The invention is also related to DNA sequences, vectors and host cells for use in expressing polypeptides in yeasts.

(54) Title: MATING FACTOR ALPHA PRO-PEPTIDE VARIANTS

(57) Abstract: The present invention is related to Mating Factor a pro-peptide variants useful for the recombinant expression of polypeptides comprising a GLP-1 peptide in yeasts. The invention is also related to DNA sequences, vectors and host cells for use in expressing polypeptides in yeasts.
MATING FACTOR ALPHA PRO-PEPTIDE VARIANTS

TECHNICAL FIELD

The present invention relates to the technical fields of protein expression and protein chemistry where a polypeptide is prepared by recombinant expression in yeast.

BACKGROUND

The techniques of recombinant polypeptide expression allow for the production of large quantities of desirable polypeptides which may be used for e.g. their biological activity. Such polypeptides are often expressed as recombinant fusion polypeptides in microbial host cells. The polypeptide of interest is often attached to a fusion partner polypeptide in order to increase the expression level, facilitate secretion, increase the solubility, promote polypeptide folding, to protect the polypeptide against unintentional proteolysis or to facilitate purification of the polypeptide of interest.

To ensure secretion of recombinantly expressed polypeptides from yeast, a pre-pro peptide (often called “leader”) is normally fused to the N-terminus of the recombinant product. The pre-sequence ensures translocation of the fusion protein into the endoplasmic reticulum (ER), which is the starting point of the secretory pathway. The pro-sequence ensures further transport from the ER to the Golgi apparatus, where an endogenous protease called Kex2p, often are used to cleave off the pro-sequence. The processed recombinant peptide is subsequently secreted to the growth media, from where it can be purified.

The pre-pro sequence from Mating Factor Alpha is often used as leader when secreting recombinant polypeptides. However, many other sequences are able to facilitate the secretory process. The leader sequence has a significant influence, not only on the amounts of secreted peptide, but also on the quality in terms of degradation and post-translational modifications such as O-glycosylation. As degradation and O-glycosylation normally are unwanted events, it is desirable with a leader sequence that reduces these modifications to the lowest possible extent, and at the same time maximizes the yield of secreted polypeptide.


Thim et al. (PNAS 83 (1986) 6766-6770) describe the use of Mating Factor Alpha for secretion and processing of insulin precursors in Saccharomyces cerevisiae.
W095/34666 describes synthetic leaders for producing secreted polypeptides in S. cerevisiae.

Rakestraw et al. (Biotech. Bioeng. 103 (2009) 1192-1201) describe mutant Mating Factor Alpha leader sequences which increase the secretion of single-chain antibody and which increase human IgG1 production levels in S. cerevisiae.

There is a need for more specific leader sequences which increase the yield of the polypeptide precursor, and which reduce the proportion of the recombinant polypeptide which is O-glycosylated. In particular there is a need for leader sequences for expressing GLP-1 peptides in yeast, which leaders increase the yield of the GLP-1 peptides or precursors thereof and lower the O-glycosylation of the GLP-1 peptides. Such more specific leader sequences may facilitate a higher yield of the recombinant polypeptide as well as a lower the amount of O-glycosylated impurities.

SUMMARY

It is an object of the present invention to provide yeast cells having increased level of expression of heterologous polypeptides. It is also an object of the present invention to provide yeast cells secreting recombinant polypeptide having a reduced amount of O-glycosylated variants of the recombinant polypeptide. In particular it is an object of the present invention to provide yeast cells having both an increased level of expression of the recombinant polypeptide and a reduced amount of O-glycosylated variants. It is an object of the present invention to provide improved expression system for recombinant expression of GLP-1 peptides in yeast cells.

According to a first aspect of the invention there is provided a method for recombinant expression of a polypeptide comprising a GLP-1 peptide in yeast comprising the culturing of a yeast strain comprising a DNA sequence encoding a processing and secretion signal upstream of the polypeptide, wherein said processing and secretion signal comprises a Mating Factor a pro-peptide variant having at least one substitution in the VIGYL sequence at positions 38-42 to comprise the amino acid sequence of the general formula (I):

\[
X_{38}X_{39}X_{40}X_{41}X_{42} \quad (SEQ \ ID \ NO:1) \quad (I)
\]

wherein

\[
X_{38} \text{ is } F, \ L, \ I \text{ or } V; \\
X_{39} \text{ is } L, \ I, \ V \text{ or } M; \\
X_{40} \text{ is } A, \ G, \ S, \ E, \ Q, \ Y, \ F, \ W, \ R, \ K, \ H, \ L, \ I, \ V \text{ or } M; \\
X_{41} \text{ is } S, \ Y, \ F, \ W, \ L, \ I, \ V \text{ or } M; \\
X_{42} \text{ is } Y, \ W, \ L, \ I, \ V, \ M \text{ or } S;
\]
with the proviso that X38-X42 is not VIGYS.

In one embodiment of the method for recombinant expression of a polypeptide in yeast the Mating Factor a pro-peptide variant has at least one substitution in the VIGYL sequence at positions 38-42 to comprise the amino acid sequence of the general formula (I):

\[ X_{38}-X_{39}-X_{40}-X_{41}-X_{42} \]

wherein \( X_{38} \) is V; \( X_{39} \) L, I, V or M; \( X_{40} \) is G or R; \( X_{41} \) is Y, and \( X_{42} \) is L.

In another embodiment the polypeptide for recombinant expression is a GLP-1 peptide.

In another embodiment the polypeptide for recombinant expression comprises GLP-1(7-37)[K34R], GLP-1(9-37)[K34R] or GLP-1(9-37)[K34R,G37K].

According to a second aspect of the invention there is provided a Mating Factor a pro-peptide variant having at least one substitution in the VIGYL sequence at positions 38-42 to comprise the amino acid sequence of the general formula (I):  

\[ X_{38}-X_{39}-X_{40}-X_{41}-X_{42} \]  

(SEQ ID NO:1)

wherein

\( X_{38} \) is F, L, I or V;
\( X_{39} \) L, I, V or M;
\( X_{40} \) is A, G, S, E, Q, Y, F, W, R, K, H, L, I, V or M;
\( X_{41} \) is S, Y, F, W, L, I, V or M;
\( X_{42} \) is Y, W, L, I, V, M or S;

with the proviso that X38-X42 is not VIGYS.

According to a third aspect of the invention there is provided a GLP-1 precursor which is a fusion polypeptide comprising:

- A pre-peptide,
- A Mating Factor a pro-peptide variant having at least one substitution in the VIGYL sequence at positions 38-42 to comprise the amino acid sequence of the general formula (I):

\[ X_{38}-X_{39}-X_{40}-X_{41}-X_{42} \]  

with the proviso that X38-X42 is not VIGYS,
<Optionally a extension peptide, and
- A GLP-1 polypeptide.

According to a fourth aspect of the invention there is provided a DNA sequence encoding the Mating Factor a pro-peptide variant or the GLP-1 precursor.

According to a fifth aspect of the invention there is provided an expression vector comprising the DNA sequence encoding the Mating Factor a pro-peptide variant or the GLP-1 precursor.

According to a sixth aspect of the invention there is provided a host cell comprising the expression vector according to the invention.

In one embodiment the host cell used for expression has a non-functional *pmt1* gene or no *pmt1* gene at all. Such a host cell has surprisingly been found to lower the amount of O-glycosylated polypeptide independently of, and in addition to, the lowering obtained from the Mating Factor a pro-peptide variant of the present invention.

**BRIEF DESCRIPTION OF DRAWINGS**

Figure 1 shows the minimal expression plasmid as used in Example 1.

Figure 2 shows the N-terminally extended GLP-1 (7-37)[K34R] precursor including the wildtype Mating Factor a pre-pro-peptide with indication of the VIGYL subsequence of the pro-peptide.

**DESCRIPTION**

According to a first aspect of the invention there is provided a method for recombinant expression of a polypeptide in yeast comprising the culturing of a yeast strain comprising a DNA sequence encoding a processing and secretion signal upstream of the polypeptide, wherein said processing and secretion signal comprises a Mating Factor a pro-peptide variant having at least one substitution in the VIGYL sequence at positions 38-42 to comprise the amino acid sequence of the general formula (I):

\[
X_{38} - X_{39} - X_{40} - X_{41} - X_{42} \quad \text{(SEQ ID NO:1)}
\]

wherein

\[
X_{38} \text{ is } F, L, I \text{ or } V;
\]
\[
X_{39} \text{ is } L, I, V \text{ or } M;
\]
\[
X_{40} \text{ is } A, G, S, E, Q, Y, F, W, R, K, H, L, I, V \text{ or } M;
\]
\[
X_{41} \text{ is } S, Y, F, W, L, I, V \text{ or } M;
\]
\[
X_{42} \text{ is } Y, W, L, I, V, M \text{ or } S;
\]
with the proviso that \( X_{38-X_{42}} \) is not VIGYS.

The term "leader sequence" as used herein is intended to mean an amino acid sequence consisting of a pre-peptide (the signal peptide) and a pro-peptide. Non-limiting examples of leader sequences are e.g. the alpha-factor signal leader from \( S. \) cerevisiae and the synthetic leader sequences for yeast described in W095/34666.

"Pre-peptide" as used herein is intended to mean a signal peptide which is present as an \( N \)-terminal sequence on the precursor form of a polypeptide to be expressed. The function of the signal peptide is to facilitate translocation of the polypeptide into the endoplasmic reticulum in the host cell. The signal peptide is normally cleaved off in the course of this process. The signal peptide may be heterologous or homologous to the host cell producing the polypeptide.

"Pro-peptide" as used herein is intended to mean a peptide sequence whose function is to allow the expressed polypeptide to be directed from the endoplasmic reticulum to the Golgi apparatus and further to a secretory vesicle for secretion into the culture medium (i.e. exportation of the polypeptide across the cell wall or at least through the cellular membrane into the periplasmic space of the yeast cell). Non-limiting examples of a pro-peptide are the yeast \( \alpha \)-factor pro-peptide (vide US 4,546,082 and 4,870,008) and the synthetic pro-peptides disclosed in US 5,395,922; 5,795,746; 5,162,498 and WO 98/32867. The pro-peptide will preferably contain an endopeptidase processing site at the C-terminal end, such as a Lys-Arg sequence or any functional analog thereof.

The term "Mating Factor a" (MFa, MFa or MFalpha) as used herein is intended to mean the \( S. \) cerevisiae prepro-sequence, comprising the Mating Factor a pre-peptide as amino acid residues 1-19 and Mating Factor a pro-peptide as amino acid residues 20-85 in the structure:

\[
\text{MRFPSIFTAVLFAASSALAAPVNTTEDAQIPAEGVIGYLDLEGDFDVAPVFSNSTNNGLL}
\]
\[
\text{FINITIASIAAKEEGVSMDKR} \quad \text{(SEQ ID NO:2).}
\]

The Mating Factor a comprises the sequence VIGYL (SEQ ID NO:3) as amino acid residues 38-42. A variant of Mating Factor a has been used for recombinant expression of polypeptides which comprises the sequence VIGYS (SEQ ID NO:4) as amino acid residues 38-42 in the Mating Factor a sequence.

In the present context the terms "polypeptide", "protein" and "peptide" may be used interchangeably to designate a polypeptide. It is to be understood that the particular term used has no limitation as to the size of the molecule (unless directly stated in the particular context).
Amino acid residues are generally designated according to single letter abbreviation according to IUPAC nomenclature, e.g. D meaning aspartic acid (Asp) and G meaning glycine. However, in some instances the corresponding three letter abbreviation is also used.

"Genetically encoded amino acids" as used herein is intended to mean the the group consisting of the following amino acids: G, P, A, V, L, I, M, C, F, Y, W, H, K, R, Q, N, E, D, S, T as well as any biological modification hereof. Non-limiting examples of such biological modifications are e.g. amidation, glycosylation and disulphide bond formation.

"Analogues" as used herein is intended to mean polypeptides which are derived from the reference polypeptide by means of substitution, deletion and/or addition of one or more amino acid residues from the polypeptide. Non-limiting examples of an analogue of GLP-1(7-37) (SEQ ID NO:5) are GLP-1(7-37)[K34R] (SEQ ID NO:6) where residue 34 has been substituted by an arginine residue and GLP-1(9-37)[K34R] (SEQ ID NO:7) where residue 34 has been substituted with an arginine residue and amino acid residues 7-8 have been deleted (using the common numbering of amino acid residues for GLP-1 peptides).

"Variant" as used herein with reference to a polypeptide is intended to mean a chemical variant of the polypeptide which retains substantially the same main function as the original protein. Hence a variant is typically a modified version of a polypeptide wherein as few modifications are introduced as necessary for the modified polypeptide to have some desirable property while preserving substantially the same main function of the original polypeptide. Non-limiting examples of polypeptide variants are e.g. extended polypeptides, truncated polypeptides, fusion polypeptides and analogues. A non-limiting example of a variant of Mating Factor a pro-peptide is L42S-Mating Factor a(20-85) (SEQ ID NO:8). A non-limiting example of a variant of GLP-1 (7-37) is GLP-1 (7-37)[K34R].

In one embodiment, a variant of a polypeptide comprises from 1-2 amino acid substitutions, deletions or additions as compared to the unmodified polypeptide. In another embodiment, a variant comprises from 1-5 amino acid substitutions, deletions or additions as compared to the unmodified polypeptide. In another embodiment, a variant comprises from 1-15 amino acid substitution, deletion or additions relative to the corresponding unmodified polypeptide.

According to a second aspect of the invention there is provided a Mating Factor a pro-peptide variant having at least one substitution in the VIGYL sequence at positions 38-42 to comprise the amino acid sequence of the general formula (I):

\[ X_{38} \cdot X_{39} \cdot X_{40} \cdot X_{41} \cdot X_{42} \]  

(SEQ ID NO:1)  

(I)

wherein

\[ X_{38} \] is F, L, I or V;
with the proviso that X38-X42 is not VIGYS. In one embodiment the Mating Factor a pro-
peptide variant does not comprise VIGYS, VIDYS, VATYL, VIGYR, or AIGYL as X38-X42.
According to a third aspect of the invention there is provided a GLP-1 precursor which
is a fusion polypeptide comprising:
- A pre-peptide,
- A Mating Factor a pro-peptide variant having at least one substitution in the VIGYL
sequence at positions 38-42 to comprise the amino acid sequence of the general
formula (I):

\[
X38-X39-X40-X41-X42 \quad (\text{SEQ ID NO:1}) \quad (I)
\]

wherein

\[
\begin{align*}
X_{38} & \text{ is F, L, I or V;} \\
X_{39} & \text{ is L, I, V or M;} \\
x_{40} & \text{ is A, G, S, E, O, Y, F, W, R, K, H, L, I, V or M;} \\
x_{41} & \text{ is S, Y, F, W, L, I, V or M;} \\
x_{42} & \text{ is Y, W, L, I, V, M or S;}
\end{align*}
\]

with the proviso that X38-X42 is not VIGYS,
- Optionally an extension peptide, and
- A GLP-1 peptide.

The term "GLP-1 peptide", as used herein, is intended to designate GLP-1 (7-37),
GLP-1 (7-36) amide as well as analogues thereof, which are capable of being produced by
conventional recombinant DNA techniques as well as conventional synthetic methods. Such
GLP-1 peptides include but are not limited to native glucagon-like peptide-1, for instance such
peptide fragments which comprises GLP-1 (7-37) and functional variants thereof as disclosed in
WO 87/06941; such peptide fragments which comprise GLP-1 (7-36) and functional derivatives
thereof as disclosed in WO 90/1296; such analogues of the active GLP-1 peptides 7-34, 7-35,
7-36, and 7-37 as disclosed in WO 91/1457; such N-terminal truncated fragments of GLP-1 as
disclosed in EP 0699686-A2; and such GLP-1 analogues and derivatives that include an N-
terminal imidazole group as disclosed in EP 0708179-A2. Non-limiting examples of a GLP-1
peptide is GLP-1 (7-37) and GLP-1 (7-37)[K34R].

The term "GLP-1 precursor" as used herein is intended to mean a polypeptide
comprising an extended GLP-1 peptide where the extension serves to facilitate the secretion,
expression or recovery of the GLP-1 peptide. Examples of GLP-1 precursors may be found in WO03/010186 and WO09/083549. GLP-1 precursors are intended to include GLP-1 peptides having a small extension, e.g. 2-5 amino acid residues, as well as GLP-1 peptides having longer extensions comprising a pre-peptide and a pro-peptide.

In one embodiment of the method for recombinant expression of a polypeptide in yeast said amino acid sequence of the general formula (I) has a sequence wherein

\[ X_{38} \text{ is } F, \ L \text{ or } V; \]
\[ X_{39} \text{ L, } I, \ V \text{ or } M; \]
\[ X_{40} \text{ is } G \text{ or } R; \]
\[ x_{41} \text{ is } S, \ Y, \ L, \ I, \ V \text{ or } M, \text{ and} \]
\[ X_{42} \text{ is } Y, \ W, \ L, \ V \text{ or } M. \]

In another embodiment said amino acid sequence of the general formula (I) has a sequence wherein

\[ X_{38} \text{ is } I \text{ or } V; \]
\[ X_{39} \text{ L, } I, \ V \text{ or } M; \]
\[ x_{40} \text{ is } A, \ G, \ S, \ E, \ Q, \ Y, \ F, \ W, \ R, \ K, \ H, \ L, \ I, \ V \text{ or } M; \]
\[ X_{41} \text{ is } Y, \ F, \text{ or } W, \text{ and} \]
\[ X_{42} \text{ is } L \text{ or } I. \]

In another embodiment said amino acid sequence of the general formula (I) has a sequence wherein

\[ X_{38} \text{ is } V; \]
\[ X_{39} \text{ L, } I, \ V \text{ or } M; \]
\[ x_{40} \text{ is } G \text{ or } R; \]
\[ x_{41} \text{ is } Y, \text{ and} \]
\[ X_{42} \text{ is } L. \]

In another embodiment said amino acid sequence of the general formula (I) has a sequence wherein \( X_{40} \text{ is } R. \) In another embodiment said amino acid sequence of the general formula (I) has a sequence wherein \( X_{40} \text{ is } R \text{ and } X_{42} \text{ is } L. \)

In yet another embodiment said amino acid sequence of the general formula (I) has a sequence wherein \( X_{38} \text{ is } V, \ X_{39} \text{ is } I, \ X_{40} \text{ is } R, \ X_{41} \text{ is } Y \text{ and } X_{42} \text{ is } L. \)

In another embodiment the polypeptide being recombinantly expressed is a GLP-1 peptide or a variant thereof, such as GLP-1 (7-37)[K34R] or GLP-1 (9-37)[K34R].

In one embodiment the polypeptide being recombinantly expressed has an N-terminal extension, i.e. positioned between the Mating Factor a variant and the polypeptide to be manufactured. This extension may facilitate the expression or secretion of the polypeptide.
from the host cell, or it may protect part of the polypeptide from undesirable proteolytical processing in the N-terminal. In another embodiment the N-terminal extension is a polypeptide having from 2-10 amino acid residues or having from about 8 to about 200 amino acid residues. Smaller N-terminal extensions are often used when the extension serves to facilitate the expression of the polypeptide in a host cell, or when the extension serves to protect a polypeptide from being proteolytically processed in the N-terminal. In another embodiment the N-terminal extension is selected from the group consisting of EEK, EEAEK, HK, EEAHK, E(EA)2HK, E(EA)3HK, EEGHK, EHPK, EEGEPK, EEAHELK, EEAHEVK, EEAHEMK, EEAHEFK, EEAHEYK, EEAHKEEGNTTPK and EELDARLEALK. In another embodiment the N-terminal extension is selected from the group consisting of QPMYKR, GOPMYK, PGOPMY, KPGQPM, LKPGQP, QLKPQG, LQLKPG, WLQLK, WHWLQL, AWHWLQ, EAWHVL, AEAWHW and EAEAWH.

When the expressed polypeptide includes an N-terminal extension, it is customary to cleave off this N-terminal extension by the use of a protease, a peptidase or by chemical cleavage. Proteases such as trypsin, Acromobacter lyticus protease and Enterokinase may be used. The particular proteolytic enzyme selected for the cleavage is often determined by the polypeptide being manufactured. Hence, the person skilled in the art will often select the proteolytic enzyme based on the polypeptides sequence, specifically the presence of any internal primary or secondary cleavage sites, as well as adapting the N-terminal extension to form a good cleavage site.

The cleavage efficiency of a protease when used to cleave a polypeptide expressed with an N-terminal extension may be determined by a simple assay as follows: A suitable aqueous solution of the polypeptide is incubated at a pH and temperature which is favourable to the protease and samples are withdrawn from the reaction mixture over time. As soon as the samples are withdrawn the enzyme activity is inactivated. After collecting all the samples covering the time-span of interest, the concentration of the corresponding polypeptide without the N-terminal extension is determined by e.g. HPLC analysis. Depicting the concentration of the cleaved polypeptide as a function of time will indicate the progress of the reaction. Comparing such reaction traces for different N-terminal extension of the polypeptide will allow a ranking of the N-terminal extensions in accordance with the ability of the protease to liberate the polypeptide without N-terminal extension.

The nucleic acid construct encoding the polypeptide may suitably be of genomic, cDNA or synthetic origin. Amino acid sequence alterations are accomplished by modification of the genetic code by well known techniques.
The DNA sequence encoding the polypeptide are usually inserted into a recombinant vector which may be any vector, which may conveniently be subjected to recombinant DNA procedures, and the choice of vector will often depend on the host cell into which it is to be introduced. Thus, the vector may be an autonomously replicating vector, i.e. a vector, which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g. a plasmid. Alternatively, the vector may be one which, when introduced into a host cell, is integrated into the host cell genome and replicated together with the chromosome(s) into which it has been integrated.

The vector is preferably an expression vector in which the DNA sequence encoding the polypeptide is operably linked to additional segments required for transcription of the DNA. The term, "operably linked" indicates that the segments are arranged so that they function in concert for their intended purposes, e.g. transcription initiates in a promoter and proceeds through the DNA sequence coding for the polypeptide until it terminates within a terminator.

Thus, expression vectors for use in expressing the polypeptide will comprise a promoter capable of initiating and directing the transcription of a cloned gene or cDNA. The promoter may be any DNA sequence, which shows transcriptional activity in the host cell of choice and may be derived from genes encoding proteins either homologous or heterologous to the host cell.

Additionally, expression vectors for use of expression of the polypeptide will also comprise a terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator which is functional in the host cell of choice may be used in the present invention.

Expression of the polypeptide is aimed for being directed into the secretory pathway for extracellular expression into the growth medium. Useful signal peptides for use as the pre-peptide in leaders for expression in yeast host cells are obtained e.g. from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase. Other examples of useful pre-peptides (signal peptides) are the yeast aspartic protease 3 (Yps1) signal peptide (Egel-Mitani et al. (1990) YEAST 6:127-137 and US 5,726,038), the alpha-factor signal of the *MFα1* gene (Thorner (1981) in The Molecular Biology of the Yeast *Saccharomyces cerevisiae*, Strathern et al., eds., pp 143-180, Cold Spring Harbor Laboratory, NY) and US 4,870,008), the signal peptide of mouse salivary amylase (O. Hagenbuchle et al., Nature 289, 1981, pp. 643-646), a modified carboxypeptidase signal
peptide (L.A. Vails et al., Cell 48, 1987, pp. 887-897) and the yeast BAR1 signal peptide (WO 87/02670).

The procedures used to ligate the DNA sequences coding for the polypeptide, the promoter, the terminator and secretory signal sequence, respectively, and to insert them into suitable vectors containing the information necessary for replication, are well known to persons skilled in the art (cf., for instance, Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, New York, 1989).

Many yeast cells contain an endogenous plasmid, called the 2 micron plasmid that contains various elements that ensure its stable maintenance in the yeast cell (Guerineau et al., 1971, Biochem. Biophys. Res. Comm. 42(3):550-557). The whole or part of this endogenous 2 micron can be used in conjunction with a recombinant gene, as a method for securing stable maintenance of the sequences necessary for the recombinant expression (Beggs J.D., 1978, Transformation of yeast by a replicating hybrid plasmid, Nature, 275:104-109). It has been found by the present inventor that when the endogenous 2 micron plasmid is present in the cell, the expression plasmid for recombinant expression only needs to comprise the replication origin and the STB region. The other factors present on the endogenous 2 micron plasmid can function in trans. The replication origin and the STB region only constitute a small part of the endogenous 2μ plasmid.

In an aspect the present invention provides an expression plasmid containing only the replication origin and the STB region from the 2 micron plasmid. Hence, this minimal expression plasmid does not contain any of the FLP region, the repeat 1 region, the REP1 region, D-protein, repeat 2 region and REP2 region. In an embodiment this plasmid comprises an expression cassette, an E.coli part including an AmpR gene, an S. pombe sequence encoding triose-phosphate-isomerase as described in Russell, PR (1985, Transcription of the triose-phosphate gene of Schizosaccharomyces pombe initiates from a start point different from that in Saccharomyces cerevisiae, Gene, 40:125-130). In another embodiment the minimal plasmid comprises no AmpR or other antibiotic resistance gene. Such an antibiotic resistance gene is useful during the cloning work in e.g. E.coli but it is preferable to eliminate the antibiotic resistance gene in the plasmid used for industrial scale recombinant protein expression. The antibiotic resistance gene can be made non-functional or removed from the host cell by well known procedures, see e.g. WO 00/04172. The minimal expression plasmid is useful for expression of a polypeptide in a yeast.

The vectors of the present invention preferably contain one or more selectable markers which permit easy selection of transformed cells. A selectable marker is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals,
complement auxotrophies, and the like. Examples of bacterial selectable markers are the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers which confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol or tetracycline resistance. Selectable markers for use in auxotroph yeast cells include *ADE2, HIS3, LEU2, LYS2, MET3, TRP1*, and *URA3*. A preferred selectable marker for yeast is the *Schizosaccharomyces pombe* TPI gene (Russell (1985) *Gene* 40:125-130).

In the vector, the polynucleotide sequence is operably connected to a suitable promoter sequence. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extra-cellular or intra-cellular polypeptides either homologous or heterologous to the host cell. Examples of useful promoters in yeast host cells are the *Saccharomyces cerevisiae* MFa1, TP/1, ADH2, TDH3 or *PGK1* promoters.

The polynucleotide construct of the invention will also typically be operably connected to a suitable terminator. In yeast an example of a suitable terminator is the TP/7 terminator (Alber et al. (1982) J. Mol. Appl. Genet. 1:41 9-434), but could be any endogenous yeast terminator.

The procedures used to ligate the polynucleotide sequence 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. It will be understood that the vector may be constructed either by first preparing a DNA construct containing the entire DNA sequence encoding the polypeptide to be expressed, 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 Mating Factor a variant of the invention, the polypeptide to be expressed optionally including a N-terminal extension) followed by assembly of the elements by ligation, seamless cloning methods or by cloning directly in the yeast cell by homologous recombination.

The present invention also relates to recombinant host cells, comprising a polynucleotide sequence encoding the Mating Factor a variant of the invention and the polypeptide to be expressed. A vector comprising such polynucleotide sequence is introduced into the host cell so that the vector is maintained as a chromosomal integrant or as a self-replicating extra-chromosomal vector.

"Host cell" as used herein is intended to mean a microorganism which is used for the expression of a polypeptide of interest. A host cell encompasses any progeny of
a parent cell that is not identical to the parent cell due to mutations that occur during replication.

A suitable host cell for the present invention is a yeast cell. "Yeast" as used herein includes ascosporogenous yeast (Endomycetales), basidiosporogenous yeast, and yeast belonging to the Fungi Imperfecti (Blastomycetes). The ascosporogenous yeasts are divided into the families Spermophthoraeeae and Saccharomycetaceae. The latter is comprised of four subfamilies, Schizosaccharomycoideae (e.g., genus Schizosaccharomyces), Nadsonioideae, Lipomycoideae, and Saccharomycoideae (e.g., genera Pichia, Kluyveromyces and Saccharomyces). The basidiosporogenous yeasts include the genera Leucosporidium, Rhodosporidium, Sporidiobolus, Filobasidium, and Filobasidiella. Yeast belonging to the Fungi Imperfecti are divided into two families, Sporobolomyctaceae (e.g., genera Sorobolomyces and Bullera) and Cryptococaceae (e.g., genus Candida). Since the classification of yeast may change in the future, for the purposes of this invention, yeast shall be defined as described in *Biology and Activities of Yeast* (Skinner, F.A., Passmore, S.M., and Davenport, R.R., eds, Soc. App. Bacteriol. Symposium Series No. 9, 1980. The biology of yeast and manipulation of yeast genetics are well known in the art (see, e.g., *Biochemistry and Genetics of Yeast*, Bacil, M., Horecker, B.J., and Stopani, A.O.M., editors, 2nd edition, 1987; *The Yeasts*, Rose, A.H., and Harrison, J.S., editors, 2nd edition, 1987; and *The Molecular Biology of the Yeast Saccharomyces*, Strathern et al., editors, 1981).

The yeast host cell used in the process of the invention may be any suitable yeast organism which, on cultivation, produces large amounts of the polypeptide to be expressed.

Examples of suitable yeast organisms are strains selected from a cell of a species of Candida, Kluyveromyces, Saccharomyces, Schizosaccharomyces, Pichia, Hansenula, and Yarrowia. In one embodiment, the yeast host cell is selected from a *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluveri*, *Saccharomyces norbensis*, *Saccharomyces oviformis*, *Schizosaccharomyces pombe*, *Saccharomyces uvarum*, *Pichia kluveri*, *Yarrowia lipolytica*, *Candida utilis*, *Candida cacaoi*, and *Geotrichum fermentans*. Other useful yeast host cells are a *Kluyveromyces lactis*, *Kluyveromyces fragilis*, *Hansenula polymorpha*, *Pichia pastoris Yarrowia lipolytica*, *Schizosaccharomyces pombe*, *Ustilgo maylis*, *Candida maltose*, *Pichia guillermondii* and *Pichia methanolol* (cf. Gleeson et al., *J. Gen. Microbiol.* 132, 1986, pp. 3459-3465; US 4,882,279 and US 4,879,231). The transformation of the yeast cells may for instance be effected by protoplast formation followed by transformation in a manner known per se.
The host cell for expressing the polypeptide is preferably a cell free from any functional antibiotic resistance genes. Although such antibiotic resistance genes are useful during initial cloning steps in e.g. *E. coli*, the antibiotic resistance genes can be made non-functional or removed from the host cell by well known procedures, see e.g. WO 00/04172.

"Medium" as used herein is intended to mean a liquid solution for cultivating the host cell, i.e. supporting the growth and product formation of the yeast. A suitable medium for yeast is e.g. YPD or as described in WO2008/037735. The medium contains at least one carbon source, one or several nitrogen sources, essential salts including salts of potassium, sodium, magnesium, phosphate and sulphate, trace metals, water soluble vitamins, and process aids including but not limited to antifoam agents, protease inhibitors, stabilizers, ligands and inducers. Typical carbon sources are e.g. mono- or disaccharides. Typical nitrogen sources are, e.g. ammonia, urea, amino acids, yeast extract, corn steep liquor and fully or partially hydrolysed proteins. Typical trace metals are e.g. Fe, Zn, Mn, Cu, Mo and H₃BO₃. Typical water soluble vitamins are e.g. biotin, pantothenate, niacin, thiamine, p-aminobenzoic acid, choline, pyridoxine, folic acid, riboflavin and ascorbic acid.

By "fermentation" as used herein is intended to mean an aseptic process used for propagating microorganisms submerged in a liquid medium. The fermentation is preferably carried out in aseptic, stirred tanks with supply lines for addition of compressed, sterile gasses consisting of but not limited to air, oxygen and ammonia. A fermentation tank can contain sensors and devices for monitoring pH, temperature, pressure, agitation rate, dissolved oxygen level, liquid content, foam level, feed addition rates and rates of adding acid and base. Furthermore, the fermentation tank can be equipped with optical devices for monitoring levels of cell density, concentrations of metabolites and products regardless of their physio-chemical form.

The desired product produced during the fermentation is present as soluble extracellular material or as intracellular material either in the form of soluble material or as insoluble material including aggregated material. It is preferable that it is present as soluble extracellular material. A fermentation process is typically carried out in tanks with a working volume ranging from 100 mL to 200,000 L. A fermentation process can be operated as a batch process, a fed-batch process, a repeated fed-batch process or a continuous process.

The secreted polypeptides, 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, filtration or catching the polypeptide by an ion-exchange matrix or by a reverse phase absorption matrix, precipitating the protein 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.

The novel Mating Factor a variants of the present invention also facilitate reduced O-glycosylation of the polypeptide during expression in yeast. As such the Mating Factor a variants of the present invention can be used in an improved method for making polypeptides such as GLP-1 peptides in yeast. Expressing the polypeptide in a yeast cell having reduced capacity for O-glycosylation may maintain the improved yield of said precursor while at the same time reducing even further the fraction of said polypeptide that is O-glycosylated during expression.

Protein O-mannosyltransferases (PMTs) initiate the assembly of O-mannosyl glycans, an essential protein modification in fungi. PMTs are conserved in fungi and the PMT family is phylogenetically classified into PMT1, PMT2 and PMT4 subfamilies, which differ in protein substrate specificity. The protein O-mannosyltransferases Pmtlp and Pmt2p are catalyzing the O-glycosylation of serine and threonine residues in proteins in the endoplasmic reticulum of yeast by transfer of a mannosyl residue from Dolichyl phosphate-D-mannose (Gentzsch et al., FEBS Lett 1995, 18, pp128-130). In Saccharomyces cerevisiae as well as in many other yeasts the PMT family is highly redundant, and only the simultaneous deletion of PMT1/PMT2 and PMT4 subfamily members is lethal (Girrbach and Strahl, J. Biol. Chem. 2003, 278, pp12554-62). US 5,714,377 describe that yeast cells having reduced O-glycosylation capacity from PMT1/PMT2 modification are still viable and show good growth characteristics in industrial fermentation conditions.

**Non-limiting embodiments**

The invention is further described by the following non-limiting embodiments:

1. A method for recombinant expression of a polypeptide in yeast comprising the culturing of a yeast strain comprising a DNA sequence encoding a processing and secretion signal upstream of the polypeptide, wherein said processing and secretion signal comprises a Mating Factor a pro-peptide variant having at least one substitution in the VIGYL sequence at positions 38-42 to comprise the amino acid sequence of the general formula (I):

\[
X_{38}X_{39}X_{40}X_{41}X_{42} \quad (\text{SEQ ID NO:1}) \quad (I)
\]

wherein

- \(X_{38}\) is F, L, I or V;
- \(X_{39}\) is L, I, V or M;
- \(X_{40}\) is A, G, S, E, Q, Y, F, W, R, K, H, L, I, V or M;
- \(X_{41}\) is S, Y, F, W, L, I, V or M;
$X_{42}$ is Y, W, L, I, V, M or S;
with the proviso that $X_{38}$-$X_{42}$ is not VIGYS.

2. The method according to embodiment 1 wherein
$X_{38}$ is F, L or V;
$X_{39}$ is L, I, V or M;
$X_{40}$ is G or R;
$X_{41}$ is S, Y, L, I, V or M, and
$X_{42}$ is Y, W, L, V or M.

3. The method according to claim 1 wherein
$X_{38}$ is I or V;
$X_{39}$ is L, I, V or M;
$X_{40}$ is A, G, S, E, Q, Y, F, W, R, K, H, L, I, V or M;
$X_{41}$ is Y, F, or W, and
$X_{42}$ is L or I.

4. The method according to any of embodiments 1-3 wherein
$X_{38}$ is V;
$X_{39}$ is L, I, V or M;
$X_{40}$ is G or R;
$X_{41}$ is Y, and
$X_{42}$ is L.

5. The method according to any of embodiments 1-4, wherein four of the amino acid residues in $X_{38}$-$X_{39}$-$X_{40}$-$X_{41}$-$X_{42}$ are identical to the corresponding amino acid residue in VIGYL or VIGYS.

6. The method according to any of embodiments 1-5, wherein at least three of the amino acid residues in $X_{38}$-$X_{39}$-$X_{40}$-$X_{41}$-$X_{42}$ are identical to the corresponding amino acid residue in VIGYL or VIGYS.

7. The method according to any of embodiments 1-6, wherein $X_{41}$ is Y.

8. The method according to any of embodiments 1-6, wherein $X_{41}$ is L.

9. The method according to any of embodiments 1-8, wherein $X_{42}$ is L.

10. The method according to any of embodiments 1 and 5-8, wherein $X_{42}$ is S.

11. The method according to any of embodiments 1 and 5-8, wherein $X_{42}$ is Y, W, L, I, V or M.

12. The method according to any of embodiments 1-11, wherein $X_{40}$ is R.

13. The method according to embodiment 12, wherein $X_{38}$ is V, $X_{39}$ is I, $X_{41}$ is Y and $X_{42}$ is L.

14. The method according to embodiment 12, wherein $X_{38}$ is V, $X_{39}$ is I, $X_{41}$ is Y and $X_{42}$ is S.
15. The method according to any of embodiments 1 and 12, wherein formula (I) is VI-X_{40}-YL or VI-X_{40}-YS.

16. The method according to embodiment 15 wherein X_{40} is A, Y, F, W, R, K, L, I, V or M.

17. The method according to any of embodiments 1-16, wherein said Mating Factor a pro-peptide variant has less than 10 amino acid residue changes outside of the X_{38}-X_{1,2} sequence as compared to the Mating Factor a pro-peptide as set out in SEQ ID NO:2 (amino acid residues 20-85).

18. The method according to any of embodiments 1-17, wherein said Mating Factor a pro-peptide variant has less than 5 amino acid residue changes outside of the X_{38}-X_{1,2} sequence as compared to the Mating Factor a pro-peptide as set out in SEQ ID NO:2 (amino acid residues 20-85).

19. The method according to any of embodiments 1-18, wherein said Mating Factor a pro-peptide variant has less than 2 amino acid residue changes outside of the X_{38}-X_{1,2} sequence as compared to the Mating Factor a pro-peptide as set out in SEQ ID NO:2 (amino acid residues 20-85).

20. The method according to any of embodiments 1-19, wherein said Mating Factor a pro-peptide variant is part of a Mating Factor a prepro-peptide comprising a pre-peptide as the N-terminal part fused to said Mating Factor a pro-peptide variant as the C-terminal part.

21. The method according to embodiment 20, wherein said pre-peptide is from yeast aspartic protease 3 (YAP3) signal peptide, from the ofactor signal of the MFa1 gene from S. cerevisiae or a variant thereof.

22. The method according to any of embodiments 1-21, wherein said yeast carries at least one genetic modification reducing its capacity for O-glycosylation.

23. The method according to embodiment 22, wherein said yeast carries at least one genetic modification within the genes for PMT1 or PMT2 reducing its capacity for O-glycosylation.

24. The method according to any of embodiments 22-23, wherein said yeast carries at least one genetic modification reducing its capacity for O-glycosylation by the protein O-mannosyltransferase 1 (PMT1) of the polypeptide GLP-1 (7-37)[K34R] when expressed with the alpha leader as compared to the yeast carrying the corresponding unmodified genes.

25. The method according to any of embodiments 22-24, wherein said yeast carries at least one genetic modification reducing its capacity for O-glycosylation by protein O-mannosyltransferase 2 (PMT2) of the polypeptide GLP-1 (7-37)[K34R] when expressed
with the alpha leader as compared to the yeast carrying the corresponding unmodified genes.

26. The method according to any of embodiments 22-25, wherein said capacity for O-glycosylation is reduced by at least a factor 2.

27. The method according to any of embodiments 22-26, wherein said capacity for O-glycosylation is reduced by at least a factor 4.

28. The method according to any of embodiments 22-27, wherein said at least one genetic modification is located in the coding region of PMT1 or PMT2.

29. The method according to any of embodiments 22-27, wherein said at least one genetic modification is located in the regions responsible for or involved in the expression and/or transcriptional regulation of PMT1 or PMT2.

30. The method according to any of embodiments 22-27, wherein the PMT1 gene in said yeast is deleted.

31. The method according to any of embodiments 22-27, wherein the PMT1 and PMT2 genes in said yeast are both deleted.

32. The method according to any of embodiments 1-31, wherein said polypeptide comprises a GLP-1 peptide.

33. The method according to embodiment 32, wherein said polypeptide comprises GLP-1 (9-37)[K34R] or GLP-1 (9-37)[K34R].

34. The method according to embodiment 33, wherein said polypeptide is GLP-1 (7-37)[K34R], GLP-1 (9-37)[K34R] or GLP-1 (9-37)[K34R,G37K] (SEQ ID NO:44).

35. The method according to any of embodiments 32-34, wherein said polypeptide has an N-terminal extension.

36. The method according to embodiment 35, wherein said N-terminal extension is selected from the group consisting of EKE, EEAKE, HK, EEHK, E(AE)2HK, E(AE)3HK, EEHK, HEPK, EEPEK, EEHEK, EEHEK, EEAHEK, EEAHEY, EEAHEKE, EEAHELK, EEAHEVK, EEAHEMK, EEAHEFK, EEAHEYK, EEAHEKEEGNTTPK and EELDARLEALK.

37. The method according to embodiment 35, wherein said N-terminal extension is selected from the group consisting of DV, DVKPGQPLA, DVKPGQPEY, DVKPGEPLY, DVKPGQPLY, DVKPGQPLE, DVKPGQPMY and DVKPGQPMYDDDDK.

38. The method according to embodiment 35, wherein said N-terminal extension is selected from the group consisting of QPMYKR, QPMPYK, QPQMPY, KPGQPM, LKPGQ, QLKPG, QLKPG, WLQLKP, HWLQLK, HWLQL, AWHWLQ, AWHWL, AEAHW and EAEAWH.
39. Mating Factor a pro-peptide variant having at least one substitution in the VIGYL sequence at positions 38-42 to comprise the amino acid sequence of the general formula (I):

$$\text{X}_{38}\text{-X}_{39}\text{-X}_{40}\text{-X}_{41}\text{-X}_{42} \quad \text{(SEQ ID NO:1)} \quad \text{(I)}$$

wherein

- $X_{38}$ is F, L, I or V;
- $X_{39}$ is L, V or M;
- $X_{40}$ is A, G, S, E, Q, Y, F, W, R, K, H, L, I, V or M;
- $X_{41}$ is S, Y, F, W, L, I, V or M;
- $X_{42}$ is Y, W, L, I, V, M or S;

with the proviso that $X_{38}\text{-X}_{42}$ is not VIGYS.

40. The Mating Factor a pro-peptide variant according to embodiment 39 wherein $X_{38}\text{-X}_{42}$ is not VIGYS, VIDYS, VATYL, VIGYR, or AIGYL.

41. GLP-1 precursor which is a fusion polypeptide comprising:

- A pre-peptide,
- A Mating Factor a pro-peptide variant having at least one substitution in the VIGYL sequence at positions 38-42 to comprise the amino acid sequence of the general formula (I):

$$\text{X}_{38}\text{-X}_{39}\text{-X}_{40}\text{-X}_{41}\text{-X}_{42} \quad \text{(SEQ ID NO:1)} \quad \text{(I)}$$

wherein

- $X_{38}$ is F, L, I or V;
- $X_{39}$ is L, V or M;
- $X_{40}$ is A, G, S, E, Q, Y, F, W, R, K, H, L, I, V or M;
- $X_{41}$ is S, Y, F, W, L, I, V or M;
- $X_{42}$ is Y, W, L, I, V, M or S;

with the proviso that $X_{38}\text{-X}_{42}$ is not VIGYS,

- Optionally an extension peptide, and
- A GLP-1 peptide.

42. The GLP-1 precursor according to embodiment 41 wherein the peptides comprised by said GLP-1 precursor are fused according to the order in which they are listed, i.e. pre-peptide - Mating Factor a pro-peptide variant, the optional extension peptide and the GLP-1 peptide.

43. DNA sequence encoding the polypeptide according to any of embodiments 39-42.

44. Expression vector comprising a DNA sequence according to embodiment 43.
45. Expression vector according to embodiment 44, wherein said DNA sequence encoding a
polypeptide to be expressed is operatively linked to an upstream promoter and a
downstream terminator.

46. Host cell comprising the expression vector according to any of embodiments 44-45.

5 47. The host cell according to embodiment 46, which is selected from the group consisting of
Saccharomyces spp., Pichia spp., Hansenula spp., Arxula spp., Kluyveromyces spp.,
Yarrowia spp. and Schizosaccharomyces spp.

All references, including publications, patent applications, and patents, cited herein
are hereby incorporated by reference in their entirety and to the same extent as if each
reference were individually and specifically indicated to be incorporated by reference and
were set forth in its entirety herein (to the maximum extent permitted by law). All headings
and sub-headings are used herein for convenience only and should not be construed as
limiting the invention in any way. The use of any and all examples, or exemplary language
(e.g., "such as") provided herein, is intended merely to better illuminate the invention and
does not pose a limitation on the scope of the invention unless otherwise claimed. No
language in the specification should be construed as indicating any non-claimed element as
essential to the practice of the invention. The citation and incorporation of patent documents
herein is done for convenience only and does not reflect any view of the validity,
patentability, and/or enforceability of such patent documents. This invention includes all
modifications and equivalents of the subject matter recited in the claims appended hereto as
permitted by applicable law.

EXAMPLES

25

Examples 1-45.

We constructed a plasmid, that contains the TDH3 promoter, the gene encoding the
MFalpha pre-pro-peptide, the gene encoding an N-terminal extended GLP-1
(DVKPGQPMYDDDK-GLP-1 {7-37}[K34R]) (SEQ ID NO:45), a minimal 2 micron region for
maintenance in yeast, and a selectable marker, namely the TPI gene from S.pombe (see
Figure 1).

In this plasmid background we introduced various single mutations in the region
encoding position 38-42 of the MFalpha-pro-peptide region, the _X38-X42 = VIGYL sequence
(see Figure 2). The experiment using the wt VIGYL sequence is Example 1, c.f. tables 1-5. Other reference experiments are Examples 5-6, 10-11, 28-29, 37-38 and 44-45.

These plasmids were introduced into a *Saccharomyces cerevisiae* strain lacking the TP/1 gene, allowing for selection of transformants harbouring the plasmid on media containing glucose as sole carbon source. Transformants were thereafter cultivated for 3 days in shake flasks in 5 ml relevant media and the cultivation supernatants were analyzed by LCMS for concentration of secreted GLP-1 peptide, and degree of O-glycosylation of the peptide. In tables 1-5 are shown the yield and O-glycosylation degree of the wild type and of the single amino acid residue mutants of the X38-X42 sequence (a couple of reference examples on each position being included).

Table 1. Data from expression of N-terminal extended GLP-1 under control of Mating Factor a propeptide mutants having a single amino acid mutation in position 38 (X38).

Concentrations of N-terminal extended GLP-1 and O-glyco impurity are normalised against the data from same expression but under control of the Mating Factor a propeptide having the wildtype Val as X38.

<table>
<thead>
<tr>
<th>Position 38</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

Table 2. Data from expression of N-terminal extended GLP-1 under control of Mating Factor a propeptide mutants having a single amino acid mutation in position 39 (X39).

Concentrations of N-terminal extended GLP-1 and O-glyco impurity are normalised against the data from same expression but under control of the Mating Factor a propeptide having the wildtype Ile as X39.

<table>
<thead>
<tr>
<th>Position 39</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
</tr>
</tbody>
</table>


Table 3. Data from expression of N-terminal extended GLP-1 under control of Mating Factor a propeptide mutants having a single amino acid mutation in position 40 (X$_{40}$).

Concentrations of N-terminal extended GLP-1 and O-glyco impurity are normalised against the data from same expression but under control of the Mating Factor a propeptide having the wildtype Gly as X$_{40}$.

<table>
<thead>
<tr>
<th>Example</th>
<th>Mutation</th>
<th>% cone. relative to wt</th>
<th>% O-glyco relative to wt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>lie</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>Leu</td>
<td>113</td>
<td>82</td>
</tr>
<tr>
<td>8</td>
<td>Val</td>
<td>108</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>Met</td>
<td>126</td>
<td>88</td>
</tr>
<tr>
<td>10</td>
<td>Glu</td>
<td>40</td>
<td>337</td>
</tr>
<tr>
<td>11</td>
<td>His</td>
<td>25</td>
<td>313</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example</th>
<th>Mutation</th>
<th>% cone. relative to wt</th>
<th>% O-glyco relative to wt</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Ala</td>
<td>84</td>
<td>48</td>
</tr>
<tr>
<td>13</td>
<td>Ser</td>
<td>79</td>
<td>89</td>
</tr>
<tr>
<td>14</td>
<td>Thr</td>
<td>80</td>
<td>67</td>
</tr>
<tr>
<td>15</td>
<td>Asp</td>
<td>51</td>
<td>90</td>
</tr>
<tr>
<td>16</td>
<td>Glu</td>
<td>58</td>
<td>92</td>
</tr>
<tr>
<td>17</td>
<td>Gin</td>
<td>87</td>
<td>98</td>
</tr>
<tr>
<td>18</td>
<td>Tyr</td>
<td>60</td>
<td>49</td>
</tr>
<tr>
<td>19</td>
<td>Phe</td>
<td>50</td>
<td>23</td>
</tr>
<tr>
<td>20</td>
<td>Trp</td>
<td>53</td>
<td>18</td>
</tr>
<tr>
<td>21</td>
<td>Arg</td>
<td>106</td>
<td>14</td>
</tr>
<tr>
<td>22</td>
<td>Lys</td>
<td>93</td>
<td>36</td>
</tr>
<tr>
<td>23</td>
<td>His</td>
<td>74</td>
<td>83</td>
</tr>
<tr>
<td>24</td>
<td>Leu</td>
<td>42</td>
<td>13</td>
</tr>
<tr>
<td>25</td>
<td>lie</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>Val</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>Met</td>
<td>60</td>
<td>36</td>
</tr>
<tr>
<td>28</td>
<td>Asn</td>
<td>81</td>
<td>112</td>
</tr>
</tbody>
</table>
Table 4. Data from expression of N-terminal extended GLP-1 under control of Mating Factor a propeptide mutants having a single amino acid mutation in position 41 (X_{41}). Concentrations of N-terminal extended GLP-1 and O-glyco impurity are normalised against the data from same expression but under control of the Mating Factor a propeptide having the wildtype Tyr as X_{41}.

<table>
<thead>
<tr>
<th>Example</th>
<th>Mutation</th>
<th>% conc. relative to wt</th>
<th>% O-glyco relative to wt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tyr</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>30</td>
<td>Phe</td>
<td>69</td>
<td>74</td>
</tr>
<tr>
<td>31</td>
<td>Trp</td>
<td>79</td>
<td>73</td>
</tr>
<tr>
<td>32</td>
<td>Ser</td>
<td>107</td>
<td>182</td>
</tr>
<tr>
<td>33</td>
<td>Leu</td>
<td>129</td>
<td>104</td>
</tr>
<tr>
<td>34</td>
<td>Ile</td>
<td>111</td>
<td>171</td>
</tr>
<tr>
<td>35</td>
<td>Val</td>
<td>134</td>
<td>136</td>
</tr>
<tr>
<td>36</td>
<td>Met</td>
<td>131</td>
<td>151</td>
</tr>
<tr>
<td>37</td>
<td>Gly</td>
<td>50</td>
<td>366</td>
</tr>
<tr>
<td>38</td>
<td>Asp</td>
<td>27</td>
<td>590</td>
</tr>
</tbody>
</table>

Table 5. Data from expression of N-terminal extended GLP-1 under control of Mating Factor a propeptide mutants having a single amino acid mutation in position 42 (X_{42}). Concentrations of N-terminal extended GLP-1 and O-glyco impurity are normalised against the data from same expression but under control of the Mating Factor a propeptide having the wildtype Leu as X_{42}.

<table>
<thead>
<tr>
<th>Example</th>
<th>Mutation</th>
<th>% cone, relative to wt</th>
<th>% O-glyco relative to wt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Leu</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>39</td>
<td>Ile</td>
<td>98</td>
<td>64</td>
</tr>
<tr>
<td>40</td>
<td>Tyr</td>
<td>123</td>
<td>156</td>
</tr>
<tr>
<td>41</td>
<td>Trp</td>
<td>116</td>
<td>163</td>
</tr>
<tr>
<td>42</td>
<td>Val</td>
<td>103</td>
<td>111</td>
</tr>
</tbody>
</table>
Example 46.

A plasmid similar to those in Examples 1-45, containing the TDH3 promoter, the gene encoding the MFalpha pre-pro-peptide, the gene encoding an N-terminal extended GLP-1 (DVKPGQPMYDDDDK-GLP-1 (7-37)[K34R]), a minimal 2 micron region for maintenance in yeast, and a selectable marker (Figure 1) was transformed into two different strain backgrounds - one containing the PMT1 gene (PMT+), the other deleted for the PMT1 gene (PMT-). The PMT1 gene is encoding a protein mannosyltransferase involved in O-glycosylation.

The MFalpha pro-peptide was either wildtype (40G) or mutated (G40R). The strains were cultivated under identical conditions in continuous cultures, with a dilution rate = 0.1, and pH 5.8. Samples were analyzed by HPLC to determine the concentration of GLP-1 precursor in the spent medium and by LCMS to determine the proportion of O-glycosylated GLP-1 precursor.

The results demonstrate that the effects indeed are additive. The G40R mutation reduced O-glycosylation with more than 80%. Deletion of PMT1 reduces O-glycosylation by further more than 80%, in this case giving levels of O-glycosylation close to the detection limit. Furthermore, the G40R mutation gives a surprising increase in yield of the N-terminally extended GLP-1 as compared to the 40G wildtype version, both in PMT+ strain as well as in pmt1 deleted strain (see Table 6).

These results demonstrate that the combination of a beneficial mutation in the MFalpha propeptide, e.g. G40R, and a PMT1 deleted host strain results in an additive reduction in O-glycosylation.

Table 6. Normalised maximum concentration (yield) of the peptide DVKPGQPMYDDDDK-GLP-1(7-37)[K34R] in continuous cultures when using 40R mutated MFalpha pro-peptide as compared to the corresponding wildtype 40G.

<table>
<thead>
<tr>
<th></th>
<th>PMT+ strain</th>
<th>PMT- strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>40G</td>
<td>40R</td>
<td>40G</td>
</tr>
</tbody>
</table>
Table 7. Data from expression of specific N-terminal extended GLP-1 peptide under control of Mating Factor a propeptide mutants with or without a single amino acid substitution to Arg in position 40 (X40). Concentrations of N-terminal extended GLP-1 peptide (yield) and O-glycosylated impurity are normalised against the data from expression of the same GLP-1 peptide but under control of the Mating Factor a propeptide having the wildtype Gly as X40.

<table>
<thead>
<tr>
<th>GLP-1 peptide</th>
<th>X40 of MFalpha pre-pro-peptide</th>
<th>Yield (% relative to wt)</th>
<th>O-glycosylation (% relative to wt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLP-1(9-37)[K34R]</td>
<td>Gly (wt)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>GLP-1(9-37)[K34R]</td>
<td>Arg</td>
<td>127</td>
<td>&lt; 1 (BDL)</td>
</tr>
<tr>
<td>GLP-1(9-37)[K34R,G37K]</td>
<td>Gly (wt)</td>
<td>100</td>
<td>BDL</td>
</tr>
<tr>
<td>GLP-1(9-37)[K34R,G37K]</td>
<td>Arg</td>
<td>143</td>
<td>BDL</td>
</tr>
</tbody>
</table>

BDL: below detection limit

Example 47.

To examine the effect of one of the mutations during expression of other GLP-1 peptides, we constructed plasmids, that contained the TDH3 promoter, the gene encoding the MFalpha pre-pro-peptide in the form of either the wildtype (40G) or a 40R mutation, the gene encoding N-terminal extended GLP-1 (DVKPGQPMYDDDDK-GLP-1 (9-37)[K34R] or DVKPGQPMYDDDDK-GLP-1 (9-37)[K34R,G37K]), a minimal 2 micron region for maintenance in yeast, and a selectable marker, namely the TPI gene from S. pombe (see Figure 1).

These plasmids were introduced into a Saccharomyces cerevisiae strain lacking the TP1 gene, allowing for selection of transformants harbouring the plasmid on media containing glucose as sole carbon source. Transformants were thereafter cultivated for 3 days in shake flasks in 5 ml relevant media and the cultivation supernatants were analyzed by LCMS for concentration of secreted GLP-1 peptide, and degree of O-glycosylation of the peptide. Table 7 shows the results including the yield of GLP-1 peptide and O-glycosylation degree of GLP-1 peptide when expressed of the wild type (wt) and of the 40R mutation.

While certain features of the invention have been illustrated and described herein, many modifications, substitutions, changes, and equivalents will now occur to those of
ordinary skill in the art. It is, therefore, to be understood that the appended claims are intended to cover all such modifications and changes as fall within the true spirit of the invention.
CLAIMS

1. A method for recombinant expression of a polypeptide comprising a GLP-1 peptide in yeast comprising the culturing of a yeast strain comprising a DNA sequence encoding a processing and secretion signal upstream of the polypeptide, wherein said processing and secretion signal comprises a Mating Factor a pro-peptide variant having at least one substitution in the VIGYL sequence at positions 38-42 to comprise the amino acid sequence of the general formula (I):

   \[ X_{38}-X_{39}-X_{40}-X_{41}-X_{42} \] (SEQ ID NO:1) (I)

   wherein
   
   \( X_{38} \) is F, L, I or V;
   \( X_{39} \) L, I, V or M;
   \( X_{40} \) is A, G, S, E, Q, Y, F, W, R, K, H, L, I, V or M;
   \( X_{41} \) is S, Y, F, W, L, I, V or M;
   \( X_{42} \) is Y, W, L, I, V, M or S;

   with the proviso that \( X_{38}-X_{42} \) is not VIGYS.

2. The method according to claim 1 wherein

   \( X_{38} \) is F, L or V;
   \( X_{39} \) L, I, V or M;
   \( X_{40} \) is G or R;
   \( X_{41} \) is S, Y, L, I, V or M, and
   \( X_{42} \) is Y, W, L, V or M.

3. The method according to claim 1 wherein

   \( X_{38} \) is I or V;
   \( X_{39} \) L, I, V or M;
   \( X_{40} \) is A, G, S, E, Q, Y, F, W, R, K, H, L, I, V or M;
   \( X_{41} \) is Y, F, or W, and
   \( X_{42} \) is L or I.

4. The method according to any of claims 1-3 wherein

   \( X_{38} \) is V;
   \( X_{39} \) L, I, V or M;
   \( X_{40} \) is G or R;
5. The method according to any of claims 1-4, wherein $X_{40}$ is R.

6. The method according to any of claims 1-5, wherein said Mating Factor a pro-peptide variant has less than 10 amino acid residue changes outside of the $X_{38}$-$X_{42}$ sequence as compared to the Mating Factor a pro-peptide as set out in SEQ ID NO:2 (amino acid residues 20-85).

7. The method according to any of claims 1-6, wherein said yeast carries at least one genetic modification reducing its capacity for O-glycosylation.

8. The method according to claim 7, wherein the PMT1 gene in said yeast is deleted.

9. The method according to any of claims 1-8, wherein said polypeptide comprises GLP-1(9-37)[K34R] or GLP-1(9-37)[K34R,G37K].

10. The method according to any of claims 1-9, wherein said polypeptide consists of GLP-1(9-37)[K34R] or GLP-1(9-37)[K34R,G37K].

11. The method according to any of claims 9-10, wherein said polypeptide has an N-terminal extension.

12. Mating Factor a pro-peptide variant having at least one substitution in the VIGYLY sequence at positions 38-42 to comprise the amino acid sequence of the general formula (I):

$$X_{38}-X_{39}-X_{40}-X_{41}-X_{42} \text{ (SEQ ID NO:1) (I)}$$

wherein

- $X_{38}$ is F, L, I or V;
- $X_{39}$ is L, V or M;
- $X_{40}$ is A, G, S, E, Q, Y, F, W, R, K, H, L, I, V or M;
- $X_{41}$ is S, Y, F, W, L, I, V or M;
- $X_{42}$ is Y, W, L, I, V, M or S;

with the proviso that $X_{38}$-$X_{42}$ is not VIGYS, VIDYS, VATYL, VIGY, or AIGYL.
13. GLP-1 precursor which is a fusion polypeptide comprising:
- A pre-peptide,
- A Mating Factor a pro-peptide variant having at least one substitution in the VIGYL sequence at positions 38-42 to comprise the amino acid sequence of the general formula (I):

\[ X_{38}-X_{39}-X_{40}-X_{41}-X_{42} \quad (\text{SEQ ID NO:1}) \quad (I) \]

wherein

- \( X_{38} \) is F, L, I or V;
- \( X_{39} \) is L, I, V or M;
- \( X_{40} \) is A, G, S, E, Q, Y, F, W, R, K, H, L, I, V or M;
- \( X_{41} \) is S, Y, F, W, L, I, V or M;
- \( X_{42} \) is Y, W, L, I, V, M or S;

with the proviso that \( X_{38}-X_{42} \) is not VIGYS,

- Optionally an extension peptide, and
- A GLP-1 peptide.

14. Expression vector comprising a DNA sequence encoding the polypeptide according to any of claims 12-13.

15. Host cell comprising the expression vector according to claim 14.
Fig. 1
Fig. 2

1 19 20 VIGYL

Pre-peptide Pro-peptide DVKPGQPMYDDDK-GLP1(7-37)K34R

MFalpha pre-pro peptide
1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
   a. [X] forming part of the international application as filed:
      - [ ] in the form of an Annex C/ST.25 text file.
      - [X] on paper or in the form of an image file.
   b. [ ] furnished together with the international application under PCT Rule 13fer1 (a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
   c. [ ] furnished subsequent to the international filing date for the purposes of international search only:
      - [ ] in the form of an Annex C/ST.25 text file (Rule 13fer1 (a)).
      - [ ] on paper or in the form of an image file (Rule 13fer1 (b) and Administrative Instructions, Section 7.13).

2. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:
A. CLASSIFICATION OF SUBJECT MATTER

INV. C07K14/435 C12N15/00
ADD.

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)

C07K C12N

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

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

EPO-Internal, Sequence Search, BIOSIS, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>RAKESTRAW J ANDY ET AL: &quot;Directed Evolution of a Secretary Leader for the Improved Expression of Heterologous Proteins and Full-Length Anti bodies in Saccharomyces cerevisiae”, BIOTECHNOLOGY AND BIOENGINEERING, vol. 103, no. 6, August 2009 (2009-08) , pages 1192-1201, XP002727251, ISSN: 0006-3592 figure 1 and Conclusion</td>
<td>1-3,6-15</td>
</tr>
<tr>
<td>X</td>
<td>Wo 03/062431 A2 (DEBIOPHARM SA [CH]; GERARD CHRISTINE [BE]; PONCIN ALAIN [BE]; SAUDUBRA) 31 July 2003 (2003-07-31) claims 1, 8</td>
<td>1,12</td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C.

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” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) one or more of 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

“Z” document member of the same patent family

Date of the actual completion of the international search

30 April 2015

Date of mailing of the international search report

11/05/2015

Name and mailing address of the ISA:

European Patent Office, P.B. 5818 Patentlaan 2
NL-2280 HV Rijswijk
Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016

Gomez Ortiz, Mario
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>WO 03062431</td>
<td>31-07-2003</td>
<td>AU 2003210220 A1</td>
</tr>
</tbody>
</table>