Polypeptides comprising a modified human serum albumin secretion pre-sequence with improved secretion yield

Polypeptide erhaltend eine modifizierte menschliche Serum-Albumin Sekretions-Pre-Sequenz mit erhöhtem Ausscheidungsertrag

Polypeptides avec une pré-séquence de sécrétion de la sérum albumine humaine modifiée avec rendement de sécrétion amélioré

References cited:

WO-A-1-96/29415

• DATABASE REGISTRY [Online] retrieved from REGISTRY Database accession no. 473511-23-8 XP002264012

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).
• GIERASCH L.M.: "SIGNAL SEQUENCES",
  BIOCHEMISTRY, vol. 28, no. 3, 7 February 1989
  (1989-02-07), pages 923-930, XP007906484, US

• GIERASCH L.M.: "SIGNAL SEQUENCES"
  BIOCHEMISTRY, vol. 28, no. 3, 7 February 1989
  (1989-02-07), pages 923-930, XP007906484 US
The present invention relates to polypeptide and polynucleotide sequences for secreting proteins from host cells.

Numerous natural or artificial polypeptide signal sequences (also called secretion pre regions) have been used or developed for secreting desired peptides, polypeptides and proteins (these terms are used interchangeably from hereon in) from host cells. The signal sequence directs the nascent protein towards the machinery of the cell that exports proteins from the cell into the surrounding medium or, in some cases, into the periplasmic space. The signal sequence is usually, although not necessarily, located at the N-terminus of the primary translation product and is generally, although not necessarily, cleaved off the desired protein during the secretion process, to yield the "mature" protein.

In the case of some desired proteins the entity that is initially secreted, after the removal of the signal sequence, includes additional amino acids at its N-terminus called a "pro" sequence, the intermediate entity being called a "pro-protein". These pro sequences may assist the final protein to fold and become functional, and are usually then cleaved off. In other instances, the pro region simply provides a cleavage site for an enzyme to cleave off the pre-pro region and is not known to have another function.

The pro sequence can be removed either during the secretion of the desired protein from the cell or after export from the cell into the surrounding medium or periplasmic space.

Polypeptide sequences which direct the secretion of proteins, whether they resemble signal (i.e. pre) sequences or pre-pro secretion sequences, are sometimes also referred to as leader sequences. The secretion of proteins is a dynamic process involving translation, translocation and post-translational processing, and one or more of these steps may not necessarily be completed before another is either initiated or completed. Gierasch (Biochemistry (1989) Vol 28 (3): 923-930 reviews leader sequences.

For production of proteins in eukaryotic species such as the yeasts Saccharomyces cerevisiae and Pichia pastoris, known leader sequences include those from the S. cerevisiae acid phosphatase protein (Pho5p) (see EP 366 400), the invertase protein (Suc2p) (see Smith et al. (1985) Science, 229, 1219-1224) and heat-shock protein-150 (Hsp150p) (see WO 95/33833). Additionally, leader sequences from the S. cerevisiae mating factor alpha-1 protein (MFα-1) and from the human lysozyme and human serum albumin (HSA) protein have been used, the latter having been used especially, although not exclusively, for secreting human albumin. WO 90/01063 discloses a fusion of the MFα-1 and HSA leader sequences, which advantageously reduces the production of a contaminating fragment of human albumin relative to the use of the MFα-1 leader sequence.

Unexpectedly, we have found that the yield of secreted protein can be increased by the introduction of an amino acid sequence motif, preferably by modification of the pre-sequence of a human serum albumin leader sequence.

In a first aspect of the present invention there is provided a polypeptide comprising (i) a leader sequence, the leader sequence comprising (a) a human serum albumin secretion pre sequence and (b) the following motif:

\[-X_1\cdot X_2\cdot \cdot X_3\cdot X_4\cdot X_5\-]

where \(X_1\) is phenylalanine, tryptophan, or tyrosine, \(X_2\) is isoleucine, leucine, valine, alanine or methionine, \(X_3\) is leucine, valine, alanine or methionine, \(X_4\) is serine or threonine and \(X_5\) is isoleucine, valine, alanine or methionine; and (ii) a desired protein, heterologous to the leader sequence, and is located at positions -20, -19, -18, -17, and -16 of the albumin pre-sequence, respectively, in place of the naturally occurring amino acids at those positions, wherein the numbering is such that the -1 residue is the C-terminal amino acid of the native albumin pro sequence; and wherein when the polypeptide is expressed from a host cell, the secretion yield of the heterologous protein is increased relative to the secretion yield of the heterologous protein resulting from a polypeptide comprising the human serum albumin secretion pre-sequence.

In other words, the polypeptide includes a sequence according to SEQ ID NO 1 -
In a preferred embodiment of the first aspect of the present invention, X1 is phenylalanine. Thus a preferred polypeptide includes the sequence of SEQ ID NO 2 -

N-Phe-(Ile/Leu/Val/Ala/Met)-(Leu/Val/Ala/Met)-(Ser/Thr)-(Ile/Val/Ala/Met)-C

SEQ ID No 2

In another preferred embodiment of the first aspect of the present invention, X2 is isoleucine. Thus another preferred polypeptide includes the sequence of SEQ ID NO 3 -

N-(Phe/Trp/Tyr)-Ile-(Leu/Val/Ala/Met)-(Ser/Thr)-(Ile/Val/Ala/Met)-C

SEQ ID No 3

In another preferred embodiment of the first aspect of the present invention, X3 is valine. Thus another preferred polypeptide includes the sequence of SEQ ID NO 4 -

N-(Phe/Trp/Tyr)-(Ile/Leu/Val/Ala/Met)-Val-(Ser/Thr)-(Ile/Val/Ala/Met)-C

SEQ ID No 4

In another preferred polypeptide X4 is serine and so includes the sequence of SEQ ID NO 5 -

N-(Phe/Trp/Tyr)-(Ile/Leu/Val/Ala/Met)-(Leu/Val/Ala/Met)-Ser-(Ile/Val/Ala/Met)-C

SEQ ID No 5

In another preferred embodiment of the first aspect of the present invention, X4 is threonine. Thus another preferred polypeptide includes the sequence of SEQ ID NO 29 -
In another preferred embodiment of the first aspect of the present invention, X₅ is isoleucine. Thus another preferred polypeptide includes the sequence of SEQ ID NO 6 -

N-(Phe/Trp/Tyr)-(Ile/Leu/Val/Ala/Met)-(Leu/Val/Ala/Met)-Thr-(Ile/Val/Ala/Met)-C

SEQ ID No 29

More preferably at least 2, even more preferably at least 3, yet more preferably at least 4 of X₁, X₂, X₃, X₄ and X₅ are as defined in the preferred embodiments above.

The motif may be inserted into the leader sequence (i.e. as an addition), or can be included as a substitute for 1, 2, 3, 4, 5 or more contiguous amino acids within the leader sequence.

In one preferred embodiment, the motif is included in the leader sequence as a substitution for naturally occurring amino acids. In other words, the amino acids of the motif are included in the place of five contiguous amino acids that were, or would have been, present in the leader sequence prior to its optimisation according to the present invention.

It is preferable that, where the motif is included in the leader sequence as a substitution then X₄ is the naturally occurring amino acid, or a variant thereof. In other words, preferably only X₁, X₂, X₃ and X₅ are substituted, whilst X₄ is maintained unchanged, or simply changed to a variant, preferably as a conservative substitution as defined below, of the natural amino acid at that position.

In a particularly preferred embodiment of the first aspect of the present invention, X₁ is phenylalanine, X₂ is isoleucine, X₃ is valine, X₄ is serine and X₅ is isoleucine. Thus in a particularly preferred embodiment of the first aspect of the invention, there is provided a polypeptide which includes the sequence of SEQ ID No 7 -

N-Phe-Ile- Val-Ser-Ile-C

SEQ ID No 7

In the above schemes, "N" and "C" denote the orientation of the polypeptide sequence, and are not intended to be limited in their interpretation to the actual termini; in other words, the polypeptide sequence may be joined (e.g. fused, conjugated or ligated), to one or more other polypeptide sequences at either the N-, or C- ends, or most usually at both ends.

A polypeptide according to the first aspect of the invention comprises the sequence of a mature desired protein, heterologous to the leader sequence. A mature desired protein sequence is the primary amino acid sequence that will be present in the expression product following post-translational processing by the expression system in which the polypeptide of the invention is expressed. The desired protein is preferably suitable for secretion from a cell in which the polypeptide of the invention is expressed.

The desired protein is heterologous to the leader sequence. In other words, the polypeptide of the first aspect of the present invention does not include naturally occurring proteins that have, in their leader sequences, the motif -X₁-X₂-X₃-X₄-X₅- as defined above. In a preferred embodiment, the polypeptide of the first aspect of the present invention does not include any naturally occurring protein that has the motif -X₁-X₂-X₃-X₄-X₅- as defined above at any position. In this context, the term "naturally occurring" refers to proteins encoded by naturally occurring organisms that have not been modified by recombinant technology, site-directed mutagenesis or equivalent artificial techniques that requires human intervention.

The desired protein may comprise any sequence, be it natural protein (including a zymogen), polypeptide or peptide, or a variant, or a fragment (which may, for example, be a domain) of a natural protein, polypeptide or peptide; or a totally synthetic protein, polypeptide or peptide; or a single or multiple fusion of different proteins, polypeptides or peptides (natural or synthetic). Such proteins can be taken, but not exclusively, from the lists provided in WO 01/79258,
WO 01/79271, WO 01/79442, WO 01/79443, WO 01/79444 and WO 01/79480, or a variant or fragment thereof. Although these patent applications present the list of proteins in the context of fusion partners for albumin, the present invention is not so limited and, for the purposes of the present invention, any of the proteins listed therein may be presented alone or as fusion partners for albumin, the Fc region of immunoglobulin, transferrin, or any other protein as a desired polypeptide.

[0026] Preferred examples of a desired protein for expression by the present invention includes albumin, transferrin, lactoferrin, endostatin, angiostatin, collagens, immunoglobulins, Fab' fragments, F(ab')2, ScAb, ScFv, interferons, IL 10, IL11, IL2, interferon-α species and sub-species, interferon-β species and sub-species, interferon-γ species and sub-species, IL1-receptor antagonist, EPO, TPO, prosapptide, cytochrome-N, 5-helix, T20 peptide, T1249 peptide, HIV gp41, HIV gp120, fibrinogen, urokinase, prourokinase, tPA (tissue plasminogen activator), hirudin, platelet derived growth factor, parathyroid hormone, proinsulin, insulin, insulin-like growth factor, calcitonin, growth hormone, transforming growth factor-β, tumour necrosis factor, G-CSF, GM-CSF, M-CSF, coagulation factors in both pre and active forms, including but not limited to plasminogen, fibrinogen, thrombin, pre-thrombin, pro-thrombin, von Willebrand's factor, α1-antithrypsin, plasminogen activators, Factor VII, Factor VIII, Factor IX, Factor X and Factor XIII, nerve growth factor, LACI (lipoprotein associated coagulation inhibitor, also known as tissue factor pathway inhibitor or extrinsic pathway inhibitor), platelet-derived endothelial cell growth factor (PD-ECGF), glucose oxidase, serum cholinesterase, apoprotein, amyloid precursor, inter-alpha trypsin inhibitor, antithrombin III, apo-lipoprotein species, Protein C, Protein S, a variant or fragment of any of the above.

[0027] A "variant", in the context of a desired protein, refers to a protein wherein at one or more positions there have been amino acid insertions, deletions, or substitutions, either conservative or non-conservative, provided that such changes result in a protein whose basic properties, for example enzymatic activity or receptor binding (type of and specific activity), thermostability, activity in a certain pH-range (pH-stability) have not significantly been changed. "Significantly" in this context means that one skilled in the art would say that the properties of the variant may still be different but not would be unobvious over the ones of the original protein.

[0028] By "conservative substitutions" is intended combinations such as Val, Ile, Leu, Ala, Met; Asp, Glu; Asn, Gln; Ser, Thr, Gly, Ala; Lys, Arg, His; and Phe, Tyr. Preferred conservative substitutions include Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr, Lys, Arg; and Phe, Tyr.

[0029] A "variant" typically has at least 25%, at least 50%, at least 60% or at least 70%, preferably at least 80%, more preferably at least 90%, even more preferably at least 95%, yet more preferably at least 99%, preferably at least 99.5% sequence identity to the polypeptide from which it is derived.

[0030] The percent sequence identity between two polypeptides may be determined using suitable computer programs, for example the GAP program of the University of Wisconsin Genetic Computing Group and it will be appreciated that percent identity is calculated in relation to polypeptides whose sequence has been aligned optimally.

[0031] The alignment may alternatively be carried out using the Clustal W program (Thompson et al., (1994) Nucleic Acids Res., 22(22), 4673-80). The parameters used may be as follows:

- Fast pairwise alignment parameters: K-tuple(word) size; 1, window size; 5, gap penalty; 3, number of top diagonals; 5. Scoring method: x percent.
- Multiple alignment parameters: gap open penalty; 10, gap extension penalty; 0.05.
- Scoring matrix: BLOSUM.

[0032] Such variants may be natural or made using the methods of protein engineering and site-directed mutagenesis as are well known in the art.

[0033] A "fragment", in the context of a desired proteins, refers to a protein wherein at one or more positions there have been deletions. Thus the fragment may comprise at most 5, 10, 20, 30, 40 or 50% of the complete sequence of the full mature polypeptide. Typically a fragment comprises up to 60%, more typically up to 70%, preferably up to 80%, more preferably up to 90%, even more preferably up to 95%, yet more preferably up to 99% of the complete sequence of the full desired protein. Particularly preferred fragments of a desired protein comprise one or more whole domains of the desired protein. For example, the desired protein may be albumin. Albumin has three domains. A particularly preferred fragment of albumin may contain one or two domains and will thus typically comprise at least 33% or at least 66% of the complete sequence of albumin.

[0034] Albumin and transferrin, or variants or fragments thereof, are particularly preferred as a desired protein, especially when they are of human origin, i.e. they have same sequence as that found in the naturally produced human protein.

[0035] The term "human albumin" is used herein to denote material which is indistinguishable from human serum albumin or which is a variant or fragment thereof. By "variant" we include insertions, deletions and substitutions, either conservative or non-conservative, where such changes do not substantially alter the oncotic, useful ligand-binding or immunogenic properties of albumin. For example we include naturally-occurring polymorphic variants of human albumin or human albumin analogues disclosed in EP-A-322 094. Generally, variants or fragments of human albumin will have at least 10% (preferably at least 50%, 80%, 90% or 95%) of human serum albumin's ligand binding activity (for example...
that is secreted, compared with any mature protein that is produced intracellularly. The host cell with appropriate DNA constructs and measuring the amount of the mature protein (for example, human albumin) secreted from the cell. Secretion of a mature polypeptide from a cell can be determined, for example, by providing a leader sequence characterised in that it includes a secretion pre sequence that includes the motif as defined above without a leader sequence will not. However, the present invention contemplates circumstances wherein different leader sequences will have different levels of efficiency. Thus a leader sequence may cause at least 10%, 20%, 30 or 40% or 50%, yet more preferably at least 98%, most preferably at least 99% of the mature protein produced by the cell to be secreted, compared with any variant or fragment thereof. A "variant" includes insertions, deletions and substitutions, either conservative or non-conservative, where such changes do not substantially alter the useful ligand-binding or immunogenic properties of transferrin. For example we include naturally-occurring polymorphic variants of human transferrin or human transferrin analogues. Generally, variants or fragments of human transferrin will have at least 50% (preferably at least 80%, 90% or 95%) of human transferrin's ligand binding activity (for example iron-binding), weight for weight. The iron binding activity of transferrin or a test sample can be determined spectrophotometrically by 470nm:280nm absorbance ratios for the proteins in their iron-free and fully iron-loaded states. Reagents should be iron-free unless stated otherwise. Iron can be removed from transferrin or the test sample by dialysis against 0.1M citrate, 0.1M acetate, 10mM EDTA pH4.5. Protein should be at approximately 20mg/mL in 100mM HEPES, 10mM NaHCO3 pH8.0. Measure the 470nm: 280nm absorbance ratio of apo-transferrin (Calbiochem, CN Biosciences, Nottingham, UK) diluted in water so that absorbance at 280nm can be accurately determined spectrophotometrically (0% iron binding). Prepare 20mM iron-nitrilotriacetate (FeNTA) solution by dissolving 191mg nitrotriacetic acid in 2mL 1M NaOH, then add 2mL 0.5M ferric chloride. Dilute to 50mL with deionised water. Fully load apo-transferrin with iron (100% iron binding) by adding a sufficient excess of freshly prepared 20mM FeNTA, then dialyse the holo-transferrin preparation completely against 100m HEPES, 10mM NaHCO3 pH8.0 to remove remaining FeNTA before measuring the absorbance ratio at 470nm:280nm. Repeat the procedure using test sample, which should initially be free from iron, and compare final ratios to the control. Additionally, single or multiple heterologous fusions of any of the above; or single or multiple heterologous fusions to albumin, transferrin or immunoglobulins or a variant or fragment of any of these may be used. Such fusions include albumin N-terminal fusions, albumin C-terminal fusions and co-N-terminal and C-terminal albumin fusions as exemplified by WO 01/79271, and transferrin N-terminal fusions, transferrin C-terminal fusions, and co-N-terminal and C-terminal transferrin fusions. In a more preferred embodiment a polypeptide according to a first aspect of the present invention comprises a leader sequence characterised in that it includes a secretion pre sequence that includes the motif as defined above by the first aspect of the present invention. The leader sequence is usually, although not necessarily, located at the N-terminus of the primary translation product and is generally, although not necessarily, cleaved off the protein during the secretion process, to yield the mature "desired" protein. A secretion leader sequence is usually, although not necessarily, an N-terminal sequence of amino acids that causes the polypeptide of which it forms part to be secreted from a host cell in which it is produced. Secretion is defined by the co-translational of post-translation translocation of a protein from the cytoplasmic compartment across a phospholipid bilayer, typically, but not exclusively the endoplasmic reticulum of eukaryotic organisms or the plasma membrane of prokaryotic organisms. The secreted protein may be retained within the confines of the cell (typically, but not exclusively, within the endoplasmic reticulum, Golgi apparatus, vacuole, lysosome or periplasmic space) or it may be secreted from the cell into the culture medium. A sequence acts as a secretion leader sequence if, in comparison to an equivalent polypeptide without the secretion pre sequence, it causes more of that polypeptide to be secreted from the host cell in which it is produced. Generally speaking, a polypeptide with a leader sequence will be secreted whereas a polypeptide without a leader sequence will not. However, the present invention contemplates circumstances wherein different leader sequences will have different levels of efficiency. Thus a leader sequence may cause at least 10%, 20%, 30 or 40% or 50%, typically at least 60% or 70%, preferably at least 80%, more preferably at least 90%, even more preferably at least 95%, yet more preferably at least 98%, most preferably at least 99% of the mature protein produced by the cell to be secreted from the cell. Secretion of a mature polypeptide from a cell can be determined, for example, by providing a host cell with appropriate DNA constructs and measuring the amount of the mature protein (for example, human albumin) that is secreted, compared with any mature protein that is produced intracellularly.
A leader sequence according to the present invention will provide for the above mentioned levels of secretion when the host cell is a yeast cell (e.g. *Saccharomyces cerevisiae* or *Pichia pastoris*). Secretion of a mature polypeptide from a yeast host cell can be determined, for example, by methods such as those set out in the examples below.

Accordingly, secretion can be measured by comparing the level of secretion of a protein comprising a test leader sequence to the level of secretion of a protein comprising a control leader sequence. In order to determine whether a given sequence (the "test sequence") is able to achieve a given level of secretion, a first protocol may be used in which a 'starter' plasmid, typically a yeast disintegration vector of the type described in EP 0 286 422, having the LEU2 gene and a polynucleotide encoding rHA with a modified leader sequence as defined by SEQ ID NO:26 operably linked to functional yeast regulatory regions, such as a PRB1 promoter and an ADH1 terminator as described below, is modified to include a polynucleotide sequence that encodes the test sequence in place of an equivalent region of the leader sequence, thereby to provide a test plasmid. As a first control, the unmodified 'starter' plasmid encoding the leader sequence described in WO 90/01063 is used. *Saccharomyces cerevisiae* strain AH22 cit9 (Hinnen et al, 1978, Proc. Natl. Acad. Sci. USA, 75(4), 1929-33; Mead et al, 1986, Mol. Gen. Genet., 205, 417), His4 reverted, is used as a test host. A HIS4 revertant (i.e. His+ of AH22 (leu2, his4, can1) can be obtained by culturing sufficient AH22 cells on BMM agar, supplemented with 0.002% (w/v) leucine, until colonies appear. The colonies are tested to confirm that that are Leu- and His+ (i.e. AH22 His+ (leu2, can1)) by plating onto BMMD agar, supplemented with 0.002% (w/v) leucine (plate 1), BMMD agar, supplemented with 0.002% (w/v) leucine and, supplemented with 0.002% (w/v) histidine (plate 2), and BMMD agar (plate 3). AH22 His+ (leu2, can1) isolates will grow on plate 1 and plate 2, but will not grow on plate 3. The test host is transformed to leucine prototrophy with the test and control plasmids. Transformants are patched out onto Buffered Minimal Medium (BMM, described by Kerry-Williams, S.M. et al. (1998) Yeast 14, 161-169) containing 2% (w/v) glucose (BMMD) and incubated at 30°C until grown sufficiently for further analysis. Transformants are cultivated in high cell density fermentation according to a fill & draw procedure, in a medium and using control parameters as described for the fed-batch procedure in WO 96/37515: upon completion of the feed phase of the fed-batch culture procedure, 90% of the culture volume is removed from the fermenter vessel. Batch medium is added to the remaining 10% volume of the culture (maintaining pH control) prior to the initiation of feed addition, using the medium and control parameters described in WO 96/37515. The procedure of fill & draw can be repeated for an unlimited number of cycles. The human albumin productivity (YP/S) of the transformants containing test and control plasmids are assessed by scanning densitometry of SDS-PAGE of cell free whole culture. YP/S represents the ratio of human albumin protein (mg) per gram of sucrose added to the culture during fermentation.

A leader sequence according to the present invention may obtain a level of secretion, as determined by YP/S as measured by the above first protocol, that is at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200%, 210%, 220%, 230%, 240%, 250%, 260%, 270%, 280%, 290%, 300%, 310%, 320%, 330%, 340%, 350%, 360%, 370%, 380%, 390%, 400%, 410%, 420%, 430%, 440%, 450%, 460%, 470%, 480%, 490% or 500% higher than the level of secretion obtained by the first control when the test and first control transformants are cultured for comparable lengths of time with comparable fermenter configurations. Thus a leader sequence according to the present invention may demonstrate a level of secretion that is up to 400%, such as 408%, or at least 500%, 550%, 600%, 650%, 700%, 750% or more higher than the first control. It is particularly preferred that a leader sequence according to the present invention can obtain a level of secretion at least 5%, such as 6%, or at least 10%, such as 12%, 13%, 14%, 15% or 16% higher than the level of secretion obtained by the second control.
In one embodiment, a leader sequence according to the present invention will obtain a level of secretion as defined above in respect of the first protocol. In another embodiment, a leader sequence according to the present invention will obtain a level of secretion as defined above in respect of the second protocol. In a particularly preferred embodiment, a leader sequence according to the present invention will obtain a level of secretion as defined above in respect of both the first and second protocols.

Solubilised proteins from the cell biomass and secreted proteins in the culture supernatant can be analysed by:

1. Gel permeation high pressure liquid chromatography.
2. Densitometry of SDS-PAGE
3. Rocket immunoelectrophoresis

The amount of the desired protein, secreted and intracellular, can be quantified relative to a standard curve of the desired protein and normalised to the amount of biomass as known by those skilled in the art.

The secretion pre sequence is derived from a human albumin secretion pre sequence and X₁, X₂, X₃, X₄ and X₅ are at positions -20, -19, -18, -17 and -16, respectively, in place of the naturally occurring amino acids at those positions, wherein the numbering is such that the -1 residue is the C-terminal amino acid of the native albumin secretion pre sequence and where X₁, X₂, X₃, X₄ and X₅ are amino acids as defined above.

For example, when the above mentioned numbering is applied to the sequence of the human albumin secretion pre sequence (as disclosed, for example in WO 90/01063), the following is obtained:

\[
\begin{array}{cccccccccccc}
N & - & Met & Lys & Trp & Val & Ser & Phe & Ile & Ser & Leu & Leu \\
\end{array}
\]

\[
\begin{array}{cccccccccccc}
Phe & Leu & Phe & Ser & Ser & Ala & Tyr & Ser & - & C \\
-14 & -13 & -12 & -11 & -10 & -9 & -8 & -7 \\
\end{array}
\]

The X₁-X₅ pentapeptide may be fused at its N-terminal end, directly or indirectly, to the C-terminal end of the following sequence SEQ ID NO 8 -

N-Met-Lys-Trp-Val-C

SEQ ID No 8

or a conservatively substituted variant thereof, namely -

\[
\text{N-Met-(Lys/Arg/His)-(Phe/Trp/Tyr)-(Ile/Leu/Val/Ala/Met)-C}
\]

SEQ ID No. 33

Additionally or alternatively it may be fused at its C-terminal end, directly or indirectly, to the N-terminal end of at least one of the following sequences -

N-Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-C

SEQ ID No 9

or a conservatively substituted variant thereof, namely -
N-(Ile/Leu/Val/Ala/Met)-(Phe/Trp/Tyr)-(Ile/Leu/Val/Ala/Met)-(Phe/Trp/Tyr)-(Ser/Thr/Gly/Tyr/Ala)-(Ser/Thr/Gly/Tyr/Ala)-(Ile/Leu/Val/Ala/Met)-(Phe/Trp/Tyr)-(Ser/Thr/Gly/Tyr/Ala)-C

SEQ ID No. 10

or

N-Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-Arg-Ser-Leu-Asp-Lys-Arg-C

SEQ ID No. 11

or

N-Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-Arg-Gly-Val-Phe-Arg-Arg-C

SEQ ID No. 30

[0053] The sequence given in SEQ ID No 9 represents the final nine amino acids of the natural human albumin presequence. In the case of SEQ ID No 11, this is fused to the final six amino acids of one of the two principal fused leader sequences of WO 90/01063 and, in the case of SEQ ID No. 30, SEQ ID No. 9 is fused to the final six amino acids of the natural human albumin presequence.

[0054] Preferably, in each case, X1 is F, X2 is I, X3 is V, X4 is S or T and X5 is L.

[0055] In a preferred embodiment, the pentapeptide is fused at its N-terminal to the C-terminal of sequence of SEQ ID NO 8 or a conservatively substituted variant thereof and is fused at its C-terminal to the N-terminal of the sequence of SEQ ID NO 9, a conservatively substituted variant thereof, SEQ ID No. 10, 11 or 30, thereby to form, for example, one of the following sequences -

N-Met-Lys-Trp-Val-X1-X2-X3-X4-X5-
Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-C

SEQ ID No 12

or

N-Met-Lys-Trp-Val-X1-X2-X3-X4-X5-(Ile/Leu/Val/Ala/Met)-(Phe/Trp/Tyr)-(Ile/Leu/Val/Ala/Met)-(Phe/Trp/Tyr)-(Ser/Thr/Gly/Tyr/Ala)-(Ser/Thr/Gly/Tyr/Ala)-(Ile/Leu/Val/Ala/Met)-(Phe/Trp/Tyr)-(Ser/Thr/Gly/Tyr/Ala)-C

SEQ ID No 13

or

N-Met-Lys-Trp-Val-X1-X2-X3-X4-X5-
Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-Arg-Ser-Leu-Asp-Lys-Arg-C

SEQ ID No 14
N-Met-Lys-Trp-Val-X₁-X₂-X₃-X₄-X₅-
Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-Arg-Gly-Val-Phe-Arg-Arg-C

SEQ ID No 31

wherein X₁-X₅ are as defined above, or a conservatively substituted variant thereof, as defined above. An especially preferred embodiment has, as the secretion pre sequence, the sequence of SEQ ID NO 28.

N-Met-Lys-Trp-Val-Phe-Ile-Val-Ser-Ile-Leu-Phe-Leu-Phe-Ser-Ser-Ala-
Tyr-Ser-C

SEQ ID No 28

i.e. the pre sequence is derived from the human serum albumin secretion pre sequence, X₁, X₂, X₃, X₄ and X₅ are at positions -20, -19, -18, -17 and -16, and X₁, X₂, X₃, X₄ and X₅ are as defined by SEQ ID No.7.

[0056] As is apparent from above, a secretion pre sequence as defined above, such as the sequences of SEQ ID Nos 12 or 28, may be combined with secretion pro sequences to form functional pre-pro secretion sequences. In a preferred embodiment, a pre sequence motif is fused by a peptide bond at its C-terminal end to the N-terminal amino acid of a secretion pro sequence motif, thereby to form a pre-pro sequence motif. It may be preferable to use a pro sequence derived from the immature version of the mature protein to which the leader sequence is, or is intended to be, attached. It may also be preferable to use the pro sequence that is associated in nature with the unmodified pre sequence or a pro sequence, or part thereof, from an related leader.

[0057] Preferably, the pro sequence terminates at its C-terminus in a dibasic pair of amino acids, i.e. each is Lys or Arg.

[0058] Typically the secretion pro sequence motif is an albumin secretion pro sequence or variant thereof, such a variant including the dibasic pair of amino acids and having only conservative substitutions at the other positions, usually a human albumin secretion pro sequence, i.e. having the sequence N-Arg-Gly-Val-Phe-Arg-Arg-C or variant thereof. In another preferred embodiment the pro sequence comprises the sequence of the whole or part of the yeast MFA₁ secretion pro sequence, i.e. N-Ser-Leu-Asp-Lys-Arg-C or variant thereof as defined for the albumin pro sequence.

[0059] In comparison with the corresponding parts of the leader defined in WO 90/01063 and the human albumin leader, a polypeptide of the present invention has at least four amino acid changes namely Ser-20Phe or Trp or Tyr; Phe-19Ile or Leu or Val or Ala or Met; Ile-18Leu or Val or Ala or Met; and Leu-16Ile or Val or Ala or Met, where the notation means that, taking the first-named mutation as an example, the serine residue at position -20 (i.e. minus twenty relative to the N-terminus of the mature protein that is to be secreted using the leader sequence) is changed to a phenylalanine residue. This is exemplified in Fig. 1.

[0060] One preferred pre-pro sequence comprises the sequence:
MKWVFIVSILFLFSSAYSRY₁Y₂Y₃Y₄Y₅

wherein Y₁ is Gly or Ser, Y₂ is Val or Leu, Y₃ is Phe or Asp, Y₄ is Arg or Lys and Y₅ is Arg or Lys.

[0061] In a preferred embodiment, Y₁ is Gly, Y₂ is Val and Y₃ is Phe. In another preferred embodiment Y₁ is Ser, Y₂ is Leu and Y₃ is Asp.

[0062] Typically Y₄ is Arg and Y₅ is Arg. Alternatively it is preferred if Y₄ is Arg and Y₅ is Arg. Another preferred alternative is where Y₄ is Lys and Y₅ is Lys. Y₄ may also be Arg where Y₅ is Lys.

[0063] An especially preferred embodiment has, as the secretion prepro sequence, the sequence of SEQ ID NO 32
A second aspect of the invention provides an isolated polynucleotide having a sequence that encodes the motif as defined by the first aspect of the invention.

As used herein, the term "isolated" includes the meaning that the polynucleotide, where it is a DNA molecule, is in isolation from at least most of the chromosome on which it is naturally found and, where it is an RNA molecule, is in isolation from an intact cell in which it is naturally transcribed. In other words, the polynucleotide is not claimed in a form in which it has previously existed, such as in nature. Thus, a polynucleotide according to the second aspect of the invention includes a polynucleotide that has been cloned into a bacterial or fungal vector, such as a plasmid, or into a viral vector, such as a bacteriophage. Preferably such clones are in isolation from clones constituting a DNA library of the relevant chromosome.

The linear amino acid sequence can be reverse translated into a DNA sequence using the degenerate standard genetic code (Fig.2) in which most amino acids are encoded by more than one trinucleotide codon.

For example, a DNA sequence encoding the peptide defined as SEQ ID 1 would be deduced to be:

5'-TTY/TGG/TAY)-(ATH/TTR or CTN/GTN/GCN/ATG)-(TTR or CTN/GTN/GCN/ATG)-(AGY or TCN/ACN)-(ATH or CTN/GTN/GCN/ATG)-3'

SEQ ID No 16

where " 3' " and " 5' " denote the orientation of the polynucleotide sequence, rather than the actual termini; in other words, the polynucleotide sequence may be joined (e.g. fused or ligated) to other polynucleotide sequences at either end or both ends, and wherein Y, R, H and N are as defined in Fig. 2.

Using the same conversion procedure the DNA sequence:

5'-TTY-ATH-GTN-(TCN or AGY)-ATH-3'

SEQ ID No 15

would be deduced to encode the polypeptide of SEQ ID No 7.

In the case of a polynucleotide sequence comprising a sequence that encodes a naturally occurring mature protein, such a human albumin, this can be either the naturally occurring coding sequence, such as the human albumin gene sequence, or a complementary DNA sequence (cDNA) or a cDNA containing one or more introns.

Further sequence modifications may also be introduced, for example into the coding region. A desirable way to modify the DNA encoding the polypeptide of the invention is to use the polymerase chain reaction as disclosed by Saiki et al (1988) Science 239, 487-491. In this method the DNA to be enzymatically amplified is flanked by two specific oligonucleotide primers which themselves become incorporated into the amplified DNA. The said specific primers may contain restriction endonuclease recognition sites which can be used for cloning into expression vectors using methods known in the art.

The polynucleotide encoding a leader sequence of the invention is most conveniently made by chemical synthesis of an oligonucleotide, followed by ligation to the other elements of the genetic construct, by methods that are well known in this art and described in more detail below.

Where it is desirable to modify the polynucleotide that encodes mature albumin, this may be most conveniently achieved by site-directed mutagenesis or PCR mutagenesis, starting from the natural cDNA sequence, or from assembling synthetic oligonucleotides. Again, such techniques are standard in this art and are in any case set out in more detail below.
Modification to the coding sequence can be advantageous because, within a particular organism, the polynucleotide sequences encoding some highly expressed proteins favour some codons over others for a particular amino acid; this is called codon bias. In a preferred embodiment of a second aspect of the invention the standard genetic code can be reduced to the preferred codons for the host organism of choice. In an especially preferred embodiment of the second aspect of the invention the standard genetic code can be reduced to the preferred codons of yeast. (See Table 4 of Sharp and Crowe (1991) Yeast 7, 657-678.) Advantageously this list of preferred yeast codons is modified by inclusion of the asparagine codon 5'-GAT-3' (Fig.3).

Using the peptide sequence of SEQ ID No 1 as an example, the codon biased DNA sequence encoding this peptide in yeast may be deduced to be:

5'-(TTC/TGG/TAC)-(ATY/TTG/GTY/GCT/ATG)-(TTG/GTY/GCT/ATG)-(TCY/ACY)-(ATY/GTY/GCT/ATG)-5'

SEQ ID No 17

Using the same conversion procedure the codon-biased degenerate DNA sequence:

5'-TTC-ATY-GTY-TCY-ATY-3'

SEQ ID No 18

would be deduced for the especially preferred polypeptide motif having the sequence of SEQ ID No 7, although the most preferred codon-biased DNA sequence encoding a polypeptide motif having the sequence of SEQ ID No. 7 is -TTCATCGTCTCCATT

SEQ ID No. 344

Using the genetic code given in Fig.2 or the preferred codon bias tables available for the intended host or the preferred codon bias given in Fig.3, the same conversion procedure can be used to convert any desired amino acid sequence into a partially redundant polynucleotide sequence. The amino acid sequences, which can be converted into a DNA sequence by this method can be taken from, but not limited to, polypeptides according to the first aspect of the invention. For example, the sequence of a coding region for mature human albumin can be derived in this way. EP 308 381 discloses a partially yeast-codon-optimised coding sequence for human albumin. SEQ ID No. 20 herein is further such sequence. Advantageously, where the DNA sequence redundancy permits, restriction sites can be introduced at domain and sub-domain boundaries, without perturbing the encoded amino acid sequence (or the codon bias if Fig.3 is used).

The remaining DNA sequence redundancies can be resolved and the number of occurrences of alternative codons equalised for each amino acid with redundant DNA sequences. Advantageously, DNA sequences representing possible transcription terminator sequences can be removed or reduced where possible by utilising the DNA sequence redundancy of the degenerate codons. Finally the balance of alternative codons for amino acids with redundant DNA sequences can be re-equalised but without conflicting with the previous modifications

A polynucleotide according to the second aspect of the invention can be directly or indirectly fused to one or more other nucleotide sequences at its 5' and/or 3' ends, for example to form a complete gene or expression cassette. Thus, the expression cassette will desirably also contain sites for transcription initiation and termination, and in the transcribed region, a ribosome binding site for translation initiation. (Hastings et al, WO 98/16643, published 23 April 1998.) Accordingly, the second aspect of the present invention includes a polynucleotide comprising a DNA sequence that is a contiguous or non-contiguous fusion of a DNA encoding a heterologous protein with either a DNA sequence encoding a polypeptide according to the first aspect of the present invention, particularly wherein the desired protein is a variant or fragment thereof. In this context, the term "heterologous protein" means that it is not the same as the "desired protein", i.e. does not form a homodimer.

Accordingly, the polynucleotide may be directly or indirectly fused to a promoter (an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur) at its 5' end and/or to other regulatory sequences, such as, at its 3' end, translation termination sequences. Thus a polynucleotide may be operably linked to one or more regulatory regions, usually transcription regulatory regions. By "operably linked" is meant that the regulatory region is linked in such a way that it is able to exert an effect on the polynucleotide sequence. The choice of which regulatory region to use will be partially dependent upon the expected host (i.e. the intended expression system) and the selection of the preferred sequence will be known to those skilled in the art

Many expression systems are known, including systems employing: bacteria (eg. Bacillus subtilis or Escherichia
coli) transformed, for example, recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeasts (eg. Saccharomyces cerevisiae or Pichia pastoris) transformed with, for example, yeast expression vectors; insect cell systems transformed with, for example, viral expression vectors (eg. baculovirus); plant cell systems transfected with, for example viral or bacterial expression vectors; animal cell systems, either in cell culture, transgenic or as gene therapy, transfected with, for example, adenovirus expression vectors. The host cell is preferably a yeast (and most preferably a Saccharomyces species such as S. cerevisiae or a Pichia species such as P. pastoris).

Accordingly, a third aspect of the present invention provides a host cell transformed with a polynucleotide according to the second aspect of the present invention. The host cell can be either prokaryotic or eukaryotic. Bacterial cells are preferably prokaryotic host cells, particularly if they can secrete proteins, as can some species of Bacillus and Escherichia. Preferred eukaryotic host cells include plants, fungi, yeast and animal cells, preferably vertebrate cells, more preferably mammalian cells, such as those from a mouse, rat, cow, sheep, goat, pig, buffalo, yak, horse or other domesticated animal, monkey or human. Suitable human cells include cells from a human fibroblastic cell line. Thus a host cell may be a transgenic cell of a mammal in situ, and may thus be the result of a gene therapy approach or of the production of a transgenic individual. Accordingly the present invention contemplates the production of whole transgenic plants, which further preferably retain a stable and heritable transgenic phenotype. Suitable promoters, transformation protocols and culture conditions for Pichia can be found in US 5 986 062.

Exemplary genera of bacterial hosts include E.coli and Bacillus subtilis.

Exemplary genera of plant hosts include spermatophytes, pteridophytes (e.g. ferns, clubmosses, horsetails), bryophytes (e.g. liverworts and mosses), and algae. Typically the plant host cell will be derived from a multicellular plant, usually a spermatophyte, such as a gymnosperm or an angiosperm. Suitable gymnosperms include conifers (e.g. pines, larches, firs, spruces and cedars), cycads, yews and ginkos. More typically the plant host cell is the cell of an angiosperm, which may be a monocotyledonous or dicotyledonous plant, preferably a crop plant. Preferred monocotyledonous plants include maize, wheat, barley, sorghum, onion, oats, orchard grass and other Pooidae. Preferred dicotyledonous crop plants include tomato, potato, sugarbeet, cassava, cruciferous crops (including oilseed rape), linseed, tobacco, sunflower, fibre crops such as cotton, and leguminous plants such as peas, beans, especially soybean, and alfalfa. The host cell may thus be an autonomous cell, for example the cell of a unicellular plant or a cell maintained in cell culture, or it may be a cell in situ in a multicellular plant. Accordingly the present invention contemplates the production of whole transgenic plants, which further preferably retain a stable and heritable transgenic phenotype.

Exemplary genera of fungal hosts include Aspergillus (e.g. A. niger and A. oryzae), Streptomyces, Penicillium and yeasts. Exemplary genera of yeast contemplated to be useful in the practice of the present invention are Pichia (Hansenula), Saccharomyces, Kluveromyceses, Candida, Torulopsis, Torulaspora, Schizosaccharomyces, Citeromyces, Pachysolen, Debaromyces, Metschnikowia, Rhodosporidium, Leucosporidium, Botryosphaeria, Sporidiobolus, Endomyces, and the like. Preferred genera are those selected from the group consisting of Pichia (Hansenula), Saccharomyces, Kluveromyces and Yarrowia. Examples of Saccharomyces spp. are S. cerevisiae, S. italicus and S. rouxii. Examples of Kluveromyces spp. are K. fragilis and K. lactis. Examples of Pichia (Hansenula) are P pastoris, P. anomala and P. capsulata. Y. lipolytica is an example of a suitable Yarrowia species. Yeast host cells include YPH499, YPH500 and YPH501 which are generally available from Stratagene Cloning Systems, La Jolla, CA 92037, USA.

Preferred mammalian host cells include Chinese hamster ovary (CHO) cells available from the ATCC as CCL61, NIH Swiss mouse embryo cells NIH/3T3 available from the ATCC as CRL 1658, and monkey kidney-derived COS-1 cells available from the ATCC as CRL 1650. Preferred insect cells are Sf9 cells which can be transfected with baculovirus expression vectors.

As discussed above, the choice of polynucleotide regulatory region will be partly dependent on the nature of the intended host.

Promoters suitable for use in bacterial host cells include the E. coli lacI and lacZ promoters, the T3 and T7 promoters, the gpt promoter, the phage λ PR and PL promoters, the phoA promoter and the trp promoter. Promoter sequences compatible with exemplary bacterial hosts are typically provided in plasmid vectors containing convenient restriction sites for insertion of a DNA segment of the present invention.

Eukaryotic promoters include the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters and the promoters of retroviral LTRs. Other suitable promoters will be known to those skilled in the art.

Suitable promoters for S. cerevisiae include those associated with the PGK1 gene, GAL1 or GAL10 genes, CYC1, PHO5, TRP1, ADH1, ADH2, the genes for glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, triose phosphate isomerase, phosphoglucose isomerase, glucokinase, α-mating factor pheromone, a-mating factor pheromone, the PRB1 promoter, the GPD1 promoter, and hybrid promoters involving hybrids of parts of 5′ regulatory regions with parts of 5′ regulatory regions of other promoters or with upstream activation sites (e.g. the promoter of EP-A-258 067).

Convenient regulatable promoters for use in Schizosaccharomyces pombe, another suitable host cell, are the thiamine-repressible promoter from the nmt gene as described by Maundrell (1990) J. Biol. Chem. 265, 10857-10864 and the glucose-repressible fbp1 gene promoter as described by Hoffman & Winston (1990) Genetics 124, 807-816.

Suitable promoters, transformation protocols and culture conditions for Pichia can be found in US 5 986 062.
For example, preparation of an HSA-producing host (or an HSA-producing strain) may be effected using a process in which a recombinant plasmid is introduced into chromosome (JP-A-3-72889 corresponding to EP-A-399455), a process in which HSA is expressed in yeast (JP-A-60-41487 corresponding to EP-A-123544, JP-A-63-39576 corresponding to EP-A-248657 and JP-A-63-74493 corresponding to EP-A-251744) and a process in which HSA is expressed in Pichia (JP-A-2-104290 corresponding to EP-A-344459). Culturing of an HSA-producing host (an HSA production process) may be carried out using known methods, such as those referred to in US 5,986,062, for example in accordance with a process disclosed in JP-A-3-83595 or JP-A-4-293495 (corresponding to EP-A-504823). The medium for culturing a transformed host may be prepared in accordance with US 5,986,062 and culturing of a host may be carried out preferably at 15 to 43°C (more preferably 20 to 30°C) for 1 to 1,000 hours, by means of static or shaking culturing or batch, semi-batch or continuous culturing under agitation and aeration in accordance with the disclosures of US 5,986,062.

[0093] Suitable transcription termination signals are well known in the art. Where the host cell is eukaryotic, the transcription termination signal is preferably derived from the 3’ flanking sequence of a eukaryotic gene, which contains proper signals for transcription termination and polyadenylation. Suitable 3’ flanking sequences may, for example, be those of the gene naturally linked to the expression control sequence used, i.e. may correspond to the promoter. Alternatively, they may be different in that case, and where the host is a yeast, preferably S. cerevisiae, then the termination signal of the S. cerevisiae ADH1 gene is preferred.

[0094] Thus a polynucleotide according to the second aspect of the present invention can be developed for any desired host by using methods such as those described above.

[0095] A DNA sequence encoding mature human albumin can be developed from DNA fusions between the native gene, cDNA or a cDNA containing one or more introns, as described above and a codon biased human albumin DNA sequence derived by the method described above.

[0096] SEQ ID No 19 is a polynucleotide sequence that comprises 22 nucleotides 5’ to the translation initiation site, a preferred polynucleotide coding sequence for the secretion leader sequence SEQ ID No. 32 and a mature human albumin coding region SEQ ID No 20. The coding sequence ends with a translation stop codon. Typically, this is TGA, TAG or TAA, although TAA is the most efficient in yeast. Preferably, further translation stop codons (preferably each is TAA), usually one or two, are included, preferably adjacent each other or with no more than 3 base pairs between each pair of stop codons. SEQ ID No 19 is flanked at both ends by appropriate cloning sites.

[0097] The polynucleotide of the second aspect of the invention may also be joined to a wide variety of other DNA sequences for introduction into an appropriate host. The companion sequence(s) will depend upon the nature of the host, the manner of the introduction of the polynucleotide into the host, and whether episomal maintenance or integration is desired. For example, the vectors can include a prokaryotic replicon, such as the Col E1 ori, for propagation in a prokaryote, even if the vector is to be used for expression in other, non-prokaryotic cell types.

[0098] Generally, a polynucleotide according to the second aspect of the invention is inserted into an expression vector, such as a plasmid, in proper orientation and correct reading frame for expression.

[0099] Thus, the polynucleotide may be used in accordance with known techniques, appropriately modified in view of the teachings contained herein, to construct an expression vector, including, but not limited to integration vectors, centromeric vectors and episomal vectors.

[0100] Thus in one embodiment of the second aspect of the invention, the polynucleotide is a vector.

[0101] Typical prokaryotic vector plasmids are: pUC18, pUC19, pBR322 and pBR329 available from Biorad Laboratories (Richmond, CA, USA); pTrc99A, pKK223-3, pKK233-3, pDR540 and pRITS available from Pharmacia (Piscataway, NJ, USA); pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16A, pNH18A, pNH46A available from Stratagene Cloning Systems (La Jolla, CA 92037, USA).

[0102] A typical mammalian cell vector plasmid is pSVL available from Pharmacia (Piscataway, NJ, USA). This vector uses the SV40 late promoter to drive expression of the cloned genes, the highest level of expression being found in T antigen-producing cells, such as COS-1 cells. An example of an inducible mammalian expression vector is pMSG, also available from Pharmacia (Piscataway, NJ, USA). This vector uses the glucocorticoid-inducible promoter of the mouse mammary tumour virus long terminal repeat to drive expression of the cloned gene.


[0104] Methods well known to those skilled in the art can be used to construct expression vectors containing the coding sequence and, for example appropriate transcriptional or translational controls. One such method involves ligation via cohesive ends. Compatible cohesive ends can be generated on the DNA fragment and vector by the action of suitable...
restriction enzymes. These ends will rapidly anneal through complementary base pairing and remaining nicks can be closed by the action of DNA ligase. 

A further method uses synthetic double stranded oligonucleotide linkers and adapters. DNA fragments with blunt ends are generated by bacteriophage T4 DNA polymerase or E.coli DNA polymerase I which remove protruding 3’ termini and fill in recessed 3’ ends. Synthetic linkers and pieces of blunt-ended double-stranded DNA which contain recognition sequences for defined restriction enzymes, can be ligated to blunt-ended DNA fragments by T4 DNA ligase. They are subsequently digested with appropriate restriction enzymes to create cohesive ends and ligated to an expression vector with compatible termini. Adapters are also chemically synthesised DNA fragments which contain one blunt end used for ligation but which also possess one preformed cohesive end. Alternatively a DNA fragment or DNA fragments can be ligated together by the action of DNA ligase in the presence or absence of one or more synthetic double stranded oligonucleotides optionally containing cohesive ends.

Synthetic linkers containing a variety of restriction endonuclease sites are commercially available from a number of sources including Sigma-Genosys Ltd, London Road, Pampisford, Cambridge, United Kingdom.

Vectors of the invention thus produced may be used to transform an appropriate host cell for the expression and production of a polypeptide comprising a sequence as defined in the first aspect of the invention. Such techniques include those disclosed in US Patent Nos. 4,440,859 issued 3 April 1984 to Rutter et al, 4,530,901 issued 23 July 1985 to Weissman, 4,582,800 issued 15 April 1986 to Crowl, 4,677,063 issued 30 June 1987 to Mark et al, 4,678,751 issued 7 July 1987 to Goeddel, 4,704,362 issued 3 November 1987 to Itakura et al, 4,710,463 issued 1 December 1987 to Murray, 4,757,006 issued 12 July 1988 to Toole, Jr. et al, 4,766,075 issued 23 August 1988 to Goeddel et al and 4,810,648 issued 7 March 1989 to Stalker.


Electroporation is also useful for transforming cells and is well known in the art for transforming yeast cell, bacterial cells and vertebrate cells. Methods for transformation of yeast by electroporation are disclosed in Becker & Guarente (1990) Methods Enzymol. 194, 182.

Physical methods may be used for introducing DNA into animal and plant cells. For example, microinjection uses a very fine pipette to inject DNA molecules directly into the nucleus of the cells to be transformed. Another example involves bombardment of the cells with high-velocity microprojectiles, usually particles of gold or tungsten that have been coated with DNA.

The desired mature protein may be extracted from the culture medium by many methods known in the art. For example, cells successfully transformed with an expression vector produce proteins displaying appropriate antigenicity. Samples of cells suspected of being transformed are harvested and assayed for the protein using suitable antibodies. Thus, in addition to the transformed host cells themselves, the present invention also contemplates a culture of those cells, preferably a monoclonal (clonally homogeneous) culture, or a culture derived from a monoclonal culture, in a nutrient medium.

Accordingly, in a fourth aspect of the present invention there is provided a cell culture comprising a cell according to the third aspect of the invention and culture medium. Typically the culture medium will contain mature polypeptide that results from the expression of a polypeptide according to the first aspect of the present invention wherein the expression system and, usually, by further translational processing, such as the removal of the pre and/or pro sequences. Methods for culturing prokaryotic host cells, such as E.coli, and eukaryotic host cells, such as mammalian cells are well known in the art. Methods for culturing yeast are generally taught in EP 330 451 and EP 361 991. Allowing host cells that have been transformed by the recombinant DNA of the invention to be cultured for a sufficient time and under appropriate conditions known to those skilled in the art in view of the teachings disclosed herein permits the expression of the polypeptide according to the first aspect of the present invention. The thus produced polypeptide may be further processed by the host cell, such that the pre and/or pro sequences are removed. Accordingly the "mature" desired protein may differ from the protein as originally translated.

Thus the invention also provides, as a fifth aspect, a process for producing a mature desired protein as defined above. The process comprises the step of culturing a cell according to the third aspect of the invention in a culture medium wherein the cell, as a result of the expression of a polypeptide as defined in the first aspect of the invention, secretes a mature desired protein, where it accumulates either in the periplasmic space, the culture medium or both, but preferably into the culture medium. The culture medium, which contains the secreted desired protein, may then be separated from the cell(s) in the cell culture. Secreted proteins associated with the cell wall can generally be disassociated therefrom using lytic enzymes under osmotic supporting (e.g. sorbitol) conditions (which gently release the secreted protein selectively). See Elango et al., J. Biol. Chem. 257: 1398-1400 (1982). Examples of lytic enzymes suitable for this purpose include lyticase, Zymolyase-60,000, and Glusulase, all of which are commercially available, for example, the case of the latter two, from Seikagaku Kogyo or Kirin Brewery, and from Boehringer Mannheim, respectively. Preferably, following the isolation of the culture medium, the mature desired protein is separated from the medium. Even more preferably the thus obtained mature desired protein is further purified.

The desired mature protein may be extracted from the culture medium by many methods known in the art. For example purification techniques for the recovery of recombinantly expressed albumin have been disclosed in: WO 92/04367, removal of matrix-derived dye; EP 464 590, removal of yeast-derived colorants; EP 319 067, alkaline precipitation and subsequent application of the albumin to a lipophilic phase; and WO 96/37515, US 5 728 553 and WO 00/44772, which describe complete purification processes. Proteins other than albumin may be purified from the culture medium by any technique that has been found to be useful for purifying such proteins, since the modified leader sequence of the invention will not affect the mature protein per se.
matography ("HPLC") is employed for purification.

[0124] The resulting protein may be used for any of its known utilities, which, in the case of albumin, include i.v. administration to patients to treat severe burns, shock and blood loss, supplementing culture media, and as an excipient in formulations of other proteins.

[0125] Although it is possible for a therapeutically useful desired protein obtained by a process of the invention to be administered alone, it is preferable to present it as a pharmaceutical formulation, together with one or more acceptable carriers or diluents. The carrier(s) or diluent(s) must be "acceptable" in the sense of being compatible with the desired protein and not deleterious to the recipients thereof. Typically, the carriers or diluents will be water or saline which will be sterile and pyrogen free.

[0126] Thus, a sixth aspect of the present invention provides a process wherein a desired protein, obtained by a process according to the fifth aspect of the invention, is formulated with a therapeutically acceptable carrier or diluent thereby to produce a therapeutic product suitable for administration to a human or an animal.

[0127] The therapeutic product may conveniently be presented in unit dosage form and may be prepared by any of the methods well known in the art of pharmacy. Preferred unit dosage products are those containing a daily dose or unit, daily sub-dose or an appropriate fraction thereof, of an active ingredient.

[0128] It should be understood that in addition to the ingredients particularly mentioned above the therapeutic product may include other agents conventional in the art having regard to the type of product in question.

[0129] The invention will now be described in more detail by reference to the following non-limiting Figures and Examples wherein:

Figure 1 shows a comparison of a natural HSA leader (having pre and pro regions) (top line) with a fused HSA/ MFα-1 leader sequence as disclosed in WO 90/01063 (second line) and a preferred modified leader sequence of the present invention (third line).

Figure 2 shows the standard genetic code.

Figure 3 shows a modified list of preferred S. cerevisiae codons.

Figure 4 shows a plasmid map of pAYE438.

Figure 5 shows a plasmid map of pAYE441.

Figure 6 shows a plasmid map of pAYE309.

Figure 7 shows a plasmid map of pAYE467.

Figure 8 shows a plasmid map of pAYE443.

Figure 9 shows a plasmid map of pAYE653.

Figure 10 shows a plasmid map of pAYE655.

Figure 11 shows a plasmid map of pAYE639.

Figure 12 shows a plasmid map of pAYE439.

Figure 13 shows a plasmid map of pAYE466.

Figure 14 shows a plasmid map of pAYE640.

Figure 15 shows plasmid maps of pAYE638 and pAYE642.

Figure 16 shows a plasmid map of pAYE643.

Figure 17 shows a plasmid map of pAYE645.

Figure 18 shows a plasmid map of pAYE646.
Figure 19 shows a plasmid map of pAYE647.

Figure 20 shows an analysis of rHSA productivity by rocket immunoelectrophoresis. Yeast were cultured in YEP, 2% (w/v) sucrose or B/MM, 2% (w/v) sucrose for 72 hr, 200rpm at 30°C. Quantitation was performed by reference to HSA standards (mg.L\(^{-1}\)).

Figure 21 shows the albumin productivity in high cell density fermentation. *Means that the human albumin level was too low to quantify.

Figure 22 summarises the characteristics of the constructs used in the examples.

**Example 1**

[0130] The *Saccharomyces cerevisiae* PRB1 promoter was isolated from yeast genomic DNA by PCR using two single stranded oligonucleotides PRBJM1 and PRBJM2:

PRBJM1  
5'-GCATGCGGCCGCCCGTAATGCGGTATCGTGAAAGCG-3'  
SEQ ID NO:35

PRBJM2  
5'GCATAAGCCTACCCACTCTTTGCTTTTAG-3'  
SEQ ID NO:36

[0131] The PCR conditions 40 cycles of 94°C for 30 seconds, 50°C for 40 seconds, 72°C for 120 seconds, followed by 72°C for 600 seconds, followed by a 4 °C hold. The 0.85kb DNA fragment was digested with both NotI and HindIII and ligated into pBST+, described in WO 97/24445, similarly digested with NotI and HindIII, to create plasmid pAYE438 (Figure 4). Plasmid pAYE438 was digested with HindIII and BamHI and ligated with the 0.48kb HindIII/BamHI ADH1 terminator DNA fragment from pAYE440 previously disclosed in WO 00/44772, so as to create plasmid pAYE441 (Figure 5). Plasmid pAYE441 was linearised at the unique HindIII site and ligated with the 1.8kb HindIII/BSu36I fragment from pAYE309 (Figure 6) previously disclosed (Sleep, D. et al. (1991) Bio/Technology 9, 183-187 and EP-A-0 431 880 and the double stranded oligonucleotide linker

5'-TTAGGCTTATA-3'  
SEQ ID NO: 37

3'-CCGAATATTCTGA-5'  
SEQ ID NO: 38

so as to create pAYE467 (Figure 7). The 3.2kb NotI, expression cassette from pAYE467 was ligated into NotI linearised pSAC35 (Sleep et al. (1991), Bio/Technology 9: 183-187), which had been previously treated with calf intestinal phosphatase (CIP) to create plasmid pAYE443 (Figure 8). SEQ ID No 22 shows a polynucleotide sequence that comprises the coding region of the HSA/MF\(\alpha\)-1 fusion leader sequence and the mature human albumin coding region to be found within the DNA sequence of both pAYE467 and pAYE443. The polynucleotide sequence encoding the HSA/MF\(\alpha\)-1 fusion leader sequence was modified by site directed mutagenesis with a single stranded oligonucleotide called CPK1 with the DNA sequence:

5'-CT AAA GAG AAA AAG AAT GGA GAC GAT GAA TAC CCA  
Ile\(^{16}\) Val\(^{18}\) Ile\(^{19}\) Phe\(^{20}\) 
CTT CAT CTT TGC-3'  
SEQ ID No 23

[0132] Site directed mutagenesis (SDM) was performed according to standard protocols (Botstein and Shortle, "Strategies and Applications of In Vitro Mutagenesis," Science, 229: 193-1210 (1985) although or other suitable techniques could also be used. The nucleotide sequence of CPK1 was designed to modify the amino acid sequence of the HSA/ MF\(\alpha\)-1 fusion leader sequence to introduce the following mutations Thr-20Phe, Phe-19Ile, Ile-18Val and Leu-16Ile, where the numbering (-20 etc) is such that the -1 residue is the C-terminal amino acid of HSA/MF\(\alpha\)-1 fusion leader sequence.

[0133] The DNA sequence of the mutagenised plasmid was confirmed by dideoxynucleotide sequencing which con-
firmed that the polynucleotide sequence had been mutagenised to the desired sequence and that no other DNA sequence alterations had been introduced. The new plasmid was named pAYE653 (Figure 9). SEQ ID No 24 shows a polynucleotide sequence that comprises the coding region of the modified HSA/MFα-1 fusion leader sequence and SEQ ID No 25 shows a polynucleotide sequence that comprises the coding region of the modified HSA/MFα-1 fusion leader sequence and the mature human albumin coding region to be found within the polynucleotide sequence of pAYE653.

Example 2

SEQ ID No 19 shows a DNA sequence that comprises: a non-coding region that includes a 5' UTR from the Saccharomyces cerevisiae PRB1 promoter, a polynucleotide region encoding the modified HSA/MFα-1 fusion leader sequence of the invention; a codon optimised coding region for mature human albumin and translation termination sites.

As a control with which to compare the effects of the sequence modifications provided to the leader sequence in SEQ ID No 19, SEQ ID No 40 shows a DNA sequence that is essentially the same as SEQ ID No 19, except that, instead of the 15 polynucleotide region representing the second aspect of the invention, the DNA sequence of SEQ ID No 40 comprises an 15 polynucleotide region encoding the 5 amino acids of an unmodified HSA/MFα-1 fusion leader sequence, namely SFISL.

Both DNA sequences were synthesised by Genosys, Inc (Cambridge, UK) from overlapping single-stranded oligonucleotides.

SEQ ID No 40 was synthesised as a 1.865kb SacI - HindIII DNA fragment cloned into the SacI - HindIII sites of plasmid pBSSK- (Stratagene Europe, P.O. Box 12085, Amsterdam, The Netherlands), as plasmid pAYE639 (Figure 11).

The Saccharomyces cerevisiae PRB1 promoter was isolated from yeast genomic DNA by PCR using two single stranded oligonucleotides PRBJM1 and PRBJM3:

The PCR conditions 40 cycles of 94°C for 30 seconds, 50°C for 40 seconds, 72°C for 120 seconds, followed by 72°C for 600 second, followed by a 4 °C hold. The 0.81kb DNA fragment was digested with both NotI and HindIII and ligated into pBST+, described in WO 97/24445, similarly digested with NotI and HindIII, to create plasmid pAYE439 (Figure 12). Plasmid pAYE439 was digested with HindIII and BamHI and ligated with the 0.48kb HindIII/BamHI ADH1 terminator DNA fragment from pAYE440 previously disclosed in WO 00/44772, so as to create plasmid pAYE466 (Figure 13).

A 1.865kb HindIII DNA fragment of SEQ ID No 40 was cloned into the unique HindIII site of plasmid pAYE466 to create plasmid pAYE640, which was shown to contain the 1.865kb HindIII DNA fragment of SEQ In No 40 between the PRB1 promoter and the ADH1 terminator in the correct orientation for expression from the PRB1 promoter(14).

Plasmid pAYE640 was digested to completion with NotI/PvuI and the NotI 3.2kb, PRB1 promoter/HindIII DNA fragment of SEQ ID No 40 gene/ADH1 terminator expression cassette was purified. A NotI/PvuI double digest of pAYE640 was preferable to a single NotI digestion because the expression cassette (3.2kb) and pBST+ plasmid backbone (3.15kb) were similar in size. The 3.2kb NotI, expression cassette from pAYE640 was ligated into NotI linearised pSAC35 (Sleep et al. (1991), Bio/technology 9: 183-187), which had been previously treated with calf intestinal phosphatase (CIP) to create plasmid pAYE638 (Figure 15). Plasmid pAYE638 was shown to contain the NotI HSA expression cassette inserted into the NotI site of pSAC35 and orientated so that the expression of the HSA gene was away from the LEU2 auxotrophic marker and toward the 2μm origin of replication. Plasmid pAYE642 contained the same HSA expression cassette but arranged in the opposite orientation (Figure 15).

SEQ ID No 19 was synthesised as a 1.865kb SacI - HindIII DNA fragment cloned into pBSSK- (Stratagene Europe, P.O. Box 12085, Amsterdam, The Netherlands), as plasmid pAYE643 (Figure 16). The DNA sequence which encodes for an HSA/MFα-1 fusion leader sequence-albumin fusion within pAYE643 is given in SEQ ID No 27. The
Example 3

[0144] Three different yeast strains, A, B and C, were transformed to leucine prototrophy with plasmids pAYE443, pAYE638, pAYE646 and pAYE655. The transformants were patched out onto Buffered Minimal Medium (BMM, described by Kerry-Williams, S.M. et al. (1998) Yeast 14, 161-169) containing 2% (w/v) glucose (BMMD) and incubated at 30°C until grown sufficiently for further analysis. The human albumin productivity of the transformants was analysed from 10mL YEP (1% (w/v) yeast extract; 2% (w/v) bacto peptone) containing 2% (w/v) glucose (YEPR) and BMMD shake flask culture (30°C, 200rpm, 72hr) by rocket immunoelectrophoresis of cell free culture supernatant (Figure 20).

[0145] The results showed that the human albumin productivity of all three strains transformed with pAYE638 was approximately 4-5 fold lower than that observed in the same strain transformed with pAYE443 (which both contained the HSA/MFα-1 fusion leader sequence, but encoded by different polynucleotide sequences) in both rich and defined media. Unexpectedly, the human albumin productivity of all three strains transformed with pAYE646 or pAYE655 was significantly higher than that observed with pAYE638 and similar or slightly greater than that observed for the same strains transformed with pAYE443.

Example 4

[0146] Yeast strain C [pAYE443], strain C [pAYE655], strain C [pAYE638] and strain C [pAYE646], and strain B [page443] and strain B [pAYE646] were cultivated in high cell density fermentation in both fed-batch and fill & draw procedures. The fed-batch procedure used a medium and control parameters as described in WO 96/37515. The fill & draw procedure used the fed-batch procedure as described above, but additionally included the steps that: upon completion of the feed phase of the fed-batch culture procedure, 90% of the culture volume was removed from the fermenter vessel; and batch medium was added to the remaining 10% volume of the culture (maintaining pH control) prior to the initiation of feed addition, using the medium and control parameters described in WO 96/37515. The human albumin productivity (YP/S) and human albumin concentration (g/L) were assessed by scanning densitometry of SDS-PAGE of cell free whole culture. The biomass yield (YX/S) was also calculated from gravimetric determinations. The results (Fig. 21) indicated that, as seen previously in Example 3, the human albumin productivity (YP/S) and human albumin concentration (g/L) of yeast strains containing the human albumin expression plasmid pAYE638 (native polypeptide sequence but yeast-biased codons) had significantly lower productivity than the same strains containing the human albumin expression plasmid pAYE443 (native polypeptide sequence and natural codon bias for leader and mature albumin) even though the amino acid sequences of both the HSA/MFα-1 fusion leader sequence and the mature human albumin were identical.

[0147] When the strain C fermentations were run in fed-batch mode a 16% and 12% increase in human albumin productivity (YP/S) relative to that of Strain C [pAYE443] was observed when Strain C [pAYE655] and Strain C [pAYE646] (human albumin expression plasmids incorporating a modified leader sequence in accordance with the present invention) were each cultured for a comparable length of time, respectively. When the strain B fermentations were run in fed-batch mode a 24% increase in human albumin productivity (YP/S) relative to that of Strain B [pAYE443] was observed when Strain B [pAYE646] (the human albumin expression plasmid incorporating a modified leader sequence in accordance with the present invention) was cultured for a comparable length of time.

[0148] When the strain C fermentations were run in fill and draw mode a 13% and 6% increase in human albumin productivity (YP/S) relative to that of Strain C [pAYE443] was observed when Strain C [pAYE655] and Strain C [pAYE646] (the human albumin expression plasmids incorporating modified leader sequence in accordance with the present invention) were each cultured for a comparable length of time, respectively. This increased to 442% and 408% relative to that of Strain C [pAYE638] when Strain C [pAYE655] and Strain C [pAYE646] (the human albumin expression plasmids incorporating a modified leader sequence in accordance with the present invention) were each cultured for a comparable length of time, respectively.

Summary

[0149] Plasmids pAYE443 and pAYE638 both encode human albumin having a leader sequence derived from HSA/MFα-1 fusion leader sequence, but the former uses the natural codon bias of the native polynucleotide sequences, while
the latter uses a polynucleotide sequence which is fully codon optimised for yeast expression. Expression of human albumin obtained from pAYE638 is 4-5 fold lower than that obtained using pAYE443. A polynucleotide sequence encoding a modified leader sequence in accordance with the present invention has been substituted into the polynucleotide sequence encoding the HSA/MFα-1 fusion leader sequence of both pAYE443 and pAYE638 to create the human albumin expression plasmids pAYE665 and pAYE646, respectively. The introduction of the polypeptide sequence according to the present invention led to a significant improvement in production of the desired polypeptide.

**SEQ ID No. 1**

(Phe/Trp/Tyr)-(Ile/Leu/Val/Ala/Met)-(Leu/Val/Ala/Met)-(Ser/Thr)-(Ile/Val/Ala/Met)-

**SEQ ID No. 2**

Phe-(Ile/Leu/Val/Ala/Met)-(Leu/Val/Ala/Met)-(Ser/Thr)-(Ile/Val/Ala/Met)-

**SEQ ID No. 3**

(Phe/Trp/Tyr)-Ile-(Leu/Val/Ala/Met)-(Ser/Thr)-(Ile/Val/Ala/Met)-

**SEQ ID No. 4**

(Phe/Trp/Tyr)-(Ile/Leu/Val/Ala/Met)-(Ser/Thr)-(Ile/Val/Ala/Met)-

**SEQ ID No. 5**

(Phe/Trp/Tyr)-(Ile/Leu/Val/Ala/Met)-(Leu/Val/Ala/Met)-Ser-(Ile/Val/Ala/Met)-

**SEQ ID No. 6**

(Phe/Trp/Tyr)-(Ile/Leu/Val/Ala/Met)-(Leu/Val/Ala/Met)-(Ser/Thr)-Ile-

**SEQ ID No. 7**

-Phe-Ile-Val-Ser-Ile-

**SEQ ID No. 8**

-Met-Lys-Trp-Val-

**SEQ ID No. 9**

-Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-

**SEQ ID No. 10**

-(Ile/Leu/Val/Ala/Met)-(Phe/Trp/Tyr)-(Ile/Leu/Val/Ala/Met)-(Phe/Trp/Tyr)-(Ser/Thr/Gly/Tyr/Ala)-(Ser/Thr/Gly/Tyr/Ala)-(Ile/Leu/Val/Ala/Met)-(Phe/Trp/Tyr)-(Ser/Thr/Gly/Tyr/Ala)-
SEQ ID No. 12
-Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-Arg-Ser-Leu-Asp-Lys-Arg-

SEQ ID No. 13
-Met-Lys-Trp-Val-X_1-X_2-X_3-X_4-X_5-(Ile/Leu/Val/Ala/Met)-(Phe/Trp/Tyr)-(Ile/Leu/Val/Ala/Met)-(Phe/Trp/Tyr)-(Ser/Thr/Gly/Tyr/Ala)-(Ser/Thr/Gly/Tyr/Ala)-(Ile/Leu/Val/Ala/Met)-(Phe/Trp/Tyr)-(Ser/Thr/Gly/Tyr/Ala)-

SEQ ID No. 14
-Met-Lys-Trp-Val-X_1-X_2-X_3-X_4-X_5-Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-Arg-Ser-Leu-Asp-Lys-Arg-

SEQ ID No. 15
-(TTY/TGG/TAY)-(ATH/TTR or CTN/GTN/GCN/ATG)-(TTR or CTN/GTN/GCN/ATG)-(AGY or TCN/ACN)-(ATH or CTN/GTN/GCN/ATG)-

SEQ ID No. 16
-TTY-ATH-GTN-(TCN or AGY)-ATH-

SEQ ID No. 17
-(TTC/TGG/TAC)-(ATY/TTG/GTY/GCT/ATG)-(TTG/GTY/GCT/ATG)-(TCY/ACY)-(ATY/GTY/GCT/ATG)-

SEQ ID No. 18
-TTC-ATY-GTY-TCY-ATY-
SEQ ID NO 19:

AAGCTTAACCTAATTCTAACAAGCAAGGTGAGTGTTCTTTCA
TCGTTCTCCATTTTTGTCTTCTTCTCTCTGCTTACTCTAGATCTTTG
GATAAAGAGAGACGCTCAAAGTCGGAGTCGCTCAGAGATTCA
GGACTTGGGTAGAAATCTTCAAGGCTTTTGCTTGTATCGCTTT
CGCTCAATACCTGCAACAATGTCCACTTGGAAGATCGACTCAAGTT
GGTCAAGCAGGTTACCGAATCTCGTAAAGACTTGTGGTGCTGAG
AATCTGCTGAAAAACTGTGACAAGTCCTTGGACACACCTTTGTTGCTA
ATAAGTTTGTGACTTGTGCTACCTTGGAGAAGACTACCGTGAA
ATGGCTGACTTGGTGCTAAGCAAGAACCAGAAAGAAACGAAATG
TTTCTTCAACCAAGCGACACAACCCAAACTTGCCAAGATTGG
TTAGACCAGAAGTGGACGTATGTGTACTGCTTCTGCACGACAACG
AAGAAACCTTCTTGAAAGAAGTACTTGTACGAAATTGCTAGAAGA
CACCCATACCTGCAAGATTGTGCTCTCCTCTGCATAGAAGA
TACAAGGGCTGTCTTACCCGAATGTTGTCAAGCTGCTGATAAGGCT
GCTTGTTTGTGCCAAAGTTGGATGAAATTGAGAGCAAGGTAA
GGCTTCTTCCGCTAAGCAAGAAGTGTTGCTTCTTCTGCAAAA
GGTGGGTTGAAAAGAGCTTCTGCAAGGCTTTGGGCTGTGCTGACTATTGC
TCAAAGATTCCAAAGGGCTGAAATCGCTGAAGTTTCTAAGTGTGTTG
TACTGACTTTGACTAAGGTTCACACTGAATGTGCTACGCTACGTTG
GTGGAAATGTGGTGAGTAGCAAGGCTGACTTGGCTAAGTAGATCTC
GTGAAAAACAAAGACTCTATCCTCTCTTCTCCAAGTTGAAAGATGTGTTG
AAAAGCCATTGTTGGAAAAAGTCTCACGTATTGCTGAAGTTGAA
AACGATGAAATGCCAGCTGACTTGCCATCTTTGGGCTGACTTC
GTGTAATCTAAGGACGTTTGTGAAGAACTACGCTGAAGCTAAAGGA
CGTCTCTTGGGTATGTCTCTGTACGAATACGCTAGAAGACACCC
AGACTACCTCCGTGTTGTTGGTTGGAGATTGGCTAAGACCTACGA
AACTACCTTTGAAAAGTTGTGTGTGCTGCTGCTGCAACACAGAAT
GTACGCTAAAGGGTTTTGGATGAAATCCAGCATTGGTGCAAGAAC
CACAAAACCTGGATCAAGCAAACACTGTAATTTGTTCCAACAAAT
GGTGAATACAAAGTTCCAAAACGCTTTGGTTGTAAGATACACTAA
GAAGGTCCCACAAGTCTCCAACCCCAACTTTGGTTGAAAGTCTCTAG
AAACTTGGGTAAGGTCGTTCTAAGTTGGTGAAGCACCAGAG
CTAAGAGAATGCCATGTGCTGAAGATATTACTCTGTCCGTGTTTGA
ACCAATTGTGTGTTTTGCACGAAAAAGCCCAAGTCTCTGATAGAG
TCACCAAGTTGTTGACTGAATTCTTTGGTTAAGAGAAAGCCATGT
TCTCTGTTTGGAAAGTGCAGCAAAAACCTACGTTCCAAGGAATTCA
ACGCTGAAAACCTTTCACCTTCCACGCTGATATCTGTACCTTTGGCC
AAAAGGAAAGACAAATTTAAGAAAGCACAAGCTGCTTTGGTTGAATTG
GTCAAAGCAACAGCACAAGGCTACTAAAGGAACAAATTGAAAGGCTGT
CATGGATGTTCGCTGTCTTTCCGTGAAAGTGTGTGTAAGGCTGA
TGATAAGGAAACTTGTTTTGGTCTGAAAGGTAAGAAGTTGGTGTCG
CTGCTTTCCAAGCTCTTTGGTTTGTAAATAAGCTT
SEQ ID NO 20:

AGATCTTTGGATAAGAGAGACGCTCACAAAGTCCGAAGTCGCTCA
CAGATTCAAGGACTTGGGTGAAGAAACTTCAAGGCTTTGGTCT
TGATCGCTTTGCTCAATACCTTGCAACAATGTCCATTCGAAGATC
ACGTCAAAGTTGTCAACGAAGTGTTACCGAAATTGCTAAGACTTGT
GTTGCTGACGAATCTGCTGAAAACGTGACAAAGTCCTTGCACACC
TTGTTTGGTGAAGTTGTGACTGTGTGCTACCTTGAGAGAACCC
TACGGTGAAATGGCTGACTGGTTGTGGCTAAGCAAGAAACCAGAAAG
AAACGAATGTCTTCTTGCAACACAAAGCAGACAAACCCAAACTTIGC
CAAGATTGGTAAAGCAGAAAGTTGACGTCATGTGTACTGCTTTCC
ACGACAACGAAGAAACCTCCTTGAGAGAAGTACTTTGTACGAAAATT
GCTAGAAAGACACCCATATCTCTAGCTCCAGAATTGTTGGTCTTC
GCTAAAGAGATAAAGGGCTGTCTTCCACCGAATGTGGTCAAGCTGCT
GATAAGGCTGTTGTGGTGGCCAAAGTTGGAGAATTTGGAGAGA
CGAAGGTGTAAGGTCTTCTTCGCCAAGAATGAAATGTTGTGCTT
CCTTGCAAAAAGTTGGTAAGAGCCTTTCAAGGCTGCTGCTGCT
GCTAGATTTGCTCAAAAGATTTCCCAAAGGCTGAAATTCCGTAAGTT
TCTAAGTTGTTTACTGACTTGACTAAGGTTCACACTGAATGTTGT
CAACGGTGACCTGGTTGAAATGTGTGCTGATGACAGAGCTGACTTGGCT
AAAGTACATCTGTGAAAACCAAGACTCTATCTCTTCCAAGTTGAAAG
GAATGTGGTGAAGGCAATTTGTTGGAAAGTACCTGACTGATTTGCT
GAAGTTGAAAAACGATGAAATGGCAGCTACCTGACCTGACTTGGGC
TGCTGACTTCTGTTGAATCTAAGGACGTGGTGTAAGAAACTACGCTGA
AGCTAAGGACGTCTTCTTGGGTATGTTCTTGTACGAATACGCTAG
AAGACACCCAGACTACTCCGTGTCTTTGTTGAGATTTGGCTAA
GACCTACGAAAACACTCTTTGGAAAAGTTGGTGGCTGCTGCTGACT
ACACGAAATGTGTACGCTAAGGTTTTCGTAAGTTCAAGACCCTTGGT
CGAAGAACCACAAAACCTGTGATCAAGCAAACCTGTGAAATTTGTCG
AAACAAATGGGTGAAATACAAGTCTCCAAAACGCTTGGTTGTTAGAT
ACACTAAGAAAGGTTCCACAAAGTCTCCACCCCAAACCTTTGGGA
GTTCTCAGAAAACCTGTTGGTAAAGGTCTGGTTCTAAGTGTGGTGAAGCACY
CCAGAAGCTAAGGAATGCCCATGTGTGCAAGATTAATCCTGTCCGT
CGTTTTGAAACATTGTGTTTTTGCAGAAAAGACCCCAACTTCTC
TGATAAGAGTCACAAAGTTGGTATCTGAATCTTTGTTGTTAACAGAAG
ACCATTGTCTCTGTCTTTGGGAAGTGCAGAAAAACTTACGTTCCAA
GGAACCCACGCTGAAACCTTTACCTTCCACCTGCTGATATCTGTAC
CTTGTCCGAAAAGGAAAGACAAATTTAAAGAAGCAAAACTGCTTTGG
TTGAATTGGTCGAACCAAGCCAAAGGCTACTAAAGGAACAAATTG
AAGGCTGTCTGGATGATTTCGCTGCTTTCTGTTGAAAGTGTTGT
AAGGCTGTGTAGAAAGGAGACTTTGTTTCTGCTGAAAGGTTAGAA
GTTGGTCTGCTGCTTTCCAAAGCTGCTTGGGTTTG
SEQ ID NO 21:

ATGAAGTGGGTTTTCATCGTCTCCCATTTTGTCTTTGTCTCTCTCTG
CTTAATCTAGATCTTTGGATAAGAAGAGACGTCACAAGTGCGAA
GTCGCTCAGATTTCAAGGACCTGGGTGAAAGAAAACTTCAAGGC
TTTGGTCTTGATCGCTTTGGCTCAATCTTTGCAACAAATGTCATTC
GAAGATCAGTCGAAGTTGGTCAACGAAGTTACCGAAATTCCGCTAA
GACTTGTGTGGTCTGACGAATCTGCTGAAGAAGACTGTCAGAAGTCTT
GCACACCTTTGGTGTTGATAAAGTGTTGACTGTTGCTACTCTTTGAG
AGAAACCTACGGTGAAATGCGACTGTTGCTAAAGCAAGAAC
CAGAAAGAAACGAATGTTTCTTGCAACACAAAGGACGACAACCCA
AACTTGCAAGATTGTTGAAGACCAAGTTGACGTCATGTCGACT
GCTTCCACGACAACGAAGAAACCTTTCTTGAAGAAAGTACTTCTTGA
CGAAATTGTGTTAAGACACCCATACTTCTCTACGCTCCAGAATTGT
GTCCTCGCTAAGAGATAAAGGCTGCTTTCAACCAAGTTGTCACA
AGCTGCTGATAAAGGCTGCTTTGTTTGGAGCAAGATTGGAGAATT
GAGAGACGAAGGTAAGGCTTTCTTCGGCTGAAGCAAAAGATTGAAGT
GTGCCTCTTGGCAAAAAGTTCGGTGAAAGAAGCTTTCAAGGCTGGG
CTGCTGCTAGATTGTCTCAAAGGATTTCCCCAAAGGCTGAAATTCCGT
AAATTCTAAGGTTGGTACTGACATTGACTAAGGTACACTAAGAT
GTGTGACAGGTGACCTTTGGGAATGGTGTGATGACAGAGCTGACT
TGCTAAAGTACATCTTGGAAAAACAAGACTCTATCTCTTCCAAAGT
TGAAAGAATGTTGGAAAAGGCAATTGTTGGAAGAGTCTCAGACTG
ATTGCTGAAAGTGAAAACGATGAATGCAATGCCAGCTGACTCTGCACT
TTTGGCTGCTGACCTTGGAAATCTAAGGACGTTTGTAAAGAACTA
CGCTGAAGCTAAGGACGTCTTCTTGGGTATGTCTTTGTACGAATA
CGCTAGAAGACACCCAGACTACCTCGTCTTGTGGAGATT
GGCTAAGACCTACGAAAACCTACCTTGGAAAAAGTGTGTTGCTGCTG
CTGACCCACACGATGTTACGTAAGGT GTTTTCGATGAATTCAAAGC
CATTTGGTGCAAGAACACACACAACTTTGATCAAGCAAAACGTGTAA
TTGGTGCAACAACTTGGTGAATAACAAGTTCACAAAACGCCTTTGTG
GTAGATACACTAAAGAGGTCCACCACAAGTCTCCACCCCAACATTG
GTGAACTGCTCTAGAAACTTGGGTAGAGTGCTGGTTCTAAGTGTTGT
AAGCCACCAAGCTAAGAGAACATGCCATGTGCAGAGATTACCTT
GTCCGTCGTTTTGGAAACCAAATGTGTTTTTGCAAGAAAAGACCCC
AGTCTCTGATAGAGTCAACCAAGTGTGTACTGAAACTTTGGTTAA
CAGAAGACCATGTTCCTCTGCTTTGGAAAGTGCAGAAAATTTAGT
TCCAAAGGAATTCACCGCTGAAAACCTTTACCTTCCACCGCTGATAT
CTGTACCTTGTCCGAAAAGGAAGCAAAATTAAGAAGCAAACCTG
CTTTGGTGGAATTGTGCAAGCACAAGCCAAAAAGCTACTAAGGAA
CAATTGAAGGCTGTCACTGGATGATTTCGCTGCTTTTCGTTGAAAAAG
TGTTGTAAAGGCTGATGATAAGGAAAACCTTTGGTTGCTGAGAAGG
TAAGAAGTTGGTGCTGCTCTCCCAAAGCTGTTTTGGTTTG
SEQ ID NO 22:

ATGAAGTGGGTAAGCTTTATTTCTCTTTTTCTCTTTTAGCTCGG
CTTATTCAGGAGCTTTGGATAAAAAAGAGATGCACAACAAGAGTGAG
GTTGCTCATCGGTAAAAGATTTGGGAGAAGAAAAATTCTAAAGGC
CTTGGTGTGTTGATTGCTTTGCTCAGTATCTCTCACGACGTGTCATTT
GAAGATCATGTAAAAATTAGTAATGAATGAAGTAACGTAACTGAAATTTGCCAA
AACATGTTGCTGTAGTGAATGACAGTGAATGAAAATTGTGACAAATCAC
TTCAATCCCTTGTGGAGACAAATTATGCACAGTTGCAACTCTTCTCC
GTGAAACCTATGGTGAATGGCTGACTGCTGTGCAAAACCAAGAA
CCTGAGAGAAATGAATGCTTTCTTGCAACCACAAGATGACAACCC
AAACCTCCCCCGATTGGTGAGACCAGAGGGTGATGTGTGTGCA
CTGCTTTTCATGACAATGAAGAGAAGACATTCTTTGGAAATAATACCTAT
ATGAAATTGGCCAGAAGACATCTCTACTTTTAGGACGAACTCC
TTTTCTTTGCTAAAAAGGTATAAAGCTGCTTTTCAGAATGTTGCC
AAAGCTGCTGATAAAGCTGCTGCTGCTGCTGCTGCAAGCTGATGAA
CTTCGGGATGAAAGGGAAGGCTCTGCTGCTGCAAAACAGAGACTCWA
GTGGTGCCAGTCTCCAAAAATTGGAAGAAGAGCTTCAAGCAT
GGGCAGTAGCTCGCCTGAGCCAGAGATTCCCAGAAGCTGAGTTT
GCAGAAGTTCCTCAAATCTATCTGAGAAATCAAGATTCGATCTCCA
GTAAACTGAAAGGAATGCTGTGAAAAACCTCTGTTGAAATAATCC
CACTGCTATTGCGCAAGTGGGAAATAGTGAGATGCTGTGCTGACCTT
GCCCTATTAGCTGCTGATTCTTGTGAAAAGTAAAGGATGTGTTGCAA
AAACTATGCTGAGGGCAAAGGATGTCTTCTGCGGCTAGTCTTTTGTAT
TGAATATGCAAGAAGGCATCTGTATACTCTCTGTGCTGCTGCTGCT
GAGACTTGCCCAAGACATATGAAAACCACCTCTAGAAGAGTGCTGTG
CGCTGCAGATTCCTCATGAAATGCTATGCCAAAGTGTTTCGATGAAT
TTAAAACCTCTTTGTGGAAAGAGCCTGAATAATCACAACAAAATT
GTGAGCTTTTTTGGAGCGACTGTTGGAGATCACAATTCCTAGAAGCTGCG
CTATTAGTTGGTTACCAAGAAAGATGGGCAAAGTCCGACTCCA
ACTCTTTGAGAGGCTCTCAAAGAAACTAGGAAAAAGTGGCGAGCAA
ATGGTTGAAACATCTGAAAGCAAAGAAGATGCTCTGTGCAAGAG
ACTATCTATCATCGTGTCCATGAAACAGTTATGTTGTTGTCATGAGA
AAACGCGATAGTGACAGAGTGCAAAATGCTGCAGACAGATCC
TTGCTGAAACAGGCGACCATGCTTTCAGCTCTGGAAGTGCTGATGA
AACATAGTTCCAAAGAGTTTATGCTGAAACATTCACCTCCCA
TGCAAGATAATGCAACACTTTCTGAGAAGGAGAGACAAATCAAGA
AAACAACTGCACTTTGGTGAGCTCGTGAAACACAAAGCCAAGGCA
ACAAAAAGAGCAACTGAAAGCTGTTATGGATGATTTCCGACGCTTT
TGTAAGAAGATGCTGCAAGGCTGACGATAAAGGAGACCTGCTTTG
CCAGGAGGGTAAAAAATTTGTTGTGCTGCAAGTCAAGCTGCCTTA
GGCTTA

SEQ ID NO 23
CTAAAGAGAAAAAGAATGGGAGACGATGAATACCCACTTCATCTT
TGC

SEQ ID NO 24
ATGAAGTGGGTATTCACTCGTCCTCCATTCTTTTCTTTTAGCTCGG
CTTATCCAGGAGCTTGGATAAAAGA
SEQ ID NO 25

ATGAAGTGCGGTATTCACGCTCCATATTCTTTTTCTTTTAGCTCGG
CTTATCCAGAGCTGGGATTTAAAGAGAAGATGCACACAAAGAGTGAG
GTTGCTCTACGGTTTTAAAGATTTGAGGAAAGAAATTTCAAAAGC
CTTGGGGTTGATTGCTTTGCTCAGTATCTTCAGCAGTGTCATTT
GAAGATCATGTAAAAATTAGTGAATGAAGTAACTGAATTGGCAAA
AACATGTGGTTGCTGTGAGTGCAGCTGAAAAATTGTGACAAATCAC
TTCACTACCCCTTTTGGAGACAAATTATGCAACAGTTGCAACTCTTC
GTGAAACCTATGTGAAATGGCTGACTGCTGTGCAAAAACAAGAA
CCTGAGAGAAATGAATGCTTCTTTGCAACACAAAAAGATGACAACCC
AAACCTCCCCCGATTGTTGAGACCGAGGTGGATGTGATGTGCA
CTGCTTTTCATGACAAATGAAGAGACATTTTTTGAAAAAAATACCTTAT
ATGAAATTGCGAGAAGACATCTTCTTTATGCCCCCGAACTCC
TTTTCTTGTAAAGGTATAGGCTTCTTTACAGAATGTGACC
AAGCTGCTGATAAAGCTGCCTGCCTGTGCAAGCGTCGATGAA
CTTCCGGAGAAGGGAAGGCTCGTCGTTGCCAAAACAGAGACTCAA
GTGTGCCAGTCTCCAAAATTTGGAGAAAGAGCTTTTCAAGACAT
GGGCAGTAGCTCCTGCTGACAGGAGATTTCCAAAAGCTGAGTTT
GCAGAAGTTTCAAGTTAGTGACAGATCCTTACAAAGTCCACAC
GGAATGCTGCGATAGCTGCTGTTGAATGTGCTGATGACAGGG
CGGACCTTGGCCAAAGTATATATCTGAAAATCAAGATTCGATCTCCA
GTAAACTGAAAGGAATGCTGTAAGAAACCTCTGTGTTGGAAATAATCC
CACTGCAATTGCCAAATGGAAAATGATGAGATGCGCTGTGAATT
GCCTTCATTAGCTGCTGATTTTGTGAAAAGTAAGGATGTTTGCAA
AAACTATGCTGAGGCAAAGGATGCTTTCTGCGATGTTTTTGTA
TGAATATGCAAGAAGGCTACCTGATTAATCTGTGCTGCTGCTGCT
GAGACTTGGCGAAAGCTATGAAAACCTCTAGAAAGATGCTGATG
CCGCTGCAGATCCTCATGAATGCTATGCGAAGAAGTGTTCGATGAAT
TTAAACCTCTTTGTGGAAGAGCGCTCAAGATTTAATCAAAAACAAAATT
GTGAGCTTCTTGGAGCAGCTTGGAGAGTACAATAATCCAGAATGCG
CTATATTGGTCTCTGACCACAAAGATACCCCAAGTGTCAAGCCCT
ACTCTCTGTAGAGTCTCAAGAAACCTAGGAAAGTGGGCGAGCAA
ATGTGTAACATCCGGAAGAAAAAGAATGCCCTGTGCAAGAAG
ACTATCTATCCGTGGCTCGAACAGTTATGTGTGTTGCTGATAAG
AAACGCACTAGAAGTGACAGAGTCACAAATGCTGCACAGAATCC
TTGGTGCAACAGCGACCATGCTTTTCAGCTCTGGAGAATGCTGTA
AAACATACGTTCCCAAAGAGTTAATGCTGAAACATTTACCTCTCCA
TGCAGATATATGCACACTTTCTGAGAAAGGAGAGACAAATCAAAGA
AAACAACTGCACTTTGTTGAGCTGCTGAAAACACAAGGCCAAGAGCA
ACAAAGAGCAACTGAAAGCTGTTATGATGATTCTGAGCTTT
TGTAGAGAAAGTGCTGCAAGGCTGACGAATAAGGAGACCTCTTGCTTTG
CCGAGGAGGTTAAAAACTTTGTTGCTGCAAGTCAAGCTGCTTTA
GGCTTA
ATGAAGTGGGTTCTTTCTATTTCCTTTGTTTCTTTCTCTCTTCGAA
CTTACTCTAGATCTTTGGAATAAGAGAGAGCTCAACAAGTCCGAA
GTGCCTCACAGATTCAAGGACTTGGGGAAGAAACTTCAAGGC
TTTGGTCTTGTCTCGCTTTTCGCTCAATCTTTGCACAATGTCCATTC
GAAGATCAGTCAGTGGTTGCAACGGATTACCAAATCGCTAA
GACTTGTCTTGCTACGAATCTCTCAGAAACTGTGCAAGATCCTT
GCCAACCTTGTGCGTAAGTGTTGTCTACGTGTGCTAAGCAAGAAC
CAGAAAAGAGAGATGGGTGCTAGTGCTGTGTGCTAAGCAAGAAC
AACTTGCCAAGATTGGTTAGACCAGAAAGTGGACGTCACTGGTACT
GCTTTCCACGCAAAGGAGAAACCTTCTTGGAAGAAGTACTTTGTA
CAGAAATTGTGCTAGAAGACACCACATACTCTACCGTCCAGAATTGT
GTCTCGCATAAGGATACAAGGCTGCTTTCAAGGAATGTGTCAG
AGTCTGCTGATAAGGCTGGTCTTTGTGCAAAAGTTGAGAATT
GAGAGACGAAAGGTAAAGGCTTCTCTCGCTAAGCAAGAGATTGAAGT
GTGCTCTCGTCCAAAGTCTGGAAGAGCTTCTCAAGGGCTTGGG
CTGTGCTAGATTTGTCTCAAGAATTCACAAGGCTGAATTTCGCTG
AAGTTCTAAGGTTTGTTACTGACTTGACTAAGGGTTCACACTGAAT
GTGTCACCGGTGACTTGTGGAATGTGCTGATGACAGAGCTGACT
TGGCTAAGTACATCTGTGAAACCAAGACTCTATCTCTCTCCAAAGT
TGAAGGAAATGTTGTGGAAGACCGATTGTTGGAAAGCTCAGCCTGT
ATTGCTGAAAGTTGAAACCGATGAAATGCCAGCTGACCTGCCATC
TTTGGCTGCTGACTTCTGTTAATCTAAGGACGTTGTTGTAAGAACTA
CGCTGAAAGCTAAGGACGCTTCTTGGGTATGTCTTGTAGAATA
CGCTAGAAAGACAACCAGACTACTCTCGTTGCTTTGTGTTGAGAATT
GGCTAAGACCTACGAAACTACCTTGAGAAAGTTGTGCTGCTG
CTGACCCACAGAAAGTTACGTAAGGTGTTTTCGATGAAATCCAAGC
CATTGGTCGAAAGAAACCAAAACTTGATCAAAGCAAAAACCTGTGAA
TTGGCTGAAACAAATGGGGTTGGAATACAAAGTTCCAAACAGCTTTGTG
GTAGATACAATAGAAGGGCCTCCACAAAGTCTCCACCCCAACTTTG
SEQ ID NO 27

ATGAAGTGGGTTTTTCATCGTCTCCATTTTGTTCTTGCTCCTCTCGCTTTTGGGATAG

SEQ ID NO 28

N-Met-Lys-Trp-Val-Phe-Ile-Val-Ser-Ile-Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-C

SEQ ID No 29

N-(Phe/Trp/Tyr)-(Ile/Leu/Val/Ala/Met)-(Leu/Val/Ala/Met)-Thr-(Ile/Val/Ala/Met)-C

SEQ ID No 30

N-Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-Ser-Arg-Gly-Val-Phe-Arg-Arg-C
SEQ ID No 31

N-Met-Lys-Trp-Val-X_1-X_2-X_3-X_4-X_5-Leu-Phe-Leu-Phe-Ser-Ser-Ala-Tyr-
Ser-Arg-Gly-Val-Phe-Arg-Arg-C

SEQ ID No 32

N-Met-Lys-Trp-Val-Phe-Ile-Val-Ser-Ile-Leu-Phe-Leu-Phe-Ser-Ser-Ala-
Tyr-Ser-Arg-Ser-Leu-Asp-Lys-Arg-C

SEQ ID No. 33
-Met-(Lys/Arg/His)-(Phe/Trp/Tyr)-(Ile/Leu/Val/Ala/Met)-

SEQ ID No. 34
-TTCATCGTCTCCATT-

SEQ ID No. 35
5'-GCATGCGGCCGCCCGTAATGCGGTATCGTGAAAGCG-3'

SEQ ID No. 36
5'-GCATAAGCTTACCCACTTCATCTTTGCTTGTTTAG-3'

SEQ ID No. 37
5'-TTAGGCTTATA-3'

SEQ ID No.38
5'-AGCTTATAAGCC-3'

SEQ ID No.39
5'-GTT-AGAATTAGGTAAAGCTTGTATTTTTATGCGCATGAA-3'
SEQ ID No. 40

AAGCTTAACCTAATTTCTAAAAGCAAGATGAAGTTGGTTTTCT
CATTCTTCTTGTTTCTTCTCTCTCTGCTTACTCTAGATCTTTGG
ATAAGAGAGAGCTGACAAGATGGGCTGTCGACTACAGATTTCAAG
GACTTGGTGAAAGAAAACATTTAAGGCTTTGCTTTGACTGCTT
GCTCAATACCTTGCAACAAATTGCTCATTGGAAGATCAGCTCAAGTTG
GTCAACGAAAGTTACCGAATTCGCTAAGACTGTGTTGCTGACGA
ATCTGCTGAAACTGTGACAAGCTTCTTGACAACCTTGTTGCTGTA
TAAGTTGTTGACTTTGCTACCTTGAGAGAAAACCTACGTTGAAAT
GGCTGACTGTTGCTAAGCAAGAACCAAGGAACCAAGAAGAGGA
TCTTGCAACACCAAGGAGCAACAAACCTCCCTTGGAGAGATTTGGT
AGACCGAGATTGACGTACTGTGACTGCTTTCCACGACAACCA
AGAAACCTCTTTAGAACAGTACTTTGACTGAAATTTGCTAGAAGAC
ACCCATACCTCTACGCTCAGAAATTGTTGTTTCTTCGCTAAGAGAT
ACAAAGGCTGCTTTAGCCAGAATTTGCTCAAGCTGCTGATAAGGCTG
CTTGTGTTGCTCCAAAGGATTGGAGATGAGAGAACAGAGGTAAG
GCTTCTTCGCTAAGAAAGTACCTTGAGAGAGTCTCTTTCCAAAGAG
TTGCGTCAGAAGCTTTTCAAGGGCTTTGCTGCTGCTAGATTGGTCT
CAAAGATTCCCAAGGGCTGAATTGCCCTGAAGTTTCTAAGTTGGTT
ACTGACTTTGACTAAGTTTCACACTGAATTTGTGCTACGTTGACTTG
TTGGAATGTTGATGACAGAGCTGACTTTGCTAAGTACATCTGT
GAAAACAAAGACTCTATCTCTCTCCAAAGTTGAAAGGAATTTGGTA
AAAGCCATTGTGGAAAAAGTCTACTGTATTTGTGAAGTTGAAA
ACGATGAAAATGCCCAGCTGACTTTGGCTGGCTGACTTTGG
Claims

1. A polypeptide comprising

   (i) a leader sequence, the leader sequence comprising

      (a) a human serum albumin secretion pre sequence,

      and

      (b) a motif comprising: \(-X_1-X_2-X_3-X_4-X_5-\) where \(X_1\) is phenylalanine, tryptophan, or tyrosine, \(X_2\) is isoleucine, leucine, valine, alanine or methionine, \(X_3\) is leucine, valine, alanine or methionine, \(X_4\) is serine or threonine and \(X_5\) is isoleucine, valine, alanine or methionine wherein the motif is within the human serum albumin secretion pre sequence and is located at positions \(-20, -19, -18, -17\) and \(-16\) of the albumin pre sequence, respectively, in place of the naturally occurring amino acids at those positions, wherein the numbering is such that the \(-1\) residue is the C-terminal amino acid of the native albumin secretion pro sequence; and

   (ii) a desired protein heterologous to the leader sequence, wherein when the polypeptide is expressed from a
host cell, the secretion yield of the heterologous protein is increased relative to the secretion yield of the heterologous protein resulting from a polypeptide comprising the human serum albumin secretion pre-sequence.

2. A polypeptide according to Claim 1 comprising the secretion pre-sequence MKWVHVFLFLFSSAYS.

3. A polypeptide according to any of Claims 1 or 2 wherein the leader sequence comprises a secretion pro-sequence.

4. A polypeptide according to any of Claims 1, 2 or 3 wherein the secretion pre-sequence is fused by a peptide bond at its C-terminal end to the N-terminal amino acid of a secretion pro-sequence, or variant thereof, thereby to form a pre-pro-sequence.

5. A polypeptide according to any of Claims 1 to 3 wherein the secretion pro-sequence is either:

(a) a serum albumin secretion pro-sequence, preferably a human albumin secretion pro-sequence, or a yeast MFα-1 secretion pro-sequence; or
(b) a variant of a secretion pro-sequence defined in (a) which variant has the dibasic pair of amino acids of the secretion pro-sequence upon which it is based and only conservative substitutions at one or more other positions.

6. A polypeptide according to any of Claims 4 to 6 wherein the desired protein is either:

(a) albumin (preferably human albumin);
(b) a variant of albumin which has at least 25% sequence identity to the mature albumin polypeptide;
(c) a fragment of albumin which comprises at least 5% of the complete sequence of the mature albumin polypeptide;
(d) a fusion protein comprising any of (a), (b) or (c);
(e) transferrin (preferably human transferrin);
(f) a variant of transferrin which has at least 25% sequence identity to the mature transferrin polypeptide;
(g) a fragment of transferrin which comprises at least 5% of the complete sequence of the mature transferrin polypeptide;
(h) a fusion protein comprising any of (e), (f) or (g).

An isolated polynucleotide comprising a sequence that encodes the polypeptide defined by any of Claims 4 to 10.

12. A polynucleotide according to Claim 11 comprising the sequence of SEQ 10 No. 15, SEQ ID No.16, SEQ 10 No. 17, SEQ ID No.18 or SEQ 10 No.34.

13. A polynucleotide according to Claim 12 which polynucleotide comprises the sequence of SEQ ID No. 18 or SEQ 10 No.34, wherein the polynucleotide comprises the sequence of SEQ ID No. 24.

14. A polynucleotide according to Claim 13 comprising the sequence of SEQ ID No.25 or a variant or fragment thereof, which variant has the leader sequences of SEQ ID No.24 and encodes either:

(a) a variant of the albumin encoded by SEQ ID No. 25 which has at least 25% sequence identity to the mature albumin polypeptide encoded by SEQ ID No. 25, or
(b) a fragment of the albumin encoded by SEQ ID No.25 which comprises at least 5% of the complete sequence.
of the mature albumin polypeptide encoded by SEQ ID No. 25.

15. A polynucleotide according to Claim 13 which polynucleotide comprises the sequence of SEQ ID No. 18 or SEQ ID No. 34 wherein Me polynucleotide comprises the sequence of SEQ ID No. 27.

16. A polynucleotide according to Claim 13 comprising the sequence of SEQ ID No.21 or a variant or fragment thereof, which variant has the leader sequence of SEQ ID No.27 and encodes either:

(a) a variant of the albumin encoded by SEQ ID No. 21 which has at least 25% sequence identity to the mature albumin polypeptide encoded by SEQ ID No. 21, or
(b) a fragment of the albumin encoded by SEQ ID No.21 which comprises at least 5% of the complete sequence of the mature albumin polypeptide encoded by SEQ ID No. 21.

17. A polynucleotide comprising the sequence of SEQ ID No. 21.

18. A polynucleotide according to any of Claims 14, 16 or 17 wherein the polynucleotide comprises a DNA sequence being a contiguous or non-contiguous fusion of a DNA sequence encoding a heterologous protein with either the DNA sequence SEQ ID No.25 or the DNA sequence SEQ 10 No.21.

19. A polynucleotide which is the complementary strand of a polynucleotide according to any of Claims 11 to 18.

20. A polynucleotide according to any of Claims 11 to 19 comprising an operably linked transcription regulatory region.

21. A polynucleotide according to Claim 20 wherein the transcription regulatory region comprises a transcription promoter.

22. A self-replicable polynucleotide sequence comprising a polynucleotide according to any of Claims 11 to 21.

23. A cell comprising a polynucleotide according to any of Claims 11 to 22, which cell is preferably a eukaryotic cell, more preferably a fungal cell, such as an Aspergillus cell or a yeast cell, most preferably a Saccharomyces, Kluyveromyces, Schizosaccharomyces or Pichia cell.

24. A cell culture comprising a cell according to Claim 23 and culture medium.

25. A cell culture according to Claim 24 wherein the medium contains a mature desired protein as a result of the production of a polypeptide as defined in any of Claims 2 to 21.

26. A process for producing a mature desired protein, composing (1) culturing a cell according to Claim 23 in a culture medium wherein the cell, as a result of the production of a polypeptide as defined in any of Claims 1 to 10, secretes a mature desired protein into the culture medium, and (2) separating the culture medium, containing the secreted mature protein, from the cell.

27. A process according to Claim 26 additionally comprising the step of separating the mature desired protein from the medium and optionally further purifying the mature desired protein.

28. A process according to Claim 27 additionally comprising the step of formulating the thus separated and/or purified mature desired protein with a therapeutically acceptable carrier or diluent thereby to produce a therapeutic product suitable for administration to a human or an animal.

Patentansprüche

1. Polypeptid, umfassend

(i) eine Leadersequenz, wobei die Leadersequenz umfasst

(a) eine Sekretionspräsequenz von humanem Serumalbumin, und
(b) ein Motiv, umfassend: \(-X_1\cdot X_2\cdot X_3\cdot X_4\cdot X_5\). wobei \(X_1\) Phenylalanin, Tryptophan oder Tyrosin ist, \(X_2\) Iso-
leucin, Leucin, Valin, Alanin oder Methionin ist, X₃ Leucin, Valin, Alanin oder Methionin ist, X₄ Serin oder
Threonin ist und X₅ Isoleucin, Valin, Alanin oder Methionin ist, wobei das Motiv innerhalb der Sekretions-
präsequenz von humanem Serumalbumin liegt und anstelle der natürlich vorkommenden Aminosäuren an
Positionen -20, -19, -18, -17 bzw. -16 der Albuminpräsequenz an diesen Positionen lokalisiert ist, wobei
die Nummerierung so ist, dass der -1 Rest die C-terminale Aminosäure der nativen Albuminsekretionspro-
sequenz ist; und

(ii) ein gewünschtes gegenüber der Leadersequenz heterologes Protein, wobei wenn das Polypeptid von einer
Wirtszelle exprimiert wird, die Sekretionsausbeute des heterologen Proteins gegenüber der Sekretionsausbeute
des heterologen Proteins, das von einem die humane Serumalbumin-Sekretionspräsequenz umfassenden Po-
lypeptid resultiert, erhöht ist.

2. Polypeptid nach Anspruch 1, umfassend die Sekretionspräsequenz MKWVFIVSILFLFSSAYS.

3. Polypeptid nach einem beliebigen der Ansprüche 1 oder 2, wobei die Leadersequenz eine Sekretionsprosequenz
umfasst.

4. Polypeptid nach Anspruch 3, wobei die Sekretionspräsequenz durch eine Peptidbindung an ihrem C-terminalen
Ende mit der N-terminalen Aminosäure einer Sekretionsprosequenz oder einer Variante davon fusioniert ist, um
eine Prä-Prosequenz zu bilden.

5. Polypeptid nach Anspruch 3 oder 4, wobei die Sekretionsprosequenz entweder:

(a) eine Sekretionsprosequenz aus Albumin ist, vorzugsweise eine Sekretionsprosequenz aus humanem Al-
bumin, oder eine MFα-1 Sekretionsprosequenz aus Hefe; oder
(b) eine Variante einer in (a) definierten Sekretionsprosequenz ist, wobei die Variante das dibasische Paar der
Aminosäuren der Sekretionsprosequenz, auf welchem sie basiert, und nur konservative Substitutionen an einer
oder mehreren anderen Positionen aufweist.

6. Polypeptid nach Anspruch 3, umfassend die Sequenz:

MKWVFIVSILFLFSSAYSRY₁Y₂Y₃Y₄Y₅

wobei Y₁ Gly oder Ser ist, Y₂ Val oder Leu ist, Y₃ Phe oder Asp ist, Y₄ Arg oder Lys ist und Y₅ Arg oder Lys ist.

7. Polypeptid nach Anspruch 6, wobei Y₁ Gly ist, Y₂ Val ist, und Y₃ Phe ist; oder Y₁ Ser ist, Y₂ Leu ist und Y₃ Asp ist.

8. Polypeptid nach Anspruch 6 oder 7, wobei Y₄ Arg ist und Y₅ Arg ist; Y₄ Lys ist und Y₅ Arg ist; Y₄ Lys ist und Y₅ Lys
ist; oder Y₄ Arg ist und Y₅ Lys ist.

9. Polypeptid nach einem beliebigen der Ansprüche 4 bis 8, wobei die Sequenz des gewünschten Proteins an seinem
N-terminalen Ende mit der C-terminalen Aminosäure der Leadersequenz fusioniert ist.

10. Polypeptid nach einem beliebigen der Ansprüche 4 bis 9, wobei das gewünschte Protein entweder:

(a) Albumin (vorzugsweise humanes Albumin) ist;
(b) eine Variante von Albumin ist, die mindestens 25% Sequenzidentität zu dem maturierten Albuminpolypeptid
aufweist;
(c) ein Fragment von Albumin ist, das mindestens 5% der vollständigen Sequenz des maturierten Albuminpo-
lypeptids umfasst; oder
(d) ein Fusionsprotein ist, umfassend ein beliebiges von (a), (b) oder (c);
(e) Transferrin (vorzugsweise humanes Transferrin) ist;
(f) eine Variante von Transferrin ist, die mindestens 25% Sequenzidentität zu dem maturierten Transferrinpo-
lypeptid aufweist;
(g) ein Fragment von Transferrin ist, das mindestens 5% der vollständigen Sequenz des maturierten Transferrin-
polypeptids umfasst; oder
(h) ein Fusionsprotein ist, umfassend ein beliebiges von (e), (f) oder (g).
11. Isoliertes Polynukleotid, umfassend eine Sequenz, die das durch einen beliebigen der Ansprüche 4 bis 10 definierte Polypeptid kodiert.

12. Polynukleotid nach Anspruch 11, umfassend die Sequenz von SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18 oder SEQ ID NO: 34.

13. Polynukleotid nach Anspruch 12, wobei das Polynukleotid die Sequenz von SEQ ID NO: 18 oder SEQ ID NO: 34 umfasst, wobei das Polynukleotid die Sequenz von SEQ ID NO: 24 umfasst.

14. Polynukleotid nach Anspruch 13, umfassend die Sequenz von SEQ ID NO: 25 oder eine Variante oder ein Fragment davon, wobei die Variante die Leadersequenz von SEQ ID NO: 24 aufweist und entweder:

   (a) eine Variante des durch SEQ ID NO: 25 kodierten Albumins kodiert, die mindestens 25% Sequenzidentität zu dem durch SEQ ID NO: 25 kodierten maturierten Albuminpolypeptid aufweist, oder
   (b) ein Fragment des durch SEQ ID NO: 25 kodierten Albumins kodiert, das mindestens 5% der vollständigen Sequenz des durch SEQ ID NO: 25 kodierten maturierten Albuminpolypeptids umfasst.

15. Polynukleotid nach Anspruch 13, wobei das Polypeptid die Sequenz von SEQ ID NO: 18 oder SEQ ID NO: 34 umfasst, wobei das Polynukleotid die Sequenz von SEQ ID NO: 27 umfasst.

16. Polynukleotid nach Anspruch 13, umfassend die Sequenz von SEQ ID NO: 21 oder eine Variante oder ein Fragment davon, wobei die Variante die Leadersequenz von SEQ ID NO: 27 aufweist und entweder:

   (a) eine Variante des durch SEQ ID NO: 21 kodierten Albumins kodiert, die mindestens 25% Sequenzidentität zu dem durch SEQ ID NO: 21 kodierten maturierten Albuminpolypeptid aufweist, oder
   (b) ein Fragment des durch SEQ ID NO: 21 kodierten Albumins kodiert, das mindestens 5% der vollständigen Sequenz des durch SEQ ID NO: 21 kodierten maturierten Albuminpolypeptids umfasst.

17. Polynukleotid, umfassend die Sequenz von SEQ ID NO: 21.

18. Polynukleotid nach einem beliebigen der Ansprüche 14, 16 oder 17, wobei das Polynukleotid eine DNA-Sequenz umfasst, die eine kontinuierliche oder nichtkontinuierliche Fusion einer DNA-Sequenz ist, die ein heterologes Protein mit entweder der DNA-Sequenz SEQ ID NO: 25 oder der SEQ ID NO: 21 kodiert.


20. Polynukleotid nach einem beliebigen der Ansprüche 11 bis 19, umfassend eine funktionsfähig verknüpfte transkriptionsregulatorische Region.

21. Polynukleotid nach Anspruch 20, wobei die transkriptionsregulatorische Region einen Transkriptionspromotor umfasst.


25. Zellkultur nach Anspruch 24, wobei das Medium ein maturiertes gewünschtes Protein als ein Ergebnis der Herstellung eines wie in einem beliebigen der Ansprüche 12 bis 21 definierten Polypeptids enthält.

26. Verfahren zum Herstellen eines maturierten gewünschten Proteins, umfassend (1) Kultivieren einer Zelle gemäß Anspruch 23 in einem Kulturmedium, wobei die Zelle als ein Ergebnis der Herstellung eines wie in einem beliebigen der Ansprüche 1 bis 10 definierten Polypeptids, ein maturiertes gewünschtes Protein in das Kulturmedium sezerniert,
EP 1 530 639 B1

und (2) Abtrennen des das sezernierte maturierte Protein enthaltende Kulturmedium von der Zelle.


28. Verfahren nach Anspruch 27, zusätzlich umfassend den Schritt des Formulierens des somit abgetrennten und/oder gereinigten maturierten gewünschten Proteins mit einem therapeutisch verträglichen Träger oder Verdünnungsmittel, zum Herstellen eines therapeutischen Produkts, das zur Verabreichung an einen Menschen oder an ein Tier geeignet ist.

Revendications

1. Polypeptide comprenant

(i) une séquence de tête, la séquence de tête comprenant

(a) une préséquence de sécrétion d’albumine sérique humaine, et
(b) un motif comprenant: -X₁-X₂-X₃-X₄-X₅- où X₁ représente une phénylalanine, un tryptophane ou une tyrosine, X₂ représente une isoleucine, une leucine, une valine, une alanine ou une méthionine, X₃ représente une leucine, une valine, une alanine ou une méthionine, X₄ représente une sérine ou une thréonine et X₅ représente une isoleucine, une valine, une alanine ou une méthionine, où le motif se situe au sein de la préséquence de sécrétion d’albumine sérique humaine et est localisé, respectivement, au niveau des positions -20, -19, -18, -17 et -16 de la préséquence d’albumine, à la place des acides aminés existant à l’état naturel au niveau de ces positions, où la numérotation est telle que le résidu -1 est l’acide aminé C-terminal de la proséquence de sécrétion d’albumine native ; et

(ii) une protéine souhaitée hétérologue à la séquence de tête, où lorsque le polypeptide est exprimé à partir d’une cellule hôte, le rendement de la sécrétion de la protéine hétérologue est augmenté par rapport au rendement de la sécrétion de la protéine hétérologue résultant d’un polypeptide comprenant la préséquence de sécrétion d’albumine sérique humaine.

2. Polypeptide selon la revendication 1, comprenant la préséquence de sécrétion MKWVFIVSILFLFSSAYS.

3. Polypeptide selon l’une quelconque des revendications 1 ou 2, dans lequel la séquence de tête comprend une proséquence de sécrétion.

4. Polypeptide selon la revendication 3, dans lequel la préséquence de sécrétion est fusionnée par une liaison peptidique à son extrémité C-terminale à l’acide aminé N-terminal d’une proséquence de sécrétion ou d’un variant de celle-ci, pour former de cette manière une pré-pro-séquence.

5. Polypeptide selon la revendication 3 ou 4, dans lequel la préséquence de sécrétion est l’un ou l’autre de :

(a) une proséquence de sécrétion d’albumine, de préférence une proséquence de sécrétion d’albumine sérique humaine ou une proséquence de sécrétion de MFₐ₋₁ de levure ; ou
(b) un variant d’une proséquence de sécrétion définie en (a), lequel variant possède la paire dibasique d’acides aminés de la proséquence de sécrétion sur laquelle il est basé et seulement des substitutions conservatrices au niveau d’une ou plusieurs autres positions.

6. Polypeptide selon la revendication 3 ou 4, comprenant la séquence :

MKWVFIVSILFLFSSAYSRY₁Y₂Y₃Y₄Y₅

dans laquelle Y₁ représente Gly ou Ser, Y₂ représente Val ou Leu, Y₃ représente Phe ou Asp, Y₄ représente Arg ou Lys et Y₅ représente Arg ou Lys.

7. Polypeptide selon la revendication 6, dans lequel Y₁ représente Gly, Y₂ représente Val et Y₃ représente Phe ; ou
Y1 représente Ser, Y2 représente Leu et Y3 représente Asp.

8. Polypeptide selon la revendication 6 ou 7, dans lequel Y4 représente Arg et Y5 représente Arg ; Y4 représente Lys et Y5 représente Arg ; Y4 représente Lys et Y5 représente Lys ; ou Y4 représente Arg et Y5 représente Lys.

9. Polypeptide selon l’une quelconque des revendications 4 à 8, dans lequel la séquence de la protéine souhaitée est fusionnée à son extrémité N-terminale à l’acide aminé C-terminal de la séquence de tête.

10. Polypeptide selon l’une quelconque des revendications 4 à 9, où la protéine souhaitée est l’une parmi :

(a) l’albumine (de préférence, l’albumine humaine) ;
(b) un variant de l’albumine qui présente au moins 25 % d’identité de séquence avec le polypeptide de l’albumine mature ;
(c) un fragment d’albumine qui comprend au moins 5 % de la séquence complète du polypeptide de l’albumine mature ; ou
(d) une protéine de fusion comprenant l’un quelconque de (a), (b) ou (c) ;
(e) la transférine (de préférence, la transférine humaine) ;
(f) un variant de la transférine qui présente au moins 25 % d’identité de séquence avec le polypeptide de la transférine mature ;
(g) un fragment de la transférine qui comprend au moins 5 % de la séquence complète du polypeptide de la transférine mature ; ou
(h) une protéine de fusion comprenant l’un quelconque de (e), (f) ou (g).

11. Polynucléotide isolé comprenant une séquence qui code pour le polypeptide défini par l’une quelconque des revendications 4 à 10.

12. Polynucléotide selon la revendication 11, comprenant la séquence de SEQ ID NO : 15, SEQ ID NO : 16, SEQ ID NO : 17, SEQ ID NO : 18 ou SEQ ID NO : 34.

13. Polynucléotide selon la revendication 12, lequel polynucléotide comprend la séquence de SEQ ID NO : 18 ou SEQ ID NO : 34, où le polynucléotide comprend la séquence de SEQ ID NO : 24.

14. Polynucléotide selon la revendication 13, comprenant la séquence de SEQ ID NO : 25 ou un variant ou un fragment de celle-ci, lequel variant a la séquence de tête de SEQ ID NO : 24 et code pour l’un ou l’autre de :

(a) un variant de l’albumine codée par SEQ ID NO : 25 qui présente au moins 25 % d’identité de séquence avec le polypeptide de l’albumine mature codé par SEQ ID NO : 25, ou
(b) un fragment de l’albumine codée par SEQ ID NO : 25 qui comprend au moins 5 % de la séquence complète du polypeptide de l’albumine mature codé par SEQ ID NO : 25.

15. Polynucléotide selon la revendication 13, lequel polypeptide comprend la séquence de SEQ ID NO : 18 ou SEQ ID NO : 34, où le polynucléotide comprend la séquence de SEQ ID NO : 27.

16. Polynucléotide selon la revendication 13, comprenant la séquence de SEQ ID NO : 21 ou un variant ou un fragment de celle-ci, lequel variant a la séquence de tête de SEQ ID NO : 27 et code pour l’un ou l’autre de :

(a) un variant de l’albumine codée par SEQ ID NO : 21 qui présente au moins 25 % d’identité de séquence avec le polypeptide de l’albumine mature codé par SEQ ID NO : 21, ou
(b) un fragment de l’albumine codée par SEQ ID NO : 21 qui comprend au moins 5 % de la séquence complète du polypeptide de l’albumine mature codé par SEQ ID NO : 21.

17. Polynucléotide comprenant la séquence de SEQ ID NO : 21.

18. Polynucléotide selon l’une quelconque des revendications 14, 16 ou 17, où le polynucléotide comprend une séquence d’ADN qui est une fusion contiguë ou non contiguë d’une séquence d’ADN codant pour une protéine hétérologue avec soit la séquence d’ADN SEQ ID NO : 25 soit la séquence d’ADN SEQ ID NO : 21.

20. Polynucléotide selon l’une quelconque des revendications 11 à 19, comprenant une région régulatrice de la transcription liée de manière fonctionnelle.

21. Polynucléotide selon la revendication 20, où la région régulatrice de la transcription comprend un promoteur de la transcription.


23. Cellule comprenant un polynucléotide selon l’une quelconque des revendications 11 à 22, laquelle cellule est de préférence une cellule eucaryote, de manière davantage préférée une cellule fongique, telle qu’une cellule d’Aspergillus ou une cellule de levure, de manière préférée entre toutes une cellule de Saccharomyces, Kluyveromyces, Schizosaccharomyces ou Pichia.

24. Culture cellulaire comprenant une cellule selon la revendication 23 et un milieu de culture.

25. Culture cellulaire selon la revendication 24, où le milieu contient une protéine mature souhaitée comme résultat de la production d’un polypeptide tel que défini dans l’une quelconque des revendications 12 à 21.

26. Procédé de production d’une protéine mature souhaitée, comprenant (1) la culture d’une cellule selon la revendication 23 dans un milieu de culture dans lequel la cellule, comme résultat de la production d’un polypeptide tel que défini dans l’une quelconque des revendications 1 à 10, sécrète une protéine mature souhaitée dans le milieu de culture, et (2) la séparation du milieu de culture, contenant la protéine mature souhaitée, de la cellule.

27. Procédé selon la revendication 26, comprenant en outre l’étape consistant à séparer la protéine mature souhaitée du milieu et éventuellement à purifier en outre la protéine mature souhaitée.

28. Procédé selon la revendication 27, comprenant en outre l’étape consistant à formuler la protéine mature souhaitée ainsi séparée et/ou purifiée avec un support ou un diluant thérapeutiquement acceptable pour produire de cette manière un produit thérapeutique approprié pour une administration à un être humain ou un animal.
### Fig. 1

<table>
<thead>
<tr>
<th>Pre</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSA</td>
<td>HSA</td>
</tr>
<tr>
<td>HSA</td>
<td>MFA-1</td>
</tr>
</tbody>
</table>

| | Met | Lys | Trp | Val | Thr | Phe | Ile | Ser | Leu | Leu | Phe | Leu | Phe | Ser | Ser | Ala | Tyr | Ser | Arg | Gly | Val | Phe | Arg | Arg |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| HSA | Met | Lys | Trp | Val | Thr | Phe | Ile | Ser | Leu | Leu | Phe | Leu | Phe | Leu | Phe | Leu | Phe | Ser | Ser | Ala | Tyr | Ser | Arg | Ser | Leu | Asp | Lys | Arg |
| HSA | -24 | -23 | -22 | -21 | -20 | -19 | -18 | -17 | -16 | -15 | -14 | -13 | -12 | -11 | -10 | -9  | -8  | -7  | -6  | -5  | -4  | -3  | -2  | -1  |

**Preferred mutations of the invention:**

- Phe
- Ile
- Val
- Ile
### Fig. 2

**Standard genetic code**

<table>
<thead>
<tr>
<th>T</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT Phe (F)</td>
<td>TCT Ser (S)</td>
<td>TAT Tyr (Y)</td>
<td>TGT Cys (C)</td>
</tr>
<tr>
<td>TTC Phe (F)</td>
<td>TCC Ser (S)</td>
<td>TAC Tyr (Y)</td>
<td>TGC Cys (C)</td>
</tr>
<tr>
<td>TTA Leu (L)</td>
<td>TCA Ser (S)</td>
<td>TAA Ter</td>
<td>TGA Ter</td>
</tr>
<tr>
<td>TTG Leu (L)</td>
<td>TCG Ser (S)</td>
<td>TAG Ter</td>
<td>TGG Trp (W)</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTT Leu (L)</td>
<td>CCT Pro (P)</td>
<td>CAT His (H)</td>
<td>CGT Arg (R)</td>
</tr>
<tr>
<td>CTC Leu (L)</td>
<td>CCC Pro (P)</td>
<td>CAC His (H)</td>
<td>CGC Arg (R)</td>
</tr>
<tr>
<td>CTA Leu (L)</td>
<td>CCA Pro (P)</td>
<td>CAA Gln (Q)</td>
<td>CGA Arg (R)</td>
</tr>
<tr>
<td>CTG Leu (L)</td>
<td>CCG Pro (P)</td>
<td>CAG Gln (Q)</td>
<td>CGG Arg (R)</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATT Ile (I)</td>
<td>ACT Thr (T)</td>
<td>AAT Asn (N)</td>
<td>AGT Ser (S)</td>
</tr>
<tr>
<td>ATC Ile (I)</td>
<td>ACC Thr (T)</td>
<td>AAC Asn (N)</td>
<td>AGC Ser (S)</td>
</tr>
<tr>
<td>ATA Ile (I)</td>
<td>ACA Thr (T)</td>
<td>AAA Lys (K)</td>
<td>AGA Arg (R)</td>
</tr>
<tr>
<td>ATG Met (M)</td>
<td>ACG Thr (T)</td>
<td>AAG Lys (K)</td>
<td>AGG Arg (R)</td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTT Val (V)</td>
<td>GCT Ala (A)</td>
<td>GAT Asp (D)</td>
<td>GGT Gly (G)</td>
</tr>
<tr>
<td>GTC Val (V)</td>
<td>GCC Ala (A)</td>
<td>GAC Asp (D)</td>
<td>GGC Gly (G)</td>
</tr>
<tr>
<td>GTA Val (V)</td>
<td>GCA Ala (A)</td>
<td>GAA Glu (E)</td>
<td>GGA Gly (G)</td>
</tr>
<tr>
<td>GTG Val (V)</td>
<td>GCG Ala (A)</td>
<td>GAG Glu (E)</td>
<td>GGG Gly (G)</td>
</tr>
</tbody>
</table>

**Single letter code:**

A = adenosine

C = cytidine

G = guanosine

T = thymidine

B = C or G or T

D = A or G or T

H = A or C or T

K = G or T

M = A or C

N = A or C or G or T

R = A or G

S = C or G

V = A or C or G

W = A or T

Y = C or T
### Fig. 3

**Modified list of preferred yeast codons**

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>TTC Phe (F)</td>
<td>TCT Ser (S)</td>
<td>TAC Tyr (Y)</td>
<td>TGT Cys (C)</td>
</tr>
<tr>
<td></td>
<td>TTG Leu (L)</td>
<td>TCC Ser (S)</td>
<td>TAA Ter</td>
<td>TGG Trp (W)</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>CCA Pro (P)</td>
<td>CAT His (H)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CAA Gln (Q)</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>ATT Ile (I)</td>
<td>ACT Thr (T)</td>
<td>AAC Asn (N)</td>
<td>AGA Arg (R)</td>
</tr>
<tr>
<td></td>
<td>ATC Ile (I)</td>
<td>ACC Thr (T)</td>
<td>AAG Lys (K)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATG Met (M)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>GTT Val (V)</td>
<td>GCT Ala (A)</td>
<td>GAT Asp (D)</td>
<td>GGT Gly (G)</td>
</tr>
<tr>
<td></td>
<td>GTC Val (V)</td>
<td></td>
<td>GAC Asp (D)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAA Glu (E)</td>
<td></td>
</tr>
</tbody>
</table>
**Fig. 4**

**pAYE438**

4059 bps
Fig. 5

pAYE441

4510 bps
Fig. 7

pAYE467
6327 bps
Fig. 8

pAYE443
14161 bps
Fig. 10

pAYE655
14169 bps
**Fig. 11**

![Diagram of pAYE639](image)

- **pAYE639**
- **4763 bps**
- **rHA codon optimised**
- **Amp**
- **HindIII**
- **EcoRI**
- **Fusion leader**
- **EcoRI**
- **HindIII**
- **BglIII**
Fig. 13

pAYE466
4471 bps

ADHt

BamHI, NotI

HindIII

PstI

PRBp'

ori

NotI

Amp
Fig. 14

pAYE640
6330 bps
rHA codon optimised
Fig. 16

pAYE643
4763 bps

HindIII
EcoRI
Amp
rHA codon optimised
Modified fusion leader
EcoRI
HindIII
BglII
Fig. 17

pAYE645
6330 bps
rHA codon optimised

NotI
PstI
HindIII
BglII
EcoRI

PRFlp

Modified leader sequence

ADH1

NotI
BamHI
BglII
HindIII
EcoRI
EcoRI
EcoRI

Fig. 19

pAYE647
14172 bps
Fig. 21

<table>
<thead>
<tr>
<th>Strain</th>
<th>Plasmid</th>
<th>$Y_{x/s}$ (g/g)</th>
<th>$Y_{p/s}$ (mg/g)</th>
<th>rHA (g/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>pAYE443 (1st feeds)</td>
<td>0.33</td>
<td>10.4</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>(fill and draw)</td>
<td>0.34</td>
<td>11.5</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>pAYE638 (1st feeds)</td>
<td>0.36</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>(fill and draw)</td>
<td>0.36</td>
<td>2.4</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>pAYE646 (1st feeds)</td>
<td>0.33</td>
<td>11.6</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>(fill and draw)</td>
<td>0.35</td>
<td>12.2</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>pAYE655 (1st feeds)</td>
<td>0.37</td>
<td>12.1</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>(fill and draw)</td>
<td>0.35</td>
<td>13.0</td>
<td>3.7</td>
</tr>
<tr>
<td>B</td>
<td>pAYE443 (1st feeds)</td>
<td>0.35</td>
<td>10.5</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>pAYE646 (1st feeds)</td>
<td>0.35</td>
<td>13.0</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>(fill and draw)</td>
<td>0.33</td>
<td>12.8</td>
<td>3.6</td>
</tr>
</tbody>
</table>
**Fig. 22**

<table>
<thead>
<tr>
<th>Plasmid designation</th>
<th>Leader</th>
<th>Desired Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAYE443</td>
<td>HSA/MFα-1</td>
<td>HSA</td>
</tr>
<tr>
<td>pAYE467</td>
<td>natural codon bias</td>
<td>natural codon bias</td>
</tr>
<tr>
<td>pAYE655</td>
<td>modified HSA/MFα-1</td>
<td>HSA</td>
</tr>
<tr>
<td>pAYE643</td>
<td>FIVSI fully codon biased the rest has natural codon bias</td>
<td>natural codon bias</td>
</tr>
<tr>
<td>pAYE638</td>
<td>HSA/MFα-1</td>
<td>HSA</td>
</tr>
<tr>
<td>PAYE639</td>
<td>all fully codon biased</td>
<td></td>
</tr>
<tr>
<td>pAYE640</td>
<td></td>
<td>fully codon biased</td>
</tr>
<tr>
<td>pAYE642</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAYE645</td>
<td>modified HSA/MFα-1</td>
<td>HSA</td>
</tr>
<tr>
<td>pAYE646</td>
<td>all fully codon biased</td>
<td></td>
</tr>
<tr>
<td>pAYE647</td>
<td></td>
<td>fully codon biased</td>
</tr>
</tbody>
</table>
REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader’s convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

• EP 366400 A [0006]
• WO 9533833 A [0006]
• WO 9001063 A [0006] [0008] [0042] [0044] [0050] [0053] [0059] [0103] [0108] [0129]
• WO 0179258 A [0025]
• WO 0179271 A [0025] [0038]
• WO 0179442 A [0025]
• WO 0179443 A [0025]
• WO 0179444 A [0025]
• WO 0179480 A [0025]
• EP 322094 A [0035]
• EP 0286422 A [0042]
• WO 9637515 A [0042] [0044] [0122] [0146]
• EP 308381 A [0076]
• WO 9816643 A [0078]
• EP 258067 A [0090] [0108]
• JP 3072889 A [0092]
• JP 60041487 A [0092]
• EP 504823 A [0092]
• JP 63039576 A [0092]
• EP 251744 A [0092] [0108]
• JP 2104290 A [0092]
• EP 344459 A [0092]
• EP 322094 A [0035]
• EP 424117 A [0103]
• EP 286424 A [0103]
• US 4440859 A, Rutter [0107]
• US 4530901 A, Weissman [0107]
• US 4582800 A, Crowl [0107]
• US 4677063 A, Mark [0107]
• US 4678751 A, Goeddel [0107]
• US 4704362 A, Itakura [0107]
• US 4710463 A, Murray [0107]
• US 4766075 A [0107]
• US 4810648 A, Stalker [0107]
• US 4945050 A [0111]
• EP 332581 A [0111]
• WO 9413822 A, Chang [0111]
• EP 330451 A [0118]
• EP 361991 A [0118]
• WO 9204367 A [0122]
• EP 464590 A [0122]
• EP 319067 A [0122]
• US 5728553 A [0122]
• WO 0044772 A [0122] [0131] [0140]
• WO 9724445 A [0131] [0140]
• EP 0431880 A [0131]

Non-patent literature cited in the description

• Thompson et al. Nucleic Acids Res., 1994, vol. 22 (22), 4673-80 [0031]
• Hoefs, J.C. Hepatology, 1992, vol. 16, 396-403 [0035]
• Testa. Proteins of iron metabolism. CRC Press, 2002 [0036]
• Mason et al. Biochemistry, 2002, vol. 41, 9448 [0036]
• Sharp; Crowe. Yeast, 1991, vol. 7, 657-678 [0073]
• Hoffman; Winston. Genetics, 1990, vol. 124, 807-816 [0091]
• Gietz, R.D.; Sugino. A. Gene, 1988, vol. 74, 527-534 [0103]
• Sherman et al. Methods In Yeast Genetics, A Laboratory Manual. 1986 [0108]
• Sherman et al. Methods In Yeast Genetics, A Laboratory Manual. 1986 [0108]
• Hinchee et al. Biotechnology, 1988, vol. 6, 915-921 [0111]
• Paszkowski et al. EMBO J., 1984, vol. 3, 2717-2722 [0111]
• McCabe et al. Biotechnology, 1988, vol. 6, 923-926 [0111]
• McCabe et al. Bio/Technology, 1988, vol. 6, 923-926 [0111]
• Klein et al. Bio/Technology, 1988, vol. 6, 559-563 [0111]
• Sleep et al. Biotechnology, 1991, vol. 9, 183-187 [0131] [0142]