: 16 June 1995 (16.06.95)

(30) Priority Data:
0712/94 17 June 1994 (17.06.94) DK

(71) Applicant: NOVO NORDISK A/S [DK/DK]; Novo Allé, DK-2880 Bagsvaerd (DK).

(72) Inventors: VAD, Knud; Duevej 112, 2.tv., DK-2000 Frederiksberg (DK). BRANDT, Jakob; Mågevej 84, 2.th., DK-2400 København NV (DK). KJELDSEN, Thomas, Børglum; Kvædevej 73, DK-2830 Virum (DK).

(74) Agent: HOUMØLLER, Peter; Novo Nordisk A/S, Corporate Patents, Novo Allé, DK-2880 Bagsvaerd (DK).

(81) Designated States: AM, AU, BB, BG, BR, BY, CA, CN, CZ, EE, FI, GE, HU, IS, JP, KG, KP, KR, KZ, LK, LR, LT, LV, MD, MG, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TT, UA, UG, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: N-TERMINALLY EXTENDED PROTEINS EXPRESSED IN YEAST

(57) Abstract

The present invention relates to polypeptides expressed and processed in yeast, a DNA construct comprising a DNA sequence encoding polypeptide having the following structure signal peptide-leader peptide $-X_1-X_2-X_3-X_4-X_5-X_6-X_7-$ heterologous protein, X_1 is Lys or Arg; X_2 is Lys or Arg, X_1 and X_2 together defining a yeast processing site; X_3 is Glu or Asp; X_4 is a sequence of amino acids with the following structure (A-B)n wherein A is Glu or Asp, B is Ala, Val, Leu or Pro, and n is 0 or an integer from 1 to 5, and when $n > 2$ each A and B is the same or different from the other A(s) and B(s), or X_4 is a sequence of amino acids with the following structure (C)m wherein C is Glu or Asp, and m is 0 or an integer from 1 to 5; X_5 is a peptide bond or is one or more amino acids which may be the same or different; X_6 is a peptide bond or an amino acid residue selected from the group consisting of Pro, Asp, Thr, Ser, Glu, Ala and Gly; and X_7 is Lys or Arg.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

N-terminally Extended Proteins Expressed in Yeast

FIELD OF INVENTION

The present invention relates to polypeptides expressed and processed in yeast, a DNA construct comprising a DNA sequence encoding such polypeptides, vectors carrying such DNA fragments and yeast cells transformed with the vectors, as well as a process of producing heterologous proteins in yeast.

BACKGROUND OF THE INVENTION

Yeast organisms produce a number of proteins synthesized intracellularly, but having a function outside the cell. Such extracellular proteins are referred to as secreted proteins. These secreted proteins are expressed initially inside the cell in a precursor or a pre-form containing a presequence ensuring effective direction of the expressed product across the membrane of the endoplasmic reticulum (ER). The presequence, normally named a signal peptide, is generally cleaved off from the desired product during translocation. Once entered in the secretory pathway, the protein is transported to the Golgi apparatus. From the Golgi the protein can follow different routes that lead to compartments such as the cell vacuole or the cell membrane, or it can be routed out of the cell to be secreted to the external medium (Pfeffer, S.R. and Rothman, J.E. Ann.Rev.Biochem. 56 (1987), 829-852).

Several approaches have been suggested for the expression and secretion in yeast of proteins heterologous to yeast. European Publication No. 0088632A describes a process by which proteins heterologous to yeast are expressed, processed and secreted by transforming a yeast organism with an expression vector harbouring DNA encoding the desired protein and a signal peptide, preparing a culture of the transformed organism, growing the culture and recovering the protein from the culture medium. The signal peptide may be the desired proteins own signal peptide, a heterologous signal peptide or a hybrid of native and heterologous signal peptide.

A problem encountered with the use of signal peptides heterologous to yeast might be that the heterologous signal peptide does not ensure efficient translocation and/or cleavage after the signal peptide.

The Saccharomyces cerevisiae MFα1 (α-factor) is synthesized as a prepro form of 165 amino acids comprising a 19 amino acids long signal- or prepeptide followed by a 64 amino acids long "leader" or propeptide, encompassing three N-linked glycosylation sites followed by (LysArg(Asp/Glu, Ala)$_{2-3}$α-factor)$_4$ (Kurjan, J. and Herskowitz, I. Cell 30 (1982), 933-943). The signal-leader part of the preproMFα1 has been widely employed to obtain synthesis and secretion of heterologous proteins in S. cerivisiae.

Use of signal/leader peptides homologous to yeast is known from i.a. US Patent No. 4,546,082, European Publications Nos. 0116201A, 0123294A, 0123544A, 0163529A, and 0123289A and European Patent No. 0100561B.

In EP 0123289A utilization of the S. cerevisiae α-factor precursor is described whereas EP 100561 describes the utilization of the Saccharomyces cerevisiae PH05 signal and PCT Publication No. WO 95/02059 describes the utilization of YAP3 signal peptide for secretion of foreign proteins.

US Patent No. 4,546,082 and European Publications Nos. 0016201A, 0123294A, 0123544A, and 0163529A describe processes by which the α-factor signal-leader from Saccharomyces cerevisiae (MFα1 or MFα2) is utilized in the secretion process of expressed heterologous proteins in yeast. By fusing a DNA sequence encoding the S. cerevisiae MFα1 signal/leader sequence at the 5' end of the gene for the desired protein secretion and processing of the desired protein was demonstrated.

EP 206,783 discloses a system for the secretion of polypeptides from S. cerevisiae whereby the α-factor leader sequence has been truncated to eliminate the four α-factor peptides present on the native leader sequence so as to leave the leader peptide itself fused to a heterologous polypeptide via the α-factor processing site Lys-Arg-Glu-Ala-Glu-Ala. This construction is indicated to lead to an efficient process of smaller peptides (less than 50 amino acids). For the secretion and processing of larger polypeptides, the native α-factor leader sequence has been truncated to leave one or two α-factor peptides between the leader peptide and the polypeptide.

A number of secreted proteins are routed so as to be exposed to a proteolytic processing system which can cleave the peptide bond at the carboxy end of two consecutive basic amino acids. This enzymatic activity is in S. cerevisiae encoded by the KEX 2 gene (Julius, D.A. et al., Cell 37 (1984b), 1075). Processing of the product by the KEX 2 protease is needed for the secretion of active S. cerevisiae mating factor α1 (MFα1 or α-factor) but is not involved in the secretion of active S. cerevisiae mating factor a.

Secretion and correct processing of a polypeptide intended to be secreted is obtained in some cases when culturing a yeast organism which is transformed with a vector constructed as indicated in the references given above. In many cases, however, the level of secretion is very low or there is no secretion, or the proteolytic processing may be incorrect or incomplete. As described in PCT Publication No. WO 90/10075 this is believed to be ascribable, to some extent, to an insufficient exposure of the processing site present between the C-terminal end of the leader peptide and the N-terminal end of the heterologous protein so as to render it inaccessible or, at least, less accessible to proteolytic cleavage.

WO 90/10075 describes a yeast expression system with improved processing of a heterologous polypeptide obtained by providing certain modifications near the processing site at the C-terminal end of the leader peptide and/or the N-terminal end of a heterologous polypeptide fused to the leader peptide. Thus, it is possible to obtain a higher yield of the correctly processed protein than is obtainable with unmodified leader peptide-heterologous polypeptide constructions.

SUMMARY OF THE INVENTION

The present invention describes modifications of the N-terminal end of the heterologous polypeptide designed as extensions which can be cleaved off either by naturally occurring yeast proteases before purification from the culture media or by in vitro proteolysis during or subsequently to purification of the product from the culture media.

The present invention relates to a DNA construct encoding a polypeptide having the following structure
signal peptide-leader peptide-X^1-X^2-X^3-X^4-X^5-X^6-X^7-heterologous protein wherein
X^1 is Lys or Arg;
X^2 is Lys or Arg, X^1 and X^2 together defining a yeast processing site;
X^3 is Glu or Asp;
X^4 is a sequence of amino acids with the following structure

$(A - B)_n$

wherein
A is Glu or Asp,
B is Ala, Val, Leu or Pro, and
n is 0 or an integer from 1 to 5, and when $n \geq 2$

each A and B is the same or different from the other A(s) and B(s), or

X^4 is a sequence of amino acids with the following structure

$(C)_m$

wherein

C is Glu or Asp, and

m is 0 or an integer from 1 to 5;

X^5 is a peptide bond or is one or more amino acids which may be the same or different;

X^6 is a peptide bond or an amino acid residue selected from the group consisting of Pro, Asp, Thr, Glu, Ala and Gly; and

X^7 is Lys or Arg.

In the present context, the term "signal peptide" is understood to mean a presequence which is predominantly hydrophobic in nature and present as an N-terminal sequence on the precursor form of an extracellular protein expressed in yeast. The function of the signal peptide is to allow the heterologous protein to be secreted to enter the endoplasmic reticulum. The signal peptide is normally cleaved off in the course of this process. The signal peptide may be heterologous or homologous to the yeast organism producing the protein but a more efficient cleavage of the signal peptide may be obtained when it is homologous to the yeast organism in question.

The expression "leader peptide" is understood to indicate a predominantly hydrophilic peptide whose function is to allow the heterologous protein to be secreted to be directed from the endoplasmic reticulum to the Golgi apparatus and further to a secretory vesicle for secretion into the medium, (i.e. exportation of the expressed protein or polypeptide across the cell wall or at least through the cellular membrane into the periplasmic space of the cell).

The expression "heterologous protein" is intended to indicate a protein or polypeptide which is not produced by the host yeast organism in nature.

X^3-X^4-X^5-X^6-X^7 together form an extension at the N-terminal of the heterologous polypeptide. This extension not only increases the fermentation yield but is due to the presence of X^3, protected against dipeptidyl aminopeptidase (DPAP A) processing, resulting in a homogenous N-terminal of the polypeptide. The extension may be constructed in such a way that it will be cleaved off by naturally occurring yeast proteases other than DPAP such as, for instance, yeast aspartic protease 3 (YAP3) before purification of the heterologous protein product from the culture media. Alternatively, the extension may be constructed in such a way that it is resistant to proteolytic cleavage during fermentation so that the N-terminally extended heterologous protein product may be purified from the culture media for subsequent in vitro maturation, e.g. by trypsin, Achromobacter lyticus protease I or enterokinase.

In a still further aspect, the invention relates to a process for producing a heterologous protein in yeast, comprising cultivating the transformed yeast strain in a suitable medium to obtain expression and secretion of the heterologous protein, after which the protein is isolated from the medium.

DETAILED DESCRIPTION OF THE INVENTION

In the peptide structure $(A - B)_n$, n is preferably 2-4 and more preferably 3.

In preferred polypeptides according to the invention X^3 may be Glu, A may be Glu, B may be Ala, X^5 may be a peptide bond or Glu, or Glu Pro Lys Ala, or X^6 may be Pro or a peptide bond.

Examples of possible N-terminal extensions X^3-X^4-X^5-X^6-X^7 are:

Glu Glu Ala Glu Ala Glu (Pro/Ala) (Glu/Lys) (Ala/Glu/Lys/Thr) Arg Ala Pro Arg,

Glu Glu Ala Glu Ala Glu Pro Lys Ala (Thr/Pro) Arg,

Glu Glu Ala Glu Ala Glu Ala Glu Pro Arg,

Glu Glu Ala Glu Ala Glu Ala Glu Pro Lys,

Glu Glu Ala Glu Ala Glu Ala Glu Arg,

Glu Glu Ala Glu Ala Glu Ala (Asp/Ala/Gly/Glu) Lys,

Glu Glu Ala Glu Ala Glu Ala (Pro/Leu/Ile/Thr) Lys,

Glu Glu Ala Glu Ala Glu Ala Arg,

Glu Glu Ala Glu Ala Glu (Glu/Asp/Gly/Ala) Lys,

Glu Glu Ala Glu Ala Pro Lys,

Glu Glu Ala Pro Lys,

Asp Asp Ala Asp Ala Asp Ala Asp Pro Arg,

Glu Glu Glu Glu Pro Lys,

Glu Glu Glu Pro Lys,

Asp Asp Asp Asp Asp Lys, and

Glu Glu Pro Lys.

The signal peptide sequence of the polypeptide of the invention may be any signal peptide which ensures an effective direction of the expressed polypeptide into the secretory pathway of the cell. The signal peptide may be a naturally occurring signal peptide or a functional part thereof, or it may be a synthetic peptide. Suitable signal peptides have been found to be the α-factor signal peptide, the signal peptide of mouse salivary amylase, a modified carboxypeptidase signal peptide or the yeast BAR1 signal peptide and the yeast aspartic protease 3 (YAP3) signal peptide. The mouse salivary amylase signal sequence is described by O. Hagenbüchle et al., Nature 289, 1981, pp. 643-646. The carboxypeptidase signal sequence is described by L.A. Valls et al., Cell 48, 1987, pp. 887-897. The BAR1 signal peptide is disclosed in PCT Publication No. WO 87/02670. The yeast aspartic protease 3 signal peptide is described in PCT Publication No. 95/02059.

The leader peptide sequence of the polypeptide of the invention may be any leader peptide which is functional in directing the expressed polypeptide through the endoplasmic reticulum and further along the secretory pathway. Possible leader sequences which are suited for this purpose are natural leader peptides derived from yeast or other organisms, such as the α-factor leader or a functional analogue thereof. The leader peptide may also be a synthetic leader peptide, e.g. one of the synthetic leader peptides disclosed in PCT Publication No. WO 89/02463 or WO 92/11378.

The N-terminally extended heterologous protein produced by the method of the invention may be any protein which may advantageously be produced in yeast. Examples of such proteins are aprotinin, tissue factor pathway inhibitor or other protease inhibitors, and insulin or insulin precursors, insulin analogues, insulin-like growth factor I or II, human or bovine growth hormone, interleukin, tissue plasminogen activator, glucagon, glucagon-like peptide-1 (GLP-1), Factor VII, Factor VIII, Factor XIII, platelet-derived growth factor, enzymes, or a functional analogue of anyone of these proteins. In the present context, the term "functional analogue" is meant to indicate a polypeptide with a similar function as the native protein (this is intended to be understood as relating to the nature rather than the level of biological activity of the native protein). The polypeptide may be structurally similar to the native protein and may be derived from the native protein by addition of one or more amino acids to either or both the C- and N-terminal end of the native protein, substitution of one or more amino acids at one or a number of different sites in the native amino acid sequence, deletion of one or more amino acids at either or both ends of the native protein or at one or several sites in the amino acid sequence, or insertion of one or more amino acids at one or more sites in the native amino acid sequence. Such modifications are well known for several of the proteins mentioned above.

Examples of suitable insulin precursors and insulin analogues are B(1-29)-Ala-Ala-Lys-A(1-21) (as described in, e.g., EP 163 529), B(1-27)-Asp-Lys-Ala-Ala-Lys-A(1-21) (as described in, e.g., PCT Publication No. 95/00550), B(1-29)-Ala-Ala-Arg-A(1-21) (as described in , e.g., PCT Publication No. 95/07931), and B(1-29)-Ser-Asp-Asp-Ala-Arg-A(1-21).

The DNA construct of the invention encoding the polypeptide of the invention may be prepared synthetically by established standard methods, e.g. the phosphoamidite method described by S.L. Beaucage and M.H. Caruthers, Tetrahedron Letters 22, 1981, pp. 1859-1869, or the method described by Matthes et al., EMBO Journal 3, 1984, pp. 801-805. According to the phosphoamidite method, oligonucleotides are synthesized, e.g in an automatic DNA synthesizer, purified, duplexed and ligated to form the synthetic DNA construct. A currently preferred way of preparing the DNA construct is by polymerase chain reaction (PCR), e.g. as described in Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, NY, 1989).

The DNA construct of the invention may also be of genomic or cDNA origin, for instance obtained by preparing a genomic or cDNA library and screening for DNA sequences coding for all or part of the polypeptide of the invention by hybridization using synthetic oligonucleotide probes in accordance with standard techniques (cf. Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, 1989). In this case, a genomic or cDNA sequence encoding a signal and leader peptide may be joined to a genomic or cDNA sequence encoding the heterologous protein, after which the DNA sequence may be modified at a site corresponding to the amino acid sequence X^1-X^2-X^3-X^4-X^5-X^6-X^7 of the polypeptide, e.g. by inserting synthetic oligonucleotides encoding the desired amino acid sequence for homologous recombination in accordance with well-known procedures.

Finally, the DNA construct may be of mixed synthetic and genomic, mixed synthetic and cDNA or mixed genomic and cDNA origin prepared by annealing fragments of synthetic, genomic or cDNA origin (as appropriate), the fragments corresponding to various parts of the entire DNA construct, in accordance with standard techniques. Thus, it may be envisaged that the DNA sequence encoding the heterologous protein may be of genomic or cDNA origin, while the sequence encoding the signal and leader peptide as well as the sequence encoding the N-terminal extension X^1-X^2-X^3-X^4-X^5-X^6-X^7 may be prepared synthetically.

In a further aspect, the invention relates to a recombinant expression vector which is capable of replicating in yeast and which carries a DNA construct encoding the above-defined polypeptide. The recombinant expression vector may be any vector which is capable of replicating in yeast organisms. In the vector, the DNA sequence encoding the polypeptide of the invention should be operably connected to a suitable promoter sequence. The promoter may be any DNA sequence which shows transcriptional activity in yeast and may be derived from genes encoding proteins either homologous or heterologous to yeast. The promoter is preferably derived from a gene encoding a protein homologous to yeast. Examples of suitable promoters are the Saccharomyces cerevisiae Mα1, TPI, ADH or PGK promoters.

The DNA sequence encoding the polypeptide of the invention may also be operably connected to a suitable terminator, e.g. the TPI terminator (cf. T. Alber and G. Kawasaki, J. Mol. Appl. Genet. 1, 1982, pp. 419-434).

The recombinant expression vector of the invention further comprises a DNA sequence enabling the vector to replicate in yeast. Examples of such sequences are the yeast plasmid 2μ replication genes REP 1-3 and origin of replication. The vector may also comprise a selectable marker, e.g. the Schizosaccharomyces pombe TPI gene as described by P.R. Russell, Gene 40, 1985, pp. 125-130.

The procedures used to ligate the DNA sequences coding for the polypeptide of the invention, the promoter and the terminator, respectively, and to insert them into suitable yeast vectors containing the information necessary for yeast replication, are well known to persons skilled in the art (cf., for instance, Sambrook et al., op.cit.). It will be understood that the vector may be constructed either by first preparing a DNA construct containing the entire DNA sequence coding for the polypeptide of the invention and subsequently inserting this fragment into a suitable expression vector, or by sequentially inserting DNA fragments containing genetic information for the individual elements (such as the signal, leader or heterologous protein) followed by ligation.

The yeast organism used in the process of the invention may be any suitable yeast organism which, on cultivation, produces large amounts of the heterologous protein or polypeptide in question. Examples of suitable yeast organisms may be strains of the yeast species Saccharomyces cerevisiae, Saccharomyces kluyveri, Schizosaccharomyces pombe or Saccharomyces uvarum. The transformation of the yeast cells may for instance be effected by protoplast formation followed by transformation in a manner known per se. The medium used to cultivate the cells may be any conventional medium suitable for growing yeast organisms. The secreted heterologous protein, a significant proportion of which will be present in the medium in correctly processed form, may be recovered from the medium by conventional procedures including separating the yeast cells from the medium by centrifugation or filtration, precipitating the proteinaceous components of the supernatant or filtrate by means of a salt, e.g. ammonium sulphate, followed by purification by a variety of chromatographic procedures, e.g. ion exchange chromatography, affinity chromatography, or the like.

After secretion to the culture medium, the protein may be subjected to various procedures to remove the sequence X^3-X^4-X^5-X^6-X^7.

When X^6 is Pro, Thr, Ser, Gly, or Asp, the extensions are found to be stably attached to the heterologous protein during fermentation, protecting the N-terminal of the heterological protein against the proteolytic activity of yeast proteases such as DPAP. The presence of an N-terminal extension on the heterologous protein may also serve as a protection of the N-terminal amino group of the heterologous protein during chemical processing of the protein, i.e. it may serve as a substitute for a BOC or similar protecting group. In such cases the amino acid sequence X^3-X^4-X^5-X^6-X^7 may be removed from the recovered heterologous protein by means of a proteolytic enzyme which is specific for a basic amino acid (e.g. Lys or Arg) so that the terminal extension is cleaved off at X^7. Examples of such proteolytic enzymes are trypsin, Achromobacter lyticus protease I, Enterokinase, Fusarium oxysporum trypsin-like protease, and yeast aspartic protease 3 (YAP3) of Saccharomyces cerevisiae.

YAP3 has been characterized applying various prohormones and expressions of heterologous protein in Saccharomyces cerevisiae (Niamh, X.C. et al., FEBS Letters, 332, p. 273-276, 1993; Bourbonnais, Y. et al., EMBO Journal, 12, p. 285-294, 1993; Egel-Mitani, M. et al., YEAST, 6, p. 127-137, 1990).

YAP3 appears during large scale fermentations of Saccharomyces cerevisiae. It is active when the pH is from pH 3 to about pH 6. The appearance of YAP3 is most pronounced using minimal types of media for production. In such media the cells are starved of glucose and/or nitrogen, but other conditions which limit the growth conditions may be used. The YAP3 protease requires a defined motif flanking the cleavage site. Such a motif is present in the N-terminal extension when the amino acid immediately preceding X^7 is Ala or Glu (but not when X^6 is Pro, Thr, Ser, Gly or Asp).

Thus, when X^6 is a peptide bond (and the amino acid preceding X^7 is Ala or Glu) or Ala or Glu, the extensions may to be cleaved from the heterologous protein during fermentation, most likely subsequent to the secretion thereof from the yeast cells - a process which depends on YAP3 acting on its substrate (the peptide bond after lysine or arginine) in the yeast cells or in the growth media. YAP3 inside, attached to, or escaped from the yeast cells may selectively cleave off the extensions such that the non-extended product may be purified from the growth media.

The amino acid sequence X^3-X^7, wherein X^6 Ala or Glu or is a peptide bond (and the amino acid preceding X^7 is Ala or Glu) may be removed from the heterologous protein in the medium by a process involving subjecting the yeast cells to stress to make them release YAP3 into the medium, whereby the amino acid sequence X^3-X^7 is cleaved off.

The stress to which the yeast cells are subjected may comprise reducing the pH of the culture medium to below 6.0, preferably below 5, or starving the yeast cells of glucose and/or nitrogen.

In the process of the invention for producing a heterologous protein, the yeast cell may further be transformed with one or more genes encoding a protease which is specific for basic amino acid residues so that, on cultivation of the cell the gene or genes are expressed, the consequent production of protease ensuring a more complete cleavage of X^3 - X^7 from the heterologous polypeptide.

In a preferred embodiment, one or more additional genes encoding YAP3 may be used to transform the yeast cell so that, on cultivation of the cell the gene or genes are overexpressed, the consequent overproduction of YAP3 ensuring a more complete cleavage of X^3-X^7 from the heterologous polypeptide.

In the present context, yeast aspartic protease 3 (YAP3) embraces the native YAP3 enzyme as well as enzymes derived from the native enzyme, wherein the C-terminal has been modified in order to ensure efficient release of the YAP3 protein from the cell or wherein any modification has taken place maintaining the proteolytic activity.

Alternatively the protease may be A.lyticus protease I, Enterokinase or trypsin.

The present invention is described in further detail in the following examples which are not in any way intended to limit the scope of the invention as claimed.

The invention is described with reference to the drawings, wherein

Fig. 1 shows a general scheme for the construction of plasmids containing genes expressing N-terminally extended polypeptides.

In Fig. 1, the following symbols are used:

1: Denotes the TPI gene promoter sequence from _S. cerevisiae_.

2: Denotes the region encoding a signal/leader peptide (_e_.g. from the α-factor gene of _S. cerevisiae_).

3: Denotes the region encoding a heterologous polypeptide.

3*: Denotes the region encoding a N-terminal extended heterologous polypeptide.

4: Denotes the TPI gene terminator sequence of _S. cerevisiae_.

P1: Denotes a synthetic oligonucleotide PCR primer determining the structure of the N-terminal extension.

P2: Denotes a universal PCR primer for the amplification of region 3.

POT: Denotes TPI gene from _S. pombe_.

2μ Ori: Denotes a sequence from _S. cerevisiae_ 2 μ plasmid including its origin of DNA replication in _S. cerevisiae_.

Ap^R: Sequence from pBR322 /pUC13 including the ampicilli resistance gene and an origin of DNA replication in _E. coli_.

Fig. 2 shows the DNA sequence in pJB59 encoding the insulin precursor B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21) N-terminally fused to the 85 residues which make up the α-factor signal/leader peptide in which Leu in position 82 and Asp in position 83 have been substituted by Met and Ala, respectively, and

Fig. 3 the DNA sequence of pAK623 encoding GLP-$1_{7-36Ala}$ N-terminally fused to the synthetic signal/leader sequence "YAP3/$S1_{PAVA}$"

Fig. 4 the DNA sequence of pKV142 encoding B_{chain}(1-29)-Ala-Ala-Arg-A_{chain}(1-21) N-terminally fused to the 85 residues which make up the α-factor signal/leader peptide in which Leu in position 82 and Asp in position 83 have been substituted by Met and Ala, respectively.

Fig. 5 the DNA sequence of pAK679 encoding B_{chain}(1-29)-Ala-Ala-Lys-A_{chain}(1-21) N-terminally fused to the synthetic signal/leader sequence "YAP3/LA19"

Fig. 6 shows HPLC chromatograms of culture supernatants containing the insulin precursor B_{chain}(1-27)Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21) with or without N-terminal extensions; and with or without in vivo or in vitro processing of the N-terminal extensions.

Fig. 7 shows the products according to figure 4 separated by size on a 10% Tricine-SDS-PAGE gel.

Fig. 8 shows the effect of the presence of YAP3 coexpression on the yield derived from the HPLC data in pJB176 compared to pJB64.

EXAMPLES

Plasmids and DNA

All expressions plasmids are of the C-POT type (see figure 1), similar to those described in WO EP 171 142, which are characterized by containing the Schizosaccharomyces pombe triose phosphate isomerase gene (POT) for the purpose of plasmid selection and stabilization in *S. cerevisiae*. The plasmids furthermore contain the *S. cerevisiae* triose phosphate isomerase promoter (region 1 in figure 1) and terminator (region 4 in figure 1). These sequences are identical to the corresponding sequences in plasmid pKFN1003 (described in WO 90/100075) as are all sequences except the sequence of the EcoRI-XbaI fragment encoding the signal/leader/product (region 2 and 3 in figure 1). In order to express different heterologous proteins, the EcoRI-XbaI fragment of pKFN1003 is simply replaced by an EcoRI-XbaI fragment encoding the signal/leader/product of interest. Such EcoRI-XbaI fragments may be synthesized using synthetic oligonucleotides and PCR according to standard techniques (cf. Sambrook et al., 1989 supra).

Figure 1 shows the general scheme used for the construction of plasmids containing genes expressing N-terminally extended polypeptides, the scheme including the following steps.

- A sample of the C-POT plasmid vector is digested with restriction nucleases EcoRV and XbaI and the largest DNA fragment is isolated using standard molecular techniques (Sambrook J, Fritsch EL and Maniatis T, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989).

- Another sample of C-POT plasmid (which may be the same or different from the plasmid mentioned above) is digested with restriction nucleases EcoRV and NcoI and the fragment comprising region 1 and 2 is isolated.

- Polymerase Chain Reaction (PCR) is performed using the Gene Amp PCR reagent kit (Perkin Elmer, 761 Main Avewalk, CT 06859, USA) according to the manufacturer's instructions and the synthetic oligonucleotide primers P1 and P2 on a template (which may be the same or different from the plasmids mentioned above) encoding the heterological polypeptide of interest. P1 is designed in such a way that it includes a recognition site for restriction nuclease NcoI (5´CCATGG3´) followed by the sequences encoding a KEX2 processing site, the N-terminal extension and 10-15 nucleotides identical to the sequence encoding the original N-terminal of the heterologous protein of interest. P2 (e.g. 5´-AATTTATTTTACATAACACTAG-3´) is designed in such a way that it amplifies region 3 and the flanking recognition site for restriction nuclease XbaI (5´-TCTAGA-3´) in PCRs with P1 using standard techniques described in Sambrook et al., supra.

- The PCR product is digested with restriction nucleases NcoI and XbaI and the digested fragment is isolated.

- The fragments isolated are ligated together by T4 DNA ligase under standard conditions (described in Sambrook et al., supra.

- The ligation mixture is used to transform competent E.coli cells ap^{r-} and selected for ampicillin resistance according to Sambrook et al., supra.

- Plasmids are isolated from the resulting E.coli clones using standard molecular techniques (described in Sambrook et al., supra.

- DNA Sequencing is performed using enzymatic chain termination (Sequenase, United States Biochemicals) according to the manufacturer's instructions in order to verify (or determine) the DNA sequences encoding the N-terminal extended polypeptide and to ensure that it is in frame with the DNA sequence encoding the signal/leader peptide of region 2 .

- The plasmid is used to transform the yeast strain MT663 and selected for growth on glucose, as described in detail below:

Yeast transformation: S. cerevisiae strain MT663 (E2-7B XE11-36 a/α, Δtpi/Δtpi, pep 4-3/pep 4-3) (the yeast strain MT663 was deposited in the Deutsche Sammlung von Mikroorganismen und Zellkulturen in connection with filing WO 92/11378 and was given the deposit number DSM 6278) was grown on YPGaL (1% Bacto yeast extract, 2% Bacto peptone, 2% galactose, 1% lactate) to an O.D. at 600 nm of 0.6.

100 ml of culture was harvested by centrifugation, washed with 10 ml of water, recentrifugated and resuspended in 10 ml of a solution containing 1.2 M sorbitol, 25 mM Na_2EDTA pH = 8.0 and 6.7 mg/ml dithiotreitol. The suspension was incubated at 30°C for 15 minutes, centrifuged and the cells resuspended in 10 ml of a solution containing 1.2 M sorbitol, 10 mM Na_2EDTA, 0.1 M sodium citrate, pH 0 5.8, and 2 mg Novozym®234. The suspension was incubated at 30°C for 30 minutes, the cells collected by centrifugation, washed in 10 ml of 1.2 M sorbitol and 10 ml of CAS (1.2 M sorbitol, 10 mM $CaCl_2$, 10 mM Tris HCl (Tris = Tris(hydroxymethyl)aminomethane) pH = 7.5) and resuspended in 2 ml of CAS. For transformation, 1 ml of CAS-suspended cells was mixed with approx. 0.1 μg of plasmid DNA and left at room temperature for 15 minutes. 1 ml of (20% polyethylene glycol 4000, 10 mM $CaCl_2$, 10 mM Tris HCl, pH = 7.5) was added and the mixture left for a further 30 minutes at room temperature. The mixture was centrifuged and the pellet resuspended in 0.1 ml of SOS (1.2 M sorbitol, 33% v/v YPD, 6.7 mM $CaCl_2$) and incubated at 30°C for 2 hours. The suspension was then centrifuged and the pellet resuspended in 0.5 ml of 1.2 M sorbitol. Then, 6 ml of top agar (the SC

medium of Sherman et al., Methods in Yeast Genetics, Cold Spring Harbor Laboratory (1982)) containing 1.2 M sorbitol plus 2.5% agar) at 52°C was added and the suspension poured on top of plates containing the same agar-solidified, sorbitol containing medium.

Example 1

Construction of pJB108 and pJB109

Plasmid pJB59 is a derivative of pKFN1003 in which the EcoRI-XbaI fragment encodes the insulin precursor B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21) N-terminally fused to a signal/leader sequence corresponding to the 85 residues of the α-factor prepro signal peptide in which Leu in position 82 and Asp in position 83 have been substituted by Met and Ala, respectively (Figure 2). The EcoRI-XbaI fragment is synthesized in an applied biosystems DNA synthesizer (Perkin Elmer DNA Thermal Cycler) in accordance with the manufacturer's instructions.

Plasmid constructs designed to express N-terminally extended insulin precursor B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21) were obtained by means of a P1-primer with the following sequence

```
        NcoI
5'-GGGGTATCCATGGCTAAGAGAGAAGAAGCTGAAGCTGAAGCT(AC)CAAAGTTCGTTAACCAACAC-3'
   GlyLeuSerMetAlaLysArgGluGluAlaGluAlaGluAla  Xaa LysPheValAsnGlnHis
```

the P2-primer: 5'-AATTTATTTTACATAACACTAG-3' and the plasmid pJB59 according to the general scheme described above. PCR and cloning resulted in a construct wherein the DNA sequence encoding the insulin precursor B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21) is preceded by a DNA sequence encoding the N-terminal extension Glu-Glu-Ala-Glu-Ala-Glu-Ala-Xaa-Lys, where Xaa is either Pro (pJB108; Pro encoded by CCA) or Thr (pJB109; Thr encoded by ACA).

Example 2

Construction of pJB44, pJB107, and pJB126

Plasmid constructs designed to express additional N-terminally extended versions of the insulin precursor B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21) were made by means of a P1-primer with the following sequence

```
         NcoI
5'-GGGGTATCCATGGCTAAGAGAGAAGAAGCTGAAGCTGAAGCTG(CAG)(AC)AAAGTTCGTTAACCAACAC-3'
   GlyLeuSerMetAlaLysArgGluGluAlaGluAlaGluAla   Xaa     LysPheValAsnGlnHis
```

the P2-primer: 5'-AATTTATTTTACATAACACTAG-3' and the plasmid pJB59 as described in example 1. Among the resulting plasmids pJB44, pJB107 and pJB126 where isolated. These plasmids encode the insulin precursor B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21) preceded by the N-terminal extension Glu-Glu-Ala-Glu-Ala-Glu-Ala-Xaa-Lys, where Xaa is either Glu (pJB44; Glu encoded by GAA), Asp (pJB126; Asp encoded by GAC) or Gly (pJB107; Gly encoded by GGC).

Example 3.

Construction of pJB64 and pJB110

Plasmid constructs designed to express N-terminally extended insulin precursor B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21) were obtained by means of a P1-primer with the following sequence

```
         NcoI
5'-GGGGTATCCATGGCTAAGAGAGAAGAAGCTGAAGCTGAAGG(AC)AAAGTTCGTTAACCAACAC-3'
   GlyLeuSerMetAlaLysArgGluGluAlaGluAlaGlu  Xaa  LysPheValAsnGlnHis
```

the P2-primer: 5'-AATTTATTTTACATAACACTAG-3' and the plasmid pJB59 as described in example 1. Among the resulting plasmids pJB64 and pJB110 were isolated. These plasmids encode N-terminal extensions of B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21) preceded by a DNA sequence encoding the N-terminal extension Glu-Glu-Ala-Glu-Ala-Glu-Xaa-Lys, where Xaa is either Glu (pJB110; Glu encoded by GAA) or Ala (pJB64; Ala encoded by GCA).

Example 4

Construction of pAK663

Plasmid pAK623 is a derivative of pKNF 1003 in which the EcoRI-XbaI fragment encodes GLP-$1_{7\text{-}36\text{Ala}}$ N-terminally fused to a synthetic signal leader sequence YAP3/$S1_{\text{PAVA}}$ (Fig. 3).

The EcoRI-XbaI fragment was synthesized in an Applied Biosystems DNA synthesizer according to the manufacturer's instructions.

Plasmid constructs designed to express GLP-$1_{7\text{-}36\text{ALA}}$ with the N-terminal extension in form of Glu-Glu-Ala-Glu-Ala-Glu-Ala-Glu-Arg was obtained by means of a P1-primer with the following sequence

```
         NcoI
5'-AACGTTGCCATGGCTCCAGCTCCAGCTAAGAGAGAAGAAGCTGAAGCTGAAGCTGAAAGACATGCTGAAGGT-3'
   AsnValAlaMetAlaProAlaValAlaLysArgGluGluAlaGluAlaGluAlaGluArgHisAlaGluGly
```

the P2-primer: 5'-AATTTATTTTACATAACACTAG-3' and the plasmid pAK623 according to the method described above resulting in plasmid pAK663.

Example 5

Expression of N-terminal extended products and the removal of the extensions

Yeast strain MT663 transformed with the C-POT plasmids described above (pJB59, pJB108, pJB109, pJB44, pJB126, pJB107, pJB64 and pJB110) were grown on YPD (1% yeast extract, 2% peptone and 2% glucose) agar plates. Single colonies were used to start 5 ml liquid cultures in YPD broth pH = 6.0, which were shaken for 72 hours at 30°C. Yields of products were determined directly on culture supernatants by the method described by Snel, Damgaard and Mollerup, Chromatographia 24, 1987, pp. 329-332.

The results with the yeast strains expressing N-terminally extended B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21) compared to the non-extended form (pJB59) are shown below

Plasmid	N-terminal extension	Yield
pJB59		100%
pJB108	Glu-Glu-Ala-Glu-Ala-Glu-Ala-Pro-Lys	275%
pJB109	Glu-Glu-Ala-Glu-Ala-Glu-Ala-Thr-Lys	300%
pJB44	Glu-Glu-Ala-Glu-Ala-Glu-Ala-Glu-Lys	325%*
pJB126	Glu-Glu-Ala-Glu-Ala-Glu-Ala-Asp-Lys	300%
pJB107	Glu-Glu-Ala-Glu-Ala-Glu-Ala-Gly-Lys	275%
pJB64	Glu-Glu-Ala-Glu-Ala-Glu-Ala-Lys	250%*
pJB110	Glu-Glu-Ala-Glu-Ala-Glu-Glu-Lys	225%*

Yields marked by (*) asterisk denotes that the product in these cases is a mixture of N-terminated extended B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21) and non-extended B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21), the latter being a result of the in vivo cleavage of the extension described below and illustrated in Figure 4a and 5.

In case of pAK663 expressing GLP-$1_{7\text{-}36Ala}$ N-terminally extended by Glu-Glu-Ala-Glu-Ala-Glu-Ala-Arg the yield was found to be 20 fold higher than pAK623 expressing non-extended GLP-$1_{7\text{-}36Ala}$.

Example 6

Removal of N-terminal extensions in vivo

Culture supernatants obtained from cultures of yeast strains transformed with plasmid pJB59, pJB64, pJB44 and pJB108 (see above) were evaluated by HPLC chromatography (Figure 4a) and parallel samples were run on a 10% Tricine-SDS-PAGE gel (Figure 5, lanes 1, 2, 3 and 4). From the HPLC chromatograms it appears that the culture supernatants from yeasts with plasmid pJB44 (chromatogram 3) and pJB64 (chromatogram 2) contain both the N-terminally extended as well as non-extended B_{chain}(1-27)Asp-Lys-

Ala-Ala-Lys-A(1-21), whereas culture supernatants from yeast with pJB108 chromatogram 4) only contain the N-terminally extended form. In case of pJB64 about 50% of the precursor is found in non-extended form (See also Fig. 5 lane 2) whereas this form only represent a minor part of pJB44.

These results illustrate the ability of yeast to cleave off N-terminal extensions selectively <u>in vivo</u> when the extension is either Glu-Glu-Ala-Glu-Ala-Glu-Ala-Lys (pJB64) or Glu-Glu-Ala-Glu-Ala-Glu-Ala-Glu-Lys (pJB44) and the inability of yeast to cleave off an N-terminal extension in the form of Glu-Glu-Ala-Glu-Ala-Glu-Ala-Pro-Lys (pJB108). The proteolytic activity responsible for cleaving off the extensions may be associated with enzymes in the secretory pathway such as membrane-bound YAP3 in the <u>trans-</u>Golgi system of the yeast cells.

<u>Example 7</u>

<u>Removal of N-terminal extensions in vitro</u>

The culture supernatants described above were used as substrates for proteolytic cleavage with either partially purified YAP3 enzyme isolated from yeast strain ME783 overexpressing YAP3 (Egel-Mitani et al. 1990) or <u>Achromobacter lyticus</u> protease I.

YAP3 assay was performed as follows:
4μl of YAP3 enzyme
800 μl of cell free growth media
Samples were incubated for 15 h at 37°C in 0,1 M Na citrate buffer, pH 4.0

<u>Achromobacter lyticus</u> protease I assay performed as follows:
10μg <u>A. lyticus</u> protease I
1ml of cell free growth media

Samples were incubated for 1 h at 37°C in 0,1 M Tris buffer, pH 8.75

Figure 4b and 4c show the results evaluated by HPLC chromatography obtained from the YAP3 and A. lyticus protease I digestions, respectively. Parallel samples were run on 10% Tricine-SDS-PAGE (Fig. 5 lane 5-12).

From the chromatograms of Fig. 4b it appears that the YAP3 enzyme is able to cleave off N-terminal extensions selectively when these are in form of Glu-Glu-Ala-Glu-Ala-Glu-Ala-Lys (pJB64) or Glu-Glu-Ala-Glu-Ala-Glu-Ala-Glu-Lys (pJB44) but not Glu-Glu-Ala-Glu-Ala-Glu-Ala-Pro-Lys (pJB108) (See also Fig. 5 lane 6, 7 and 8). This results clearly indicates that YAP3 or YAP3-like enzyme(s) are responsible for the partial cleavage of N-terminal extensions seen in vivo (cf. Example 6).

From the chromatograms of Fig 4c it appears that digestions with A. lyticus protease I in all cases result in the same product, namely B_{chain}(1-27)-Asp-Lys-(connected by disulfide bonds)-A_{chain}(1-21) which is the end result of proteolytic cleavage after all Lys-residues found in the B_{chain}(1-27)Asp-Lys-ala-Ala-Lys-A_{chain}(1-21) insulin precursor including those found between the B and A chain in the precursor.

In the case of the digestion of the culture supernatant from yeast transformed with pJB59 (expressing the non-extended B_{chain}(1-27)Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21)) a product is seen which does not appear in the other digestions. This product, Arg-B_{chain}(1-27)-Asp-Lys-(connected by disulfide bonds)-A_{chain}(1-21), results from A. lyticus protease I cleavage of secreted leader-precursor in the growth media. A. lyticus protease I cleaves between the Lys and Arg residues in the dibasic KEX2 site of the product which has escaped KEX2 cleavage in the secretory pathway of the yeast cells.

Example 8

Removal of N-terminal extensions by over-expressing YAP3

The YAP3 gene cloned by Egel-Mitani et al. (Yeast 6 (1990) pp. 127-137) was inserted into the C-POT plasmid pJB64 encoding Glu-Glu-Ala-Glu-Ala-Glu-Ala-Lys-B_{chain}(1-27)-Pro-Lys-Ala-Ala-Lys-A_{chain}(1-21) (see example 3) in the following way:

The 2.5 kb SalI/SacI fragment containing the YAP3 gene was isolated from plasmid pME768 (Egel-Mitani et al. Yeast 6 (1990) pp. 127-137) and inserted into the SalI and SacI site of plasmid pIC19R (March et al. Gene 32 (1984) pp. 481-485). From the resulting plasmid designated pME834 a 2.5 kb SalI/XhoI fragment containing the YAP3 gene was isolated and inserted into the unique SalI site placed between the POT and the Ap^R sequences of pJB64. One resulting plasmid was designated pJB176.

Yeast strain MT663 was transformed with pJB176 and analyzed as described in example 5.

As can be seen from the HPLC data on yield shown in Figure 8 the presence of the YAP3 gene in pJB176 clearly effects a higher percent of non-extended B_{chain}(1-27)-Pro-Lys-Ala-Ala-Arg-A_{chain}(1-21) in the culture-supernatant of the corresponding yeast transformants compared to the yeast transformant of pJB64.

This result illustrate the ability to make yeast strains with an enhanced capacity to cleave off N-terminal extensions selectively in vivo by manipulating the level of proteolysis caused by YAP3.

Example 9

Construction of pKV143

Plasmid pKV142 is a derivative of pKFN1003 in which the EcoRI-XbaI fragment encodes the insulin precursor B_{chain}(1-29)-Ala-Ala-Arg-A_{chain}(1-21) N-terminally fused to a signal/leader sequence corresponding to the 85 residues of the α-factor prepro signal peptide in which Leu in position 82 and Asp in position 83 have been substituted by Met and Ala, respectively (Figure 4).

Plasmid constructs designed to express B_{chain}(1-29)-Ala-Ala-Arg-A_{chain}(1-21) with a N-terminal extension in form Asp-Asp-Ala-Asp-Ala-Asp-Ala-Asp-Pro-Arg was obtained by means of a P1-primer with the following sequence

```
        NcoI
5'-GGGGTATCCATGGCTAAGAGAGACGACGCTGACGCTGACGCTGACCCAAGATTCGTTAACCAACACTTGTGCGG-3'
   GlyLeuSerMetAlaLysArgAspAspAlaAspAlaAspAlaAspProArgPheValAsnGlnHisLeuCys
```

the P2-primer: 5'-AATTTATTTTACATAACACTAG-3' and the plasmid pKV142 according to the general scheme described above.

Example 10

Construction of pKV102

Plasmid constructs designed to express B_{chain}(1-29)-Ala-Ala-Arg-A_{chain}(1-21) with a N-terminal extension in form Glu-Glu-Ala-Glu-Ala-Glu-Ala-Glu-Pro-Lys-Ala-Thr-Arg was obtained by means of a P1-primer with the following sequence

```
        NcoI
5'-GGGGTATCCATGGCTAAGAGAGAAGAAGCTGAAGCTGAAGCTGAACCAAAGGCTACAAGATTCGTTAACCAACACTTGTGCGG-3'
   GlyLeuSerMetAlaLysArgGluGluAlaGluAlaGluAlaGluProLysAlaThrArgPheValAsnGlnHisLeuCys
```

the P2-primer: 5'-AATTTATTTTACATAACACTAG-3' and the plasmid pKV142 according to the general scheme described above.

Example 11

Expression of N-terminal extended B_{chain}(1-29)-Ala-Ala-Arg-A_{chain}(1-21)

Yeast strain MT663 transformed with the C-POT plasmids pKV142 pKV143 and pKV102 and analyzed as described in example 5.

The results with the yeast strains expressing N-terminally extended B_{chain}(1-29)-Ala-Ala-Arg-A_{chain}(1-21) compared to the non-extended form (pKV142) are shown below

Plasmid	N-terminal extension	Yield
pKV142		100%
pKV143:	Asp-Asp-Ala-Asp-Ala-Asp-Ala-Asp-Pro-Arg	263%
pKV102:	Glu-Glu-Ala-Glu-Ala-Glu-Ala-Glu-Pro-Lys-Ala-Thr-Arg	400%

Example 12

Construction and expression of pIM69 and pIM70

Plasmid pAK679 is a derivative of pKFN1003 in which the EcoRI-XbaI fragment encodes the insulin precursor B_{chain}(1-29)-Ala-Ala-Lys-A_{chain}(1-21) N-terminally fused to a synthetic signal/leader sequence YAP3/LA19 (Figure 5).

Plasmid constructs designed to express N-terminally extended insulin precursor B_{chain}(1-29)-Ala-Ala-Lys-A_{chain}(1-21) were obtained by a procedure involving two successive PCR reaction. The first PCR reaction was performed by means of the primer with the following sequence

```
5'-GAAGAAGAAGAAGAAGAACCAAAGTTCGTTAACCAACAC-3'
   GluGluGluGluGluGluProLysPheValAsnGlnHis
```

the P2-primer 5'-AATTTATTTTACATAACACTAG-3' and the plasmid pAK679.

The second PCR reaction was performed by means of the P1-primer

```
                    NcoI
5'-GTTGTTAACTTGATCTCCATGGCTAAGAGAGAAGAA-3'
   ValValAsnLeuIleSerMetAlaLysArgGluGLu
```

the P2-primer 5'-AATTTATTTTACATAACACTAG-3' using the PCR product of the first PCR reaction as the DNA-template for the second PCR reaction.

The PCR product of the second PCR reaction was cut with NcoI and XbaI and ligated into pAK679 according to the general scheme described above. Among the resulting plasmids two where identified to encode N-terminal extensions of B_{chain}(1-29)-Ala-Ala-Lys-A_{chain}(1-21) in form of Glu-Glu-Glu-Pro-Lys (pIM70) and Glu-Glu-Glu-Glu-Pro-Lys (pIM69) respectively.

Yeast strain MT663 was transformed with the C-POT plasmids pAK579, pIM69 and pIM70 and analyzed as described in example 5.

Whereas the yield of non-extended B_{chain}(1-29)-Ala-Ala-Lys-A_{chain}(1-21) from yeast with pAK579 was found to be practicably nothing, yeast with pIM69 and pIM70 was found to produce large quantity of Glu-Glu-Glu-Glu-Pro-Lys-B_{chain}(1-29)-Ala-Ala-Lys-A_{chain}(1-21) and Glu-Glu-Glu-Pro-Lys-B_{chain}(1-29)-Ala-Ala-Lys-A_{chain}(1-21) respectively.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Novo Nordisk A/S
(B) STREET: Novo Alle
(C) CITY: Bagsvaerd
(E) COUNTRY: Denmark
(F) POSTAL CODE (ZIP): 2880
(G) TELEPHONE: +45 4444 8888
(H) TELEFAX: +45 4449 3256
(I) TELEX: 37304

(ii) TITLE OF INVENTION: N-Terminally Extended Proteins Expressed in Yeast

(iii) NUMBER OF SEQUENCES: 52

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Saccharomyces cerevisiae

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
Lys Arg Glu Ala Glu Ala
1               5
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Glu Glu Ala Glu Ala Glu Xaa Xaa Xaa Arg Ala Pro Arg
1               5                   10
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
Glu Glu Ala Glu Ala Glu Pro Lys Ala Xaa Arg
1               5                   10
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Glu Glu Ala Glu Ala Glu Ala Glu Pro Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Glu Glu Ala Glu Ala Glu Ala Glu Pro Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Glu Glu Ala Glu Ala Glu Ala Glu Arg
1 5

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Glu Glu Ala Glu Ala Glu Ala Xaa Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Glu Glu Ala Glu Ala Glu Ala Xaa Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Glu Glu Ala Glu Ala Glu Ala Arg
1 5

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Glu Glu Ala Glu Ala Glu Xaa Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Glu Glu Ala Glu Ala Pro Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Glu Glu Ala Pro Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Glu Glu Pro Lys
1

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

AATTTATTTT ACATAACACT AG 22

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

GGGGTATCCA TGGCTAAGAG AGAAGAAGCT GAAGCTGAAG CTNCAAAGTT CGTTAACCAA 60

CAC 63

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 64 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

GGGGTATCCA TGGCTAAGAG AGAAGAAGCT GAAGCTGAAG CTGNNAAAGT TCGTTAACCA 60

ACAC 64

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 61 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GGGGTATCCA TGGCTAAGAG AGAAGAAGCT GAAGCTGAAG GNAAAGTTCG TTAACCAACA 60

C 61

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 72 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

AACGTTGCCA TGGCTCCAGC TCCAGCTAAG AGAGAAGAAG CTGAAGCTGA AGCTGAAAGA 60

CATGCTGAAG GT 72

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

Glu Glu Ala Glu Ala Glu Ala Pro Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Glu Glu Ala Glu Ala Glu Ala Thr Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

Glu Glu Ala Glu Ala Glu Ala Glu Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

Glu Glu Ala Glu Ala Glu Ala Asp Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

Glu Glu Ala Glu Ala Glu Ala Gly Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Glu Glu Ala Glu Ala Glu Ala Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

Glu Glu Ala Glu Ala Glu Glu Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Glu Glu Ala Glu Ala Glu Ala Pro Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

Glu Glu Ala Glu Ala Glu Ala Glu Pro Lys Ala Thr Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

Glu Glu Ala Glu Ala Glu Ala Lys
1 5

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

```
Glu Glu Ala Pro Lys
1               5
```

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 660 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 127..540

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 127..381

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 382..540

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

```
GTTTGTATTC TTTTCTTGCT TAAATCTATA ACTACAAAAA ACACATACAG GAATTCCATT      60

CAAGAATAGT TCAAACAAGA AGATTACAAA CTATCAATTT CATACACAAT ATAAACGATT     120

AAAAGA ATG AGA TTT CCT TCA ATT TTT ACT GCA GTT TTA TTC GCA GCA        168
       Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala
       -85                 -80                 -75

TCC TCC GCA TTA GCT GCT CCA GTC AAC ACT ACA ACA GAA GAT GAA ACG       216
Ser Ser Ala Leu Ala Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr
    -70                 -65                 -60

GCA CAA ATT CCG GCT GAA GCT GTC ATC GGT TAC TCA GAT TTA GAA GGG       264
Ala Gln Ile Pro Ala Glu Ala Val Ile Gly Tyr Ser Asp Leu Glu Gly
-55                 -50                 -45                 -40

GAT TTC GAT GTT GCT GTT TTG CCA TTT TCC AAC AGC ACA AAT AAC GGG       312
Asp Phe Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly
                -35                 -30                 -25
```

TTA TTG TTT ATA AAT ACT ACT ATT GCC AGC ATT GCT GCT AAA GAA GAA 360
Leu Leu Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu
-20 -15 -10

GGG GTA TCC ATG GCT AAG AGA TTC GTT AAC CAA CAC TTG TGC GGT TCC 408
Gly Val Ser Met Ala Lys Arg Phe Val Asn Gln His Leu Cys Gly Ser
-5 1 5

CAC TTG GTT GAA GCT TTG TAC TTG GTT TGC GGT GAA AGA GGT TTC TTC 456
His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe
10 15 20 25

TAC ACT GAC AAG GCT GCT AAG GGT ATC GTT GAA CAA TGC TGT ACC TCC 504
Tyr Thr Asp Lys Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser
30 35 40

ATC TGC TCC TTG TAC CAA TTG GAA AAC TAC TGC AAC TAGACGCAGC 550
Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
45 50

CCGCAGGCTC TAGAAACTAA GATTAATATA ATTATATAAA AATATTATCT TCTTTTCTTT 610

ATATCTAGTG TTATGTAAAA TAAATTGATG ACTACGGAAA GCTAGCTTTT 660

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 138 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
-85 -80 -75 -70

Ala Leu Ala Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln
-65 -60 -55

Ile Pro Ala Glu Ala Val Ile Gly Tyr Ser Asp Leu Glu Gly Asp Phe
-50 -45 -40

Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu
-35 -30 -25

Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val
-20 -15 -10

Ser Met Ala Lys Arg Phe Val Asn Gln His Leu Cys Gly Ser His Leu
-5 1 5 10

```
Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr
            15                  20                  25

Asp Lys Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys
        30                  35                  40

Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
    45                  50
```

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 516 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthetic

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 133..411

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 133..321

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 322..411

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

```
TGTTTGTATT CTTTTTCTTGC TTAAATCTAT AACTACAAAA AACACATACA GGAATTCCAT       60

TCAAGAATAG TTCAAACAAG AAGATTACAA ACTATCAATT TCATACACAA TATAAACGAC      120

GGTACCAAAA TA ATG AAA CTG AAA ACT GTA AGA TCT GCG GTC CTT TCG          168
              Met Lys Leu Lys Thr Val Arg Ser Ala Val Leu Ser
              -63         -60                 -55

TCA CTC TTT GCA TCT CAG GTC CTT GGC CAA CCA ATT GAC GAC ACT GAA        216
Ser Leu Phe Ala Ser Gln Val Leu Gly Gln Pro Ile Asp Asp Thr Glu
    -50                 -45                 -40
```

```
TCT AAC ACT ACT TCT GTC AAC TTG ATG GCT GAC GAC ACT GAA TCT ATC      264
Ser Asn Thr Thr Ser Val Asn Leu Met Ala Asp Asp Thr Glu Ser Ile
-35                 -30                 -25                 -20

AAC ACT ACT TTG GTT AAC TTG GCT AAC GTT GCC ATG GCT CCA GCT CCA      312
Asn Thr Thr Leu Val Asn Leu Ala Asn Val Ala Met Ala Pro Ala Pro
                -15                 -10                 -5

GCT AAG AGA CAT GCT GAA GGT ACC TTC ACC TCT GAC GTC TCG AGT TAC      360
Ala Lys Arg His Ala Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr
            1               5                   10

TTG GAA GGC CAA GCT GCT AAG GAG TTC ATC GCT TGG TTG GTT AAG GGC      408
Leu Glu Gly Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly
    15                  20                  25

GCT TAGTCTAGAA ACTAAGATTA ATATAATTAT ATAAAAATAT TATCTTCTTT           461
Ala
30

TCTTTATATC TAGTGTTATG TAAAATAAAT TGATGACTAC GGAAAGCTAG CTTTT         516
```

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 93 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

```
Met Lys Leu Lys Thr Val Arg Ser Ala Val Leu Ser Ser Leu Phe Ala
-63         -60                 -55                 -50

Ser Gln Val Leu Gly Gln Pro Ile Asp Asp Thr Glu Ser Asn Thr Thr
        -45                 -40                 -35

Ser Val Asn Leu Met Ala Asp Asp Thr Glu Ser Ile Asn Thr Thr Leu
    -30                 -25                 -20

Val Asn Leu Ala Asn Val Ala Met Ala Pro Ala Pro Ala Lys Arg His
-15                 -10                 -5                    1

Ala Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Gly Gln
            5                   10                  15

Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Ala
        20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```
Asp Asp Ala Asp Ala Asp Ala Asp Pro Arg
1               5                   10
```

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```
Glu Glu Glu Glu Pro Lys
1               5
```

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Glu Glu Glu Pro Lys
1 5

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Asp Asp Asp Asp Asp Lys
1 5

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 61 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Glu Glu Ala Glu Ala Glu Ala Lys Phe Val Asn Gln His Leu Cys Gly
1 5 10 15

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
20 25 30

Phe Tyr Thr Pro Lys Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr
35 40 45

Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
50 55 60

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ala Ala Arg
20 25 30

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
35 40 45

Glu Asn Tyr Cys Asn
50

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Asp Asp Ala Asp Ala Asp Ala Asp Pro Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 74 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GGGGTATCCA TGGCTAAGAG AGACGACGCT GACGCTGACG CTGACCCAAG ATTCGTTAAC 60

CAACACTTGT GCGG 74

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Glu Glu Ala Glu Ala Glu Ala Glu Pro Lys Ala Thr Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 83 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GGGGTATCCA TGGCTAAGAG AGAAGAAGCT GAAGCTGAAG CTGAACCAAA GGCTACAAGA 60

TTCGTTAACC AACACTTGTG CGG 83

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys
20 25 30

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
35 40 45

Glu Asn Tyr Cys Asn
50

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GAAGAAGAAG AAGAAGAACC AAAGTTCGTT AACCAACAC 39

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GTTGTTAACT TGATCTCCAT GGCTAAGAGA GAAGAA 36

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 59 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Glu Glu Glu Glu Pro Lys Phe Val Asn Gln His Leu Cys Gly Ser His
1 5 10 15

Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr
20 25 30

Thr Pro Lys Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile
35 40 45

Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
50 55

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 58 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Glu Glu Glu Pro Lys Phe Val Asn Gln His Leu Cys Gly Ser His Leu
1 5 10 15

Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr
20 25 30

Pro Lys Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys
35 40 45

Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
50 55

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Glu Glu Ala Glu Ala Glu Ala Glu Pro Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 660 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GTTTGTATTC TTTTCTTGCT TAAATCTATA ACTACAAAAA ACACATACAG GAATTCCATT 60

CAAGAATAGT TCAAACAAGA AGATTACAAA CTATCAATTT CATACACAAT ATAAACGATT 120

AAAAGAATGA GATTTCCTTC AATTTTTACT GCAGTTTTAT TCGCAGCATC CTCCGCATTA 180

GCTGCTCCAG TCAACACTAC AACAGAAGAT GAAACGGCAC AAATTCCGGC TGAAGCTGTC 240

ATCGGTTACT CAGATTTAGA AGGGGATTTC GATGTTGCTG TTTTGCCATT TTCCAACAGC 300

ACAAATAACG GGTTTATTGTT TATAAATACT ACTATTGCCA GCATTGCTGC TAAAGAAGAA 360

GGGGTATCCA TGGCTAAGAG ATTCGTTAAC CAACACTTGT GCGGTTCCCA CTTGGTTGAA 420

GCTTTGTACT TGGTTTGCGG TGAAAGAGGT TTCTTCTACA CTCCTAAGGC TGCTAGAGGT 480

ATTGTCGAAC AATGCTGTAC CTCCATCTGC TCCTTGTACC AATTGGAAAA CTACTGCAAC 540

TAGACGCAGC CCGCAGGCTC TAGAAACTAA GATTAATATA ATTATATAAA AATATTATCT 600

TCTTTTCTTT ATATCTAGTG TTATGTAAAA TAAATTGATG ACTACGGAAA GCTAGCTTTT 660

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 600 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

TGTTTGTATT CTTTTCTTGC TTAAATCTAT AACTACAAAA AACACATACA GGAATTCCAT 60

TCAAGAATAG TTCAAACAAG AAGATTACAA ACTATCAATT TCATACACAA TATAAACGAC 120

GGTACCAAAA TAATGAAACT GAAAACTGTA AGATCTGCGG TCCTTTCGTC ACTCTTTGCA 180

TCTCAGGTCC TTGGCCAACC AATTGACGAC ACTGAATCTA ACACTACTTC TGTCAACTTG 240

ATGGCTGACG ACACTGAATC TAGATTCGCT ACTAACACTA CTTTGGCTTT GGATGTTGTT 300

AACTTGATCT CCATGGCTAA GAGATTCGTT AACCAACACT TGTGCGGTTC CCACTTGGTT 360

GAAGCTTTGT ACTTGGTTTG CGGTGAAAGA GGTTTCTTCT ACACTCCTAA GGCTGCTAAG 420

GGTATTGTCG AACAATGCTG TACCTCCATC TGCTCCTTGT ACCAATTGGA AAACTACTGC 480

AACTAGACGC AGCCCGCAGG CTCTAGAAAC TAAGATTAAT ATAATTATAT AAAAATATTA 540

TCTTCTTTTC TTTATATCTA GTGTTATGTA AAATAAATTG ATGACTACGG AAAGCTAGCT 600

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 55 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ser Asp Asp
20 25 30

Ala Arg Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr
35 40 45

Gln Leu Glu Asn Tyr Cys Asn
50 55

CLAIMS

1. A DNA construct encoding polypeptide having the following structure

signal peptide-leader peptide-X^1-X^2-X^3-X^4-X^5-X^6-X^7-heterologous protein

X^1 is Lys or Arg;

X^2 is Lys or Arg, X^1 and X^2 together defining a yeast processing site;

X^3 is Glu or Asp

X^4 is a sequence of amino acids with the following structure

$(A - B)_n$

wherein

A is Glu or Asp,

B is Ala, Val, Leu or Pro, and

n is 0 or an integer from 1 to 5, and when $n \geq 2$ each A and B is the same or different from the other A(s) and B(s), or

X^4 is a sequence of amino acids with the following structure

$(C)_m$

wherein

C is Glu or Asp, and

m is 0 or an integer from 1 to 5;

X^5 is a peptide bond or is one or more amino acids which may be the same or different;

X^6 is a peptide bond or an amino acid residue selected from the group consisting of Pro, Asp, Thr, Ser, Glu, Ala and Gly; and

X^7 is Lys or Arg.

2. A DNA construct according to claim 1, wherein X^3 is Glu.

3. A DNA construct according to claim 1, wherein A is Glu.

4. A DNA construct according to claim 1, wherein B is Ala

5. A DNA construct according to claim 1, wherein n is preferably 2-4, more preferably 3.

6. A DNA construct according to claim 1, wherein X^5 is a peptide bond, Glu, Asp or Glu-Pro-Lys-Ala.

7. A DNA construct according to claim 1, wherein X^6 is Pro, or a peptide bond.

8. A DNA construct according to claim 1, wherein the signal peptide is the α-factor signal peptide, the yeast aspartic protease 3 signal peptide, of mouse salivary amylase, the carboxypeptidase signal peptide, or the yeast BAR1 signal peptide.

9. A DNA construct according to claim 1, wherein the leader peptide is a natural leader peptide such as the α-factor leader peptide, or a synthetic leader peptide.

10. A DNA construct according to claim 1, wherein the heterologous protein is selected from the group consisting of aprotinin, tissue factor pathway inhibitor or other protease inhibitors, insulin-like growth factor I or II, human or bovine growth hormone, interleukin, tissue plasminogen activator, glucagon, glucagon-like peptide-1, Factor VII, Factor VIII, Factor XIII, platelet-derived growth factor, enzymes, insulin or an insulin precursor, and a functional analogue of any of these proteins.

11. A DNA construct according to claim 10, wherein the heterologous protein is insulin or an insulin precursor or a functional analogue thereof.

12. A DNA construct according to any of claims 1-11, wherein the amino acid sequence X^3-X^7 is

Glu Glu Ala Glu Ala Glu Ala Pro Lys,

Glu Glu Ala Glu Ala Glu Ala Glu Pro Lys Ala Thr Arg,

Glu Glu Ala Glu Ala Glu Ala Lys,
Glu Glu Ala Pro Lys,
Glu Glu Ala Glu Ala Glu Ala Glu Pro Lys,
Asp Asp Ala Asp Ala Asp Ala Asp Pro Arg,
Glu Glu Glu Glu Pro Lys,
Glu Glu Glu Pro Lys or
Asp Asp Asp Asp Asp Lys.

13. A recombinant expression vector which is capable of replicating in yeast and which carries a DNA construct according to claim 12.

14. A yeast strain which is capable of expressing a heterologous protein and which is transformed with a vector according to claim 13.

15. A process for producing a heterologous protein, comprising cultivating a yeast cell according to claim 14 in a suitable medium to obtain expression and secretion of the heterologous protein, after which the protein is isolated from the culture medium.

16. A process according to claim 15, wherein the amino acid sequence X^3-X^7 is removed from the recovered heterologous protein by a process involving treatment with a proteolytic enzyme which is specific for a basic amino acid.

17. A process according to claim 16, wherein the proteolytic enzyme is selected from the group consisting of trypsin, <u>Achromobacter lyticus</u> protease I, Enterokinase, <u>Fusarium oxysporum</u> trypsin-like protease, and YAP3.

18. A process according to claim 16, wherein the amino acid sequence X^3-X^7, wherein X^6 is a peptide bond, Ala, or Glu, is removed from the heterologous protein in the medium by a process involving subjecting the yeast cells to stress to make

them release yeast aspartic protease-3 into the medium, whereby the amino acid sequence X^3-X^7 is cleaved off.

19. A process according to claim 18, wherein the stress to which the yeast cell are subjected comprises reducing the pH of the culture medium to below 6.0, or starving the yeast cells by subjecting them to limiting growth conditions.

20. A process according to claim 19, wherein the pH of the culture medium is reduced to below 5.

21. A process according to claim 15, wherein the yeast cell is further transformed with one or more genes encoding a protease which is specific for basic amino acid residues so that, on cultivation of the cell the gene or genes are expressed, the consequent production of protease ensuring a more complete cleavage of X^3 - X^7 from the heterologous polypeptide.

22. A process according to claim 21, wherein the gene coding for the protease is a gene coding for YAP3 so that, on cultivation of the cell the YAP3 gene or genes are overexpressed, the consequent overproduction of YAP3 ensuring a more complete cleavage of X^3-X^7 from the heterologous polypeptide.

23. A process according to claim 21, wherein the gene coding for the protease codes for <u>A.lyticus</u> protease I or trypsin.

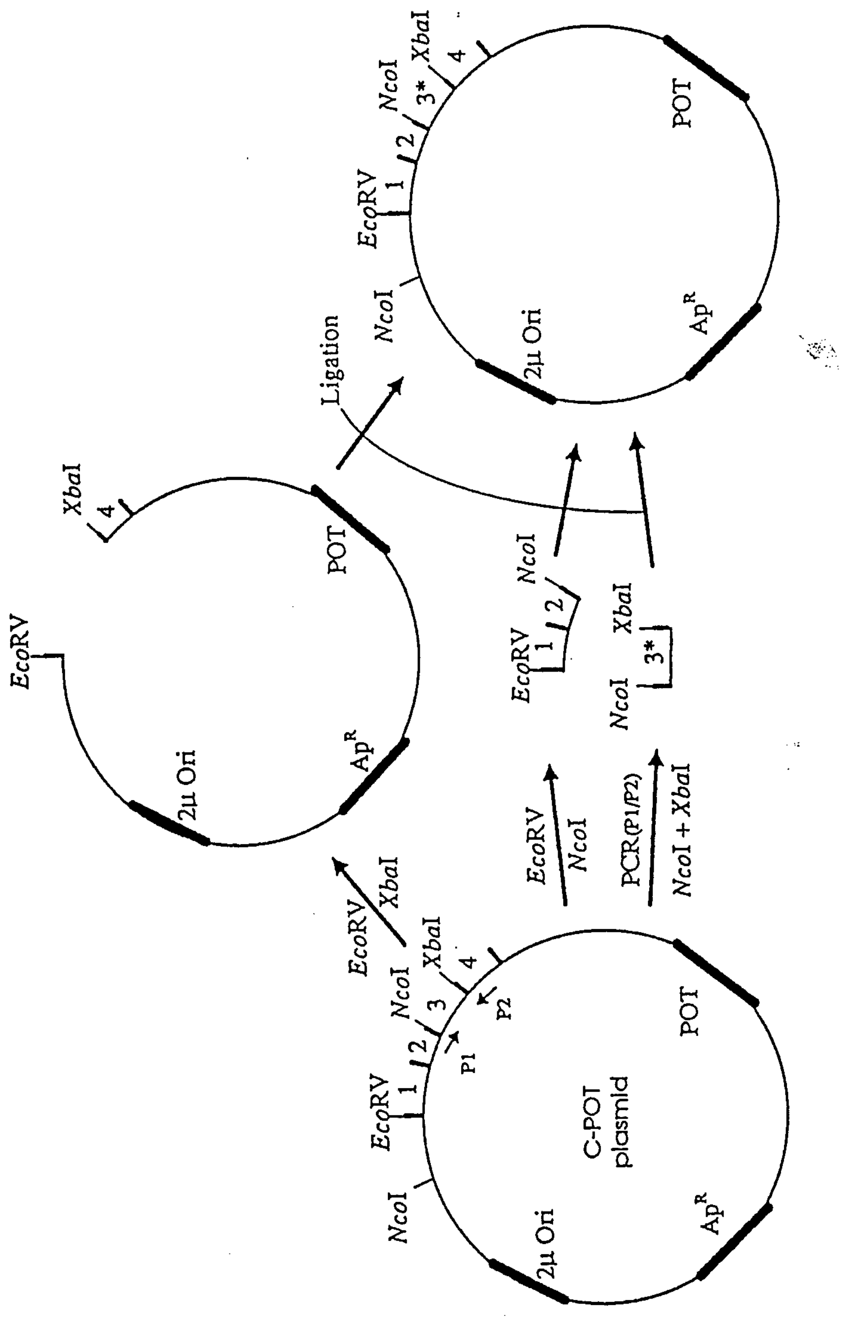

C-POT plasmid
EcoRV
NcoI
XbaI
1
2
3
3*
4
P1
P2
POT
Ap^R
2μ Ori
EcoRV XbaI
EcoRV NcoI
PCR(P1/P2) NcoI + XbaI
Ligation

Fig. 1

```
                                                             EcoRI
GTTTGTATTCTTTTCTTGCTTAAATCTATAACTACAAAAAACACATACAGGAATTCCATT
---------+---------+---------+---------+---------+---------+
CAAACATAAGAAAAGAACGAATTTAGATATTGATGTTTTTTGTGTATGTCCTTAAGGTAA

CAAGAATAGTTCAAACAAGAAGATTACAAACTATCAATTTCATACACAATATAAACGATT
---------+---------+---------+---------+---------+---------+
GTTCTTATCAAGTTTGTTCTTCTAATGTTTGATAGTTAAAGTATGTGTTATATTTGCTAA

AAAAGAATGAGATTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGCATCCTCCGCATTA
---------+---------+---------+---------+---------+---------+
TTTTCTTACTCTAAAGGAAGTTAAAAATGACGTCAAAATAAGCGTCGTAGGAGGCGTAAT
      MetArgPheProSerIlePheThrAlaValLeuPheAlaAlaSerSerAlaLeu
      <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<

GCTGCTCCAGTCAACACTACAACAGAAGATGAAACGGCACAAATTCCGGCTGAAGCTGTC
---------+---------+---------+---------+---------+---------+
CGACGAGGTCAGTTGTGATGTTGTCTTCTACTTTGCCGTGTTTAAGGCCGACTTCGACAG
AlaAlaProValAsnThrThrThrGluAspGluThrAlaGlnIleProAlaGluAlaVal
<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<

ATCGGTTACTCAGATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTTCCAACAGC
---------+---------+---------+---------+---------+---------+
TAGCCAATGAGTCTAAATCTTCCCCTAAAGCTACAACGACAAAACGGTAAAAGGTTGTCG
IleGlyTyrSerAspLeuGluGlyAspPheAspValAlaValLeuProPheSerAsnSer
<<<<<<<<<<α-factor-prepro leader sequence 1-81<<<<<<<<<<<<<<

ACAAATAACGGGTTATTGTTTATAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGAA
---------+---------+---------+---------+---------+---------+
TGTTTATTGCCCAATAACAAATATTTATGATGATAACGGTCGTAACGACGATTTCTTCTT
ThrAsnAsnGlyLeuLeuPheIleAsnThrThrIleAlaSerIleAlaAlaLysGluGlu
<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
       NcoI
GGGGTATCCATGGCTAAGAGATTCGTTAACCAACACTTGTGCGGTTCCCACTTGGTTGAA
---------+---------+---------+---------+---------+---------+
CCCCATAGGTACCGATTCTCTAAGCAATTGGTTGTGAACACGCCAAGGGTGAACCAACTT
GlyValSerMetAlaLysArgPheValAsnGlnHisLeuCysGlySerHisLeuValGlu
<<<<<<<<<           >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
HindIII
GCTTTGTACTTGGTTTGTGGTGAAAGAGGTTTCTTCTACACTGACAAGGCTGCTAAGGGT
```

Fig. 2a

```
---------+---------+---------+---------+---------+---------+
CGAAACATGAACCAAACACCACTTTCTCCAAAGAAGATGTGACTGTTCCGACGATTCCCA
AlaLeuTyrLeuValCysGlyGluArgGlyPhePheTyrThrAspLysAlaAlaLysGly
>>>>>>Insulin Precursor B1-27-Asp-Lys-Ala-Ala-Lys-A1-21>>>>>>>>

ATCGTTGAACAATGTTGTACTTCTATCTGTTCTTTGTACCAATTGGAAAACTACTGTAAC
---------+---------+---------+---------+---------+---------+
TAGCAACTTGTTACAACATGAAGATAGACAAGAAACATGGTTAACCTTTTGATGACATTG
IleValGluGlnCysCysThrSerIleCysSerLeuTyrGlnLeuGluAsnTyrCysAsn
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                    XbaI
TAGACGCAGCCCGCAGGCTCTAGAAACTAAGATTAATATAATTATATAAAAATATTATCT
---------+---------+---------+---------+---------+---------+
ATCTGCGTCGGGCGTCCGAGATCTTTGATTCTAATTATATTAATATATTTTTATAATAGA

TCTTTTCTTTATATCTAGTGTTATGTAAAATAAATTGATGACTACGGAAAGCTAGCTTTT
---------+---------+---------+---------+---------+---------+
AGAAAAGAAATATAGATCACAATACATTTTATTTAACTACTGATGCCTTTCGATCGAAAA
                            P2
```

Fig. 2b

```
                                                                    EcoRI
TGTTTGTATTCTTTTCTTGCTTAAATCTATAACTACAAAAAACACATACAGGAATTCCAT
---------+---------+---------+---------+---------+---------+
ACAAACATAAGAAAAGAACGAATTTAGATATTGATGTTTTTTGTGTATGTCCTTAAGGTA

TCAAGAATAGTTCAAACAAGAAGATTACAAACTATCAATTTCATACACAATATAAACGAC
---------+---------+---------+---------+---------+---------+
AGTTCTTATCAAGTTTGTTCTTCTAATGTTTGATAGTTAAAGTATGTGTTATATTTGCTG

GGTACCAAAATAATGAAACTGAAAACTGTAAGATCTGCGGTCCTTTCGTCACTCTTTGCA
---------+---------+---------+---------+---------+---------+
CCATGGTTTTATTACTTTGACTTTTGACATTCTAGACGCCAGGAAAGCAGTGAGAAACGT
             MetLysLeuLysThrValArgSerAlaValLeuSerSerLeuPheAla
             <<<<<<<<YAP3 (pre) signal peptid<<<<<<<<<<<<<<<<

TCTCAGGTCCTTGGCCAACCAATTGACGACACTGAATCTAACACTACTTCTGTCAACTTG
---------+---------+---------+---------+---------+---------+
AGAGTCCAGGAACCGGTTGGTTAACTGCTGTGACTTAGATTGTGATGAAGACAGTTGAAC
SerGlnValLeuGlyGlnProIleAspAspThrGluSerAsnThrThrSerValAsnLeu
<<<<<<<<<<<<<<<*********************************************
                                                        NcoI
ATGGCTGACGACACTGAATCTATCAACACTACTTTGGTTAACTTGGCTAACGTTGCCATG
---------+---------+---------+---------+---------+---------+
TACCGACTGCTGTGACTTAGATAGTTGTGATGAAACCAATTGAACCGATTGCAACGGTAC
MetAlaAspAspThrGluSerIleAsnThrThrLeuValAsnLeuAlaAsnValAlaMet
************S1PAVA(pro) leader peptide**********************

GCTCCAGCTCCAGCTAAGAGACATGCTGAAGGTACCTTCACCTCTGACGTCTCGAGTTAC
---------+---------+---------+---------+---------+---------+
CGAGGTCGAGGTCGATTCTCTGTACGACTTCCATGGAAGTGGAGACTGCAGAGCTCAATG
AlaProAlaValAlaLysArgHisAlaGluGlyThrPheThrSerAspValSerSerTyr
********************>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                        XbaI
TTGGAAGGCCAAGCTGCTAAGGAGTTCATCGCTTGGTTGGTTAAGGGCGCTTAGTCTAGA
---------+---------+---------+---------+---------+---------+
AACCTTCCGGTTCGACGATTCCTCAAGTAGCGAACCAACCAATTCCCGCGAATCAGATCT
LeuGluGlyGlnAlaAlaLysGluPheIleAlaTrpLeuValLysGlyAla
>>>>>>>>>>>>>>>>GLP-17-36Ala>>>>>>>>>>>>>>>>>>>>>>>>

AACTAAGATTAATATAATTATATAAAAATATTATCTTCTTTTCTTTATATCTAGTGTTAT
---------+---------+---------+---------+---------+---------+
TTGATTCTAATTATATTAATATATTTTTATAATAGAAGAAAAGAAATATAGATCACAATA

GTAAAATAAATTGATGACTACGGAAAGCTAGCTTTT
---------+---------+---------+------
CATTTTATTTAACTACTGATGCCTTTCGATCGAAAA
  P2
```

Fig. 3

```
                                                       EcoRI
GTTTGTATTCTTTTCTTGCTTAAATCTATAACTACAAAAAACACATACAGGAATTCCATT
---------+---------+---------+---------+---------+---------+
CAAACATAAGAAAAGAACGAATTTAGATATTGATGTTTTTTGTGTATGTCCTTAAGGTAA

CAAGAATAGTTCAAACAAGAAGATTACAAACTATCAATTTCATACACAATATAAACGATT
---------+---------+---------+---------+---------+---------+
GTTCTTATCAAGTTTGTTCTTCTAATGTTTGATAGTTAAAGTATGTGTTATATTTGCTAA

AAAAGAATGAGATTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGCATCCTCCGCATTA
---------+---------+---------+---------+---------+---------+
TTTTCTTACTCTAAAGGAAGTTAAAAATGACGTCAAAATAAGCGTCGTAGGAGGCGTAAT
      MetArgPheProSerIlePheThrAlaValLeuPheAlaAlaSerSerAlaLeu
      <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<

GCTGCTCCAGTCAACACTACAACAGAAGATGAAACGGCACAAATTCCGGCTGAAGCTGTC
---------+---------+---------+---------+---------+---------+
CGACGAGGTCAGTTGTGATGTTGTCTTCTACTTTGCCGTGTTTAAGGCCGACTTCGACAG
AlaAlaProValAsnThrThrThrGluAspGluThrAlaGlnIleProAlaGluAlaVal
<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<

ATCGGTTACTCAGATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTTCCAACAGC
---------+---------+---------+---------+---------+---------+
TAGCCAATGAGTCTAAATCTTCCCCTAAAGCTACAACGACAAAACGGTAAAAGGTTGTCG
IleGlyTyrSerAspLeuGluGlyAspPheAspValAlaValLeuProPheSerAsnSer
<<<<<<<<<<α-factor-prepro leader sequence 1-81<<<<<<<<<<<<<<

ACAAATAACGGGTTATTGTTTATAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGAA
---------+---------+---------+---------+---------+---------+
TGTTTATTGCCCAATAACAAATATTTATGATGATAACGGTCGTAACGACGATTTCTTCTT
ThrAsnAsnGlyLeuLeuPheIleAsnThrThrIleAlaSerIleAlaAlaLysGluGlu
<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
        NcoI
GGGGTATCCATGGCTAAGAGATTCGTTAACCAACACTTGTGCGGTTCCCACTTGGTTGAA
---------+---------+---------+---------+---------+---------+
CCCCATAGGTACCGATTCTCTAAGCAATTGGTTGTGAACACGCCAAGGGTGAACCAACTT
GlyValSerMetAlaLysArgPheValAsnGlnHisLeuCysGlySerHisLeuValGlu
<<<<<<<<<           >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
HindIII
GCTTTGTACTTGGTTTGCGGTGAAAGAGGTTTCTTCTACACTCCTAAGGCTGCTAGAGGT
---------+---------+---------+---------+---------+---------+
```

Fig. 4a

```
CGAAACATGAACCAAACGCCACTTTCTCCAAAGAAGATGTGAGGATTCCGACGATCTCCA
AlaLeuTyrLeuValCysGlyGluArgGlyPhePheTyrThrProLysAlaAlaArgGly
>>>>>>Insulin Precursor B1-29-Ala-Ala-Arg-A1-21>>>>>>>>>>>>>>>>

ATTGTCGAACAATGCTGTACCTCCATCTGCTCCTTGTACCAATTGGAAAACTACTGCAAC
---------+---------+---------+---------+---------+---------+
TAACAGCTTGTTACGACATGGAGGTAGACGAGGAACATGGTTAACCTTTTGATGACGTTG
IleValGluGlnCysCysThrSerIleCysSerLeuTyrGlnLeuGluAsnTyrCysAsn
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                  XbaI
TAGACGCAGCCCGCAGGCTCTAGAAACTAAGATTAATATAATTATATAAAAATATTATCT
---------+---------+---------+---------+---------+---------+
ATCTGCGTCGGGCGTCCGAGATCTTTGATTCTAATTATATTAATATATTTTTATAATAGA

TCTTTTCTTTATATCTAGTGTTATGTAAAATAAATTGATGACTACGGAAAGCTAGCTTTT
---------+---------+---------+---------+---------+-------
AGAAAAGAAATATAGATCACAATACATTTTATTTAACTACTGATGCCTTTCGATCGAAAA
                         P2
```

Fig. 4b

```
                                                                   EcoRI
TGTTTGTATTCTTTTCTTGCTTAAATCTATAACTACAAAAAACACATACAGGAATTCCAT
---------+---------+---------+---------+---------+---------+
ACAAACATAAGAAAAGAACGAATTTAGATATTGATGTTTTTTGTGTATGTCCTTAAGGTA

TCAAGAATAGTTCAAACAAGAAGATTACAAACTATCAATTTCATACACAATATAAACGAC
---------+---------+---------+---------+---------+---------+
AGTTCTTATCAAGTTTGTTCTTCTAATGTTTGATAGTTAAAGTATGTGTTATATTTGCTG

GGTACCAAAATAATGAAACTGAAAACTGTAAGATCTGCGGTCCTTTCGTCACTCTTTGCA
---------+---------+---------+---------+---------+---------+
CCATGGTTTTATTACTTTGACTTTTGACATTCTAGACGCCAGGAAAGCAGTGAGAAACGT
                  MetLysLeuLysThrValArgSerAlaValLeuSerSerLeuPheAla
                  <<<<<<<<YAP3 (pre) signal peptid<<<<<<<<<<<<<<<<

TCTCAGGTCCTTGGCCAACCAATTGACGACACTGAATCTAACACTACTTCTGTCAACTTG
---------+---------+---------+---------+---------+---------+
AGAGTCCAGGAACCGGTTGGTTAACTGCTGTGACTTAGATTGTGATGAAGACAGTTGAAC
SerGlnValLeuGlyGlnProIleAspAspThrGluSerAsnThrThrSerValAsnLeu
<<<<<<<<<<<<<<***********************************************

                  XbaI
ATGGCTGACGACACTGAATCTAGATTCGCTACTAACACTACTTTGGCTTTGGATGTTGTT
---------+---------+---------+---------+---------+---------+
TACCGACTGCTGTGACTTAGATCTAAGCGATGATTGTGATGAAACCGAAACCTACAACAA
MetAlaAspAspThrGluSerArgPheAlaThrAsnThrThrLeuAlaLeuAspValVal
************LA19 (pro) leader peptide***********************

          NcoI
AACTTGATCTCCATGGCTAAGAGATTCGTTAACCAACACTTGTGCGGTTCCCACTTGGTT
---------+---------+---------+---------+---------+---------+
TTGAACGAGAGGTACCGATTCTCTAAGCAATTGGTTGTGAACACGCCAAGGGTGAACCAA
AsnLeuIleSerMetAlaLysArgPheValAsnGlnHisLeuCysGlySerHisLeuVal
************************>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

HindIII
GAAGCTTTGTACTTGGTTTGCGGTGAAAGAGGTTTCTTCTACACTCCTAAGGCTGCTAAG
---------+---------+---------+---------+---------+---------+
CTTCGAAACATGAACCAAACGCCACTTTCTCCAAAGAAGATGTGAGGATTCCGACGATTC
GluAlaLeuTyrLeuValCysGlyGluArgGlyPhePheTyrThrProLysAlaAlaLys
>>>>>>>>>Insulin Precursor B1-29-Ala-Ala-Lys-A1-21>>>>>>>>>>>>

GGTATTGTCGAACAATGCTGTACCTCCATCTGCTCCTTGTACCAATTGGAAAACTACTGC
---------+---------+---------+---------+---------+---------+
```

Fig. 5a

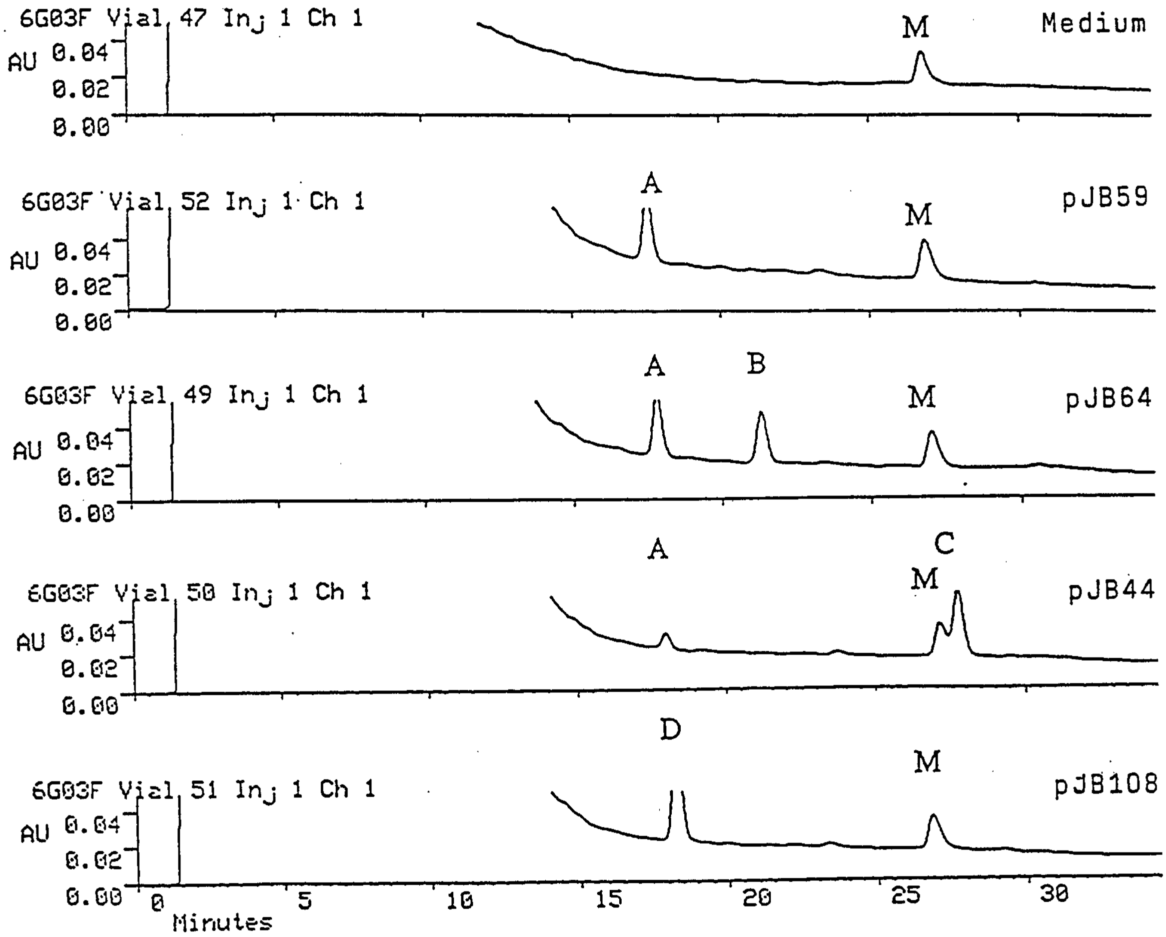
```
CCATAACAGCTTGTTACGACATGGAGGTAGACGAGGAACATGGTTAACCTTTTGATGACG
GlyIleValGluGlnCysCysThrSerIleCysSerLeuTyrGlnLeuGluAsnTyrCys
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                        XbaI
AACTAGACGCAGCCCGCAGGCTCTAGAAACTAAGATTAATATAATTATATAAAAATATTA
---------+---------+---------+---------+---------+---------+
TTGATCTGCGTCGGGCGTCCGAGATCTTTGATTCTAATTATATTAATATATTTTTATAAT
Asn
>>>

TCTTCTTTTCTTTATATCTAGTGTTATGTAAAATAAATTGATGACTACGGAAAGCTAGCT
---------+---------+---------+---------+---------+---------+
AGAAGAAAAGAAATATAGATCACAATACATTTTATTTAACTACTGATGCCTTTCGATCGA
                         P2
```

Fig. 5b

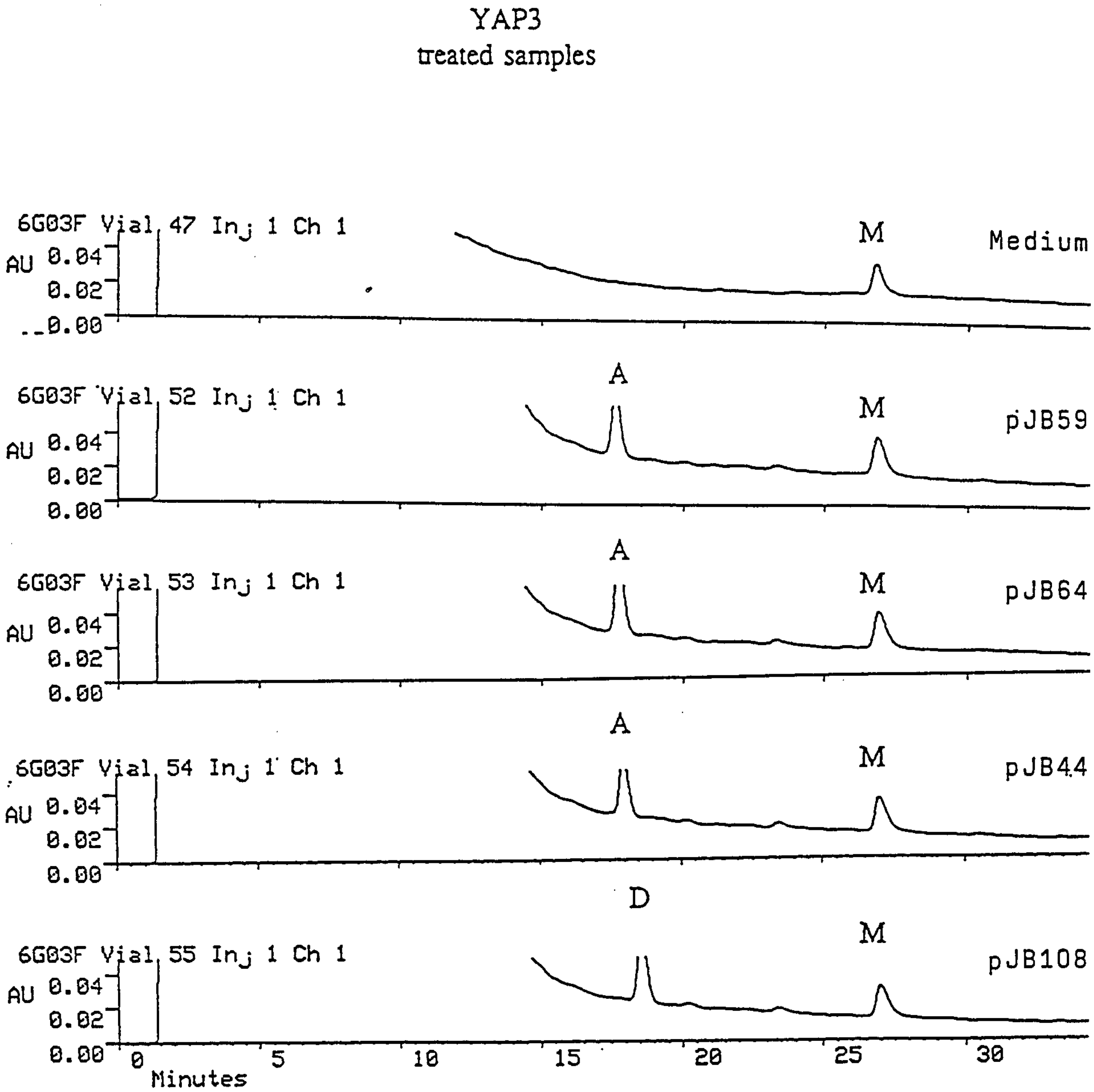
HPLC chromatograms

Control fermentations
untreated samples



Peak M:	Background profile on yeast growth media
Peak A:	B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21)
Peak B:	Glu-Glu-Ala-Glu-Ala-Glu-Ala-Lys-B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21)
Peak C:	Glu-Glu-Ala-Glu-Ala-Glu-Ala-Glu-Lys-B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21)
Peak D:	Glu-Glu-Ala-Glu-Ala-Glu-Ala-Pro-Lys-B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21)

Fig. 6a



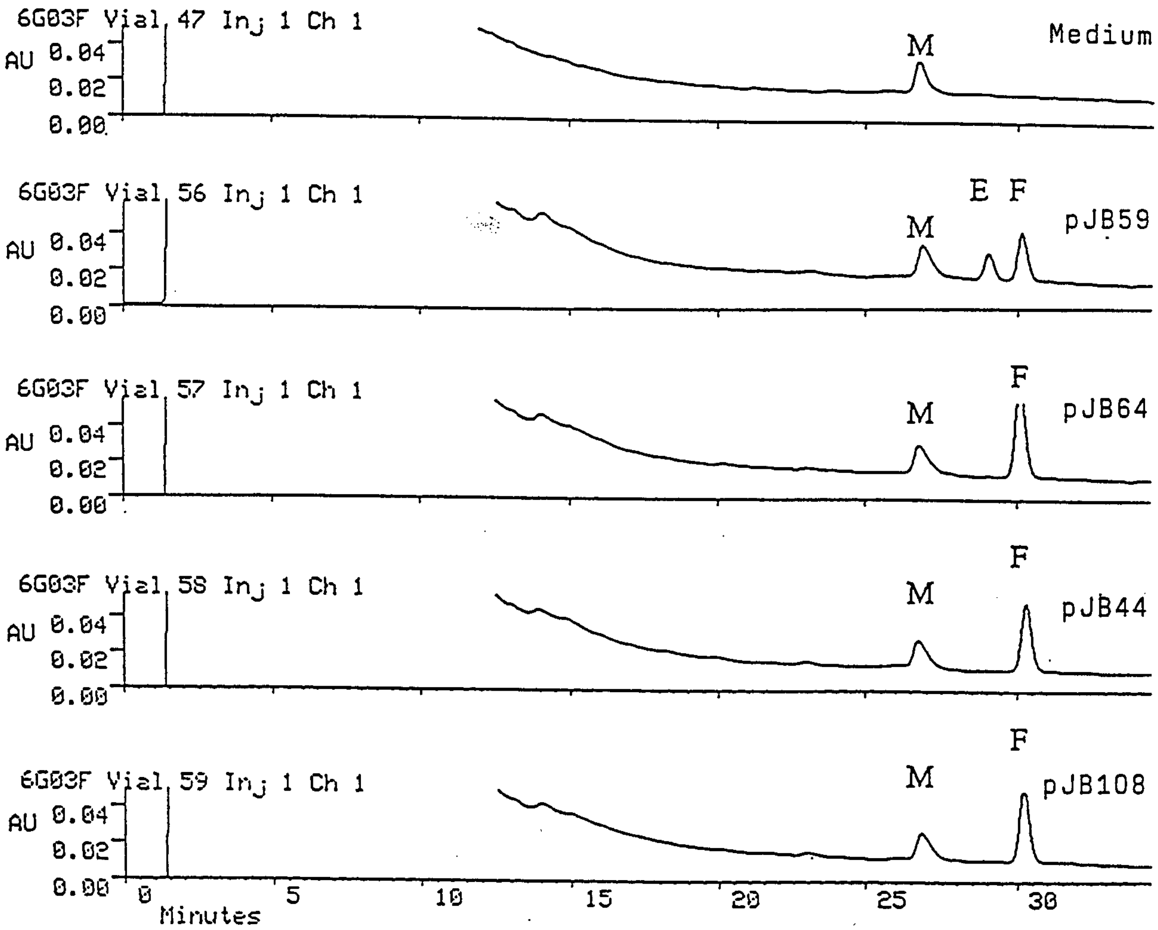
Peak M: Background profile on yeast growth media

Peak A: B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21)

Peak D: Glu-Glu-Ala-Glu-Ala-Glu-Ala-Pro-Lys-B_{chain}(1-27)-Asp-Lys-Ala-Ala-Lys-A_{chain}(1-21)

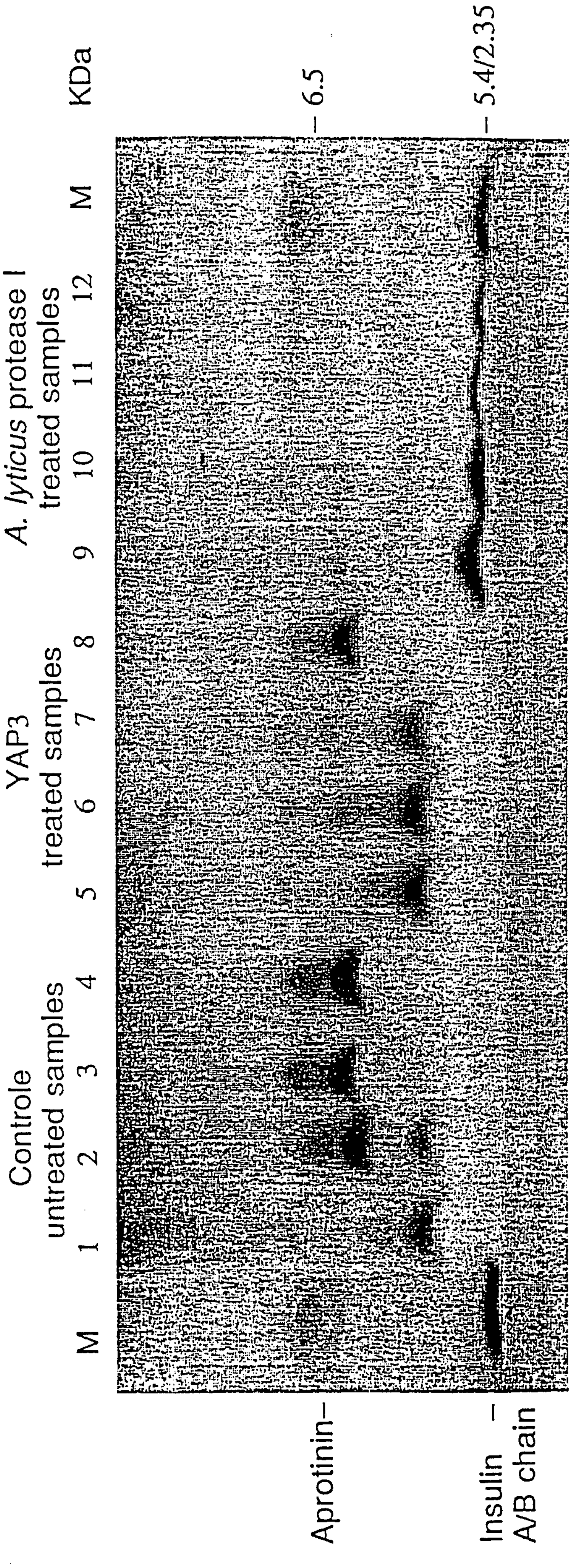
Fig. 6b

A. lyticus protease I
treated samples



Peak M: Background profile on yeast growth media
Peak E: Arg-B_{chain}(1-27)-Asp-Lys- (connected by disulfide bonds) A_{chain}(1-21)
Peak F: B_{chain}(1-27)-Asp-Lys- (connected by disulfide bonds) A_{chain}(1-21)

Fig. 6c



10% Tricine–SDS–PAGE

Fig. 7

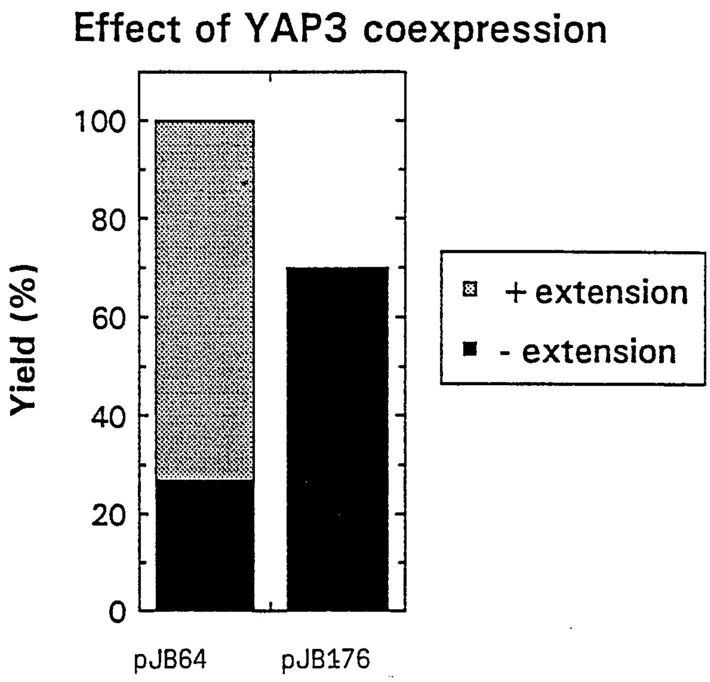

Effect of YAP3 coexpression
Yield (%)
100
80
60
40
20
0
pJB64
pJB176
+ extension
- extension

Fig. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/DK 95/00250

A. CLASSIFICATION OF SUBJECT MATTER

IPC6: C12N 15/62, C12N 15/81, C12N 15/11
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC6: C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, CLAIMS, EDOC, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 8606406 A1 (TRANSGENE S.A.), 6 November 1986 (06.11.86), page 11, line 7; page 13, line 14 - line 15 --	1-23
X	WO 9010075 A1 (NOVO-NORDISK A/S), 7 Sept 1990 (07.09.90), see the whole document especially the claims --	1-23
X	WO 9213951 A1 (THE SALK INSTITUTE BIOTECHNOLOGY/INDUSTRIAL ASSOCIATES, INC.), 20 August 1992 (20.08.92), see the whole document especially page 23, line 26 - line 32 --	1-23

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents:
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" erlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 November 1995	9.11.95
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Patrick Andersson Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

International application No.
PCT/DK 95/00250

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0301669 A1 (GIST-BROCADES N.V.), 1 February 1989 (01.02.89), page 3, line 41 - page 4, line 3	1-23
	--	
X	EP 0324274 A1 (CHIRON CORPORATION), 19 July 1989 (19.07.89), see the whole document	1-23
	--	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/DK 95/00250

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see next sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

According to rule 13.2, an international application shall relate to one invention only or a group of inventions linked by one or more of the same or corresponding "special technical features", i.e. features that define a contribution which each of the inventions makes over prior art.
Such a link between all the subject of claims 1-23 would be any DNA construct encoding a peptide as defined in Claim 1 . This *a priori* allegation however is not acceptable due to the state of the art as revealed in the attached search report, e.g. WO86/06406 (see page 11, line 7 and page 13 line 14 -15)

Accordingly, the following inventions were found:

Invention 1: DNA construct encoding a polypeptide comprising signal peptide-leader peptide-x1-x2-x3-x4-x5-x6-x7 which at least fulfils the minimum requirements of claim 1.

Invention 2, claim 12-23, partially:DNA construct encoding a polypeptide comprising signal peptide-leader peptide-x1-x2-x3-x4-x5-x6-x7, where x3-x7 has the following amino acid sequence: EEAEAEAPK

Invention 3, claim 12-23, partially:DNA construct encoding a polypeptide comprising signal peptide-leader peptide-x1-x2-x3-x4-x5-x6-x7, where x3-x7 has the following amino acid sequence: EEAEAEAEPKATR

Invention 4, claim 12-23, partially:DNA construct encoding a polypeptide comprising signal peptide-leader peptide-x1-x2-x3-x4-x5-x6-x7, where x3-x7 has the following amino acid sequence: EEAEAEAK

Invention 5, claim 12-23, partially:DNA construct encoding a polypeptide comprising signal peptide-leader peptide-x1-x2-x3-x4-x5-x6-x7, where x3-x7 has the following amino acid sequence: EEAPK

Invention 6, claim 12-23, partially:DNA construct encoding a polypeptide comprising signal peptide-leader peptide-x1-x2-x3-x4-x5-x6-x7, where x3-x7 has the following amino acid sequence: EEAEAEAEPK

Invention 7, claim 12-23, partially:DNA construct encoding a polypeptide comprising signal peptide-leader peptide-x1-x2-x3-x4-x5-x6-x7, where x3-x7 has the following amino acid sequence: DDADADADPR

Invention 8, claim 12-23, partially:DNA construct encoding a polypeptide comprising signal peptide-leader peptide-x1-x2-x3-x4-x5-x6-x7, where x3-x7 has the following amino acid sequence: EEEEPK

Invention 9, claim 12-23, partially:DNA construct encoding a polypeptide comprising signal peptide-leader peptide-x1-x2-x3-x4-x5-x6-x7, where x3-x7 has the following amino acid sequence: EEEPK

Invention 10, claim 12-23, partially:DNA construct encoding a polypeptide comprising signal peptide-leader peptide-x1-x2-x3-x4-x5-x6-x7, where x3-x7 has the following amino acid sequence: DDDDDK

INTERNATIONAL SEARCH REPORT
Information on patent family members

02/10/95

International application No.
PCT/DK 95/00250

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A1- 8606406	06/11/86	AT-T-	121778	15/05/95
		AU-B,B-	600667	23/08/90
		AU-A-	5778786	18/11/86
		DE-D-	3650306	00/00/00
		EP-A,B-	0200655	05/11/86
		FR-A,B-	2593518	31/07/87
		FR-A,A-	2601383	15/01/88
		JP-T-	62502661	15/10/87
		SU-A-	1774950	07/11/92
WO-A1- 9010075	07/09/90	AT-T-	110414	15/09/94
		AU-B,B-	624694	18/06/92
		AU-A-	5261290	26/09/90
		CA-A-	2050336	04/09/90
		DE-D,T-	69011853	15/12/94
		EP-A,B-	0461165	18/12/91
		SE-T3-	0461165	
		ES-T-	2062514	16/12/94
		JP-T-	4504846	27/08/92
		PL-B-	163532	29/04/94
		US-A-	5395922	07/03/95
WO-A1- 9213951	20/08/92	NONE		
EP-A1- 0301669	01/02/89	SE-T3-	0301669	
		AU-B,B-	623860	28/05/92
		AU-A-	2013688	02/02/89
		CA-A-	1327762	15/03/94
		CN-A-	1032367	12/04/89
		DE-A,T-	3881776	22/07/93
		ES-T-	2058238	01/11/94
		IE-B-	62261	11/01/95
		JP-A-	1124390	17/05/89
		PT-B-	88115	01/03/95
		US-A-	5010182	23/04/91
		US-A-	5217891	08/06/93
EP-A1- 0324274	19/07/89	SE-T3-	0324274	
		AU-A-	2765088	06/07/89
		DE-D,T-	3885728	10/03/94
		ES-T-	2059547	16/11/94
		HK-A-	139994	16/12/94
		IE-B-	62087	14/12/94
		JP-A-	2002339	08/01/90