In the course of *P. pastoris* genome sequencing 54 different sequences were listed as predicted signal peptides which include a cleavage site to provide for the secretion of proteins. (De Schutter et al. Nature Biotechnology doi: 10.1038/nbt.1544 2009).

US2011/0021378A1 describes a set of 54 *P. pastoris* genes identified that contain a signal sequence, among them residues 1-21 of SEQ ID 8 as listed in US2011/0021378A1, again including a cleavage site to provide for the secretion of proteins.

EP2258855A1 describes regulatory sequences of a *P. pastoris* derived expression system, wherein a signal and leader sequence of the *P. pastoris* Epx1 protein is used as a 57 amino acid precursing sequence to facilitate expression and secretion of a protein of interest POI. A cleavage site for a signal peptidase is predicted to be following the signal sequence, i.e. between position 20 and 21, indicated as a hyphen in the sequence of the cleavage site: VSA-AP (SEQ ID 6).

It is desirable to provide alternative regulatory elements suitable for expressing a POI in a recombinant eukaryotic host cell, and methods to produce secreted proteins in eukaryotic cells which are simple and efficient, and could preferably lead to a homogeneous N-terminus of the POI.

SUMMARY

The object is solved by the subject matter as claimed.

In one aspect, there is provided an isolated nucleic acid encoding a leader, which is selected from the group consisting of

a) a leader peptide with the amino acid sequence of SEQ ID 10 or a functional variant thereof with one or two point mutations, and

b) a leader peptide with the amino acid sequence selected from the group consisting of SEQ ID 11, 12, 13 and 14,

preferably which is a nucleotide sequence encoding a leader peptide, selected from the group consisting of SEQ ID 18, 19 and 20, or a codon optimized variant of any of SEQ ID 18, 19 or 20.

In one aspect, there is provided an expression cassette comprising a nucleic acid encoding a leader peptide directly fused to a nucleic acid encoding a heterologous protein of interest (POI), which leader peptide is selected from the group consisting of

a) a leader peptide consisting of the amino acid sequence SEQ ID 1 or a functional variant thereof, wherein the variation consists of one or two point mutations, and

b) a leader peptide consisting of the amino acid sequence selected from the group consisting of SEQ ID 2, 3, 4 and 5,

wherein the expression cassette excludes a nucleic acid encoding the native leader of the *P. pastoris* Epx1 protein identified as SEQ ID 8.

The disclosure provides for an isolated nucleic acid encoding a leader, which is selected from the group consisting of

a) a leader peptide with the amino acid sequence of SEQ ID 10 or a functional variant thereof with one or two point mutations,

b) a leader peptide with the amino acid sequence selected from the group consisting of SEQ ID 11, 12, 13 and 14,

c) a signal peptide with the amino acid sequence of SEQ ID 1 or a functional variant thereof with one or two point mutations, preferably excluding SEQ ID 2, and

d) a signal peptide with the amino acid sequence selected from the group consisting of SEQ ID 2, 3, 4 and 5, preferably excluding SEQ ID 2.

Specifically, the nucleic acid encoding the truncated leader comprises or consists of a nucleic acid encoding a signal peptide selected from the group consisting of SEQ ID 2, 3, 4 and 5, preferably excluding SEQ ID 2. Specifically, the nucleic acid encoding the truncated leader comprises or consists of a nucleic acid encoding a leader peptide selected from the group consisting of SEQ ID 11, 12, 13, and 14.

Specifically, the truncated leader comprises or consists of a the signal peptide selected from the group consisting of SEQ ID 2, 3, 4 and 5, preferably excluding SEQ ID 2, or of a leader peptide selected from the group consisting of SEQ ID 11, 12, 13, and 14.

In one aspect, there is provided an isolated leader encoded by a nucleic acid as defined herein, preferably consisting of an amino acid sequence acid sequence selected from the group consisting of SEQ ID 10, 11, 12, 13 and 14.

A specifically preferred nucleic acid encodes a signal peptide of the amino acid sequence of SEQ ID 3 or a leader peptide of SEQ ID 12; preferably the nucleic acid sequence encoding the signal peptide or the leader peptide is SEQ ID 16 or SEQ ID 19, respectively. A specifically preferred leader is a signal peptide having the amino acid sequence of SEQ ID 3 or a leader peptide having the amino acid sequence of SEQ ID 12.

Accordingly, a specific embodiment of the disclosure refers to a leader or a nucleic acid encoding such leader, which is

a) a signal peptide with the amino acid sequence of SEQ ID 3, preferably wherein the coding nucleic acid consists of a nucleotide sequence of SEQ ID 16, or a codon optimized variant of SEQ ID 16; or

b) a leader peptide with, or comprising or consisting of the amino acid sequence of SEQ ID 12, preferably wherein the coding nucleic acid consists of a nucleotide sequence of SEQ ID 19, or a codon optimized variant of SEQ ID 19.

Specifically, the nucleic acid according to the disclosure has

a) a nucleotide sequence encoding a signal peptide, selected from the group consisting of SEQ ID 15, 16 and 17, preferably excluding SEQ ID 15, or

b) a nucleotide sequence encoding a leader peptide, selected from the group consisting of SEQ ID 18, 19 and 20, or

c) a nucleotide sequence which is a codon optimized variant of SEQ ID 15, 16, 17, 18, 19 or 20.

A specific nucleic acid of the disclosure encodes a leader, which is a signal peptide with the amino acid sequence of SEQ ID 3, preferably wherein the nucleic acid consists of a nucleotide sequence of SEQ ID 16, or a codon optimized variant of SEQ ID 16.

CLAIMS

1. Isolated nucleic acid encoding a leader, which is selected from the group consisting of

a) a leader peptide with the amino acid sequence of SEQ ID 10 or a functional variant thereof with one or two point mutations, and

b) a leader peptide with the amino acid sequence selected from the group consisting of SEQ ID 11, 12, 13 and 14,

preferably which is a nucleotide sequence encoding a leader peptide, selected from the group consisting of SEQ ID 18, 19 and 20, or a codon optimized variant of any of SEQ ID 18, 19, or 20.

2. The nucleic acid of claim 1, encoding a leader which is a leader peptide with the amino acid sequence of SEQ ID 12, preferably wherein the nucleic acid consists of a nucleotide sequence of SEQ ID 19, or a codon optimized variant of SEQ ID 19.

3. An expression cassette comprising a nucleic acid encoding a leader peptide directly fused to a nucleic acid encoding a heterologous protein of interest (POI), which leader peptide is selected from the group consisting of

a) a leader peptide consisting of the amino acid sequence SEQ ID 1 or a functional variant thereof, wherein the variation consists of one or two point mutations, and

b) a leader peptide consisting of the amino acid sequence selected from the group consisting of SEQ ID 2, 3, 4 and 5,

wherein the expression cassette excludes a nucleic acid encoding the native leader of the *P. pastoris* Epx1 protein identified as SEQ ID 8.

4. The expression cassette of claim 3, wherein the nucleotide sequence encoding the leader peptide is selected from the group consisting of SEQ ID 15, 16 and 17, or a codon optimized variant of any of SEQ ID 15, 16, or 17.

5. The expression cassette of claim 3, wherein the leader peptide consists of the amino acid sequence SEQ ID 3, preferably wherein the nucleic acid encoding the leader peptide consists of a nucleotide sequence of SEQ ID 16, or a codon optimized variant of SEQ ID 16.

6. The nucleic acid according to any of claim 1 or 2, which is a DNA.

7. The expression cassette according to any of claims 3 to 5, which is a DNA.

8. Isolated leader encoded by a nucleic acid according to claim 1 or 2, preferably consisting of an amino acid sequence selected from the group consisting of SEQ ID 10, 11, 12, 13 and 14.

9. The leader of claim 8, which is a peptide consisting of the amino acid sequence SEQ ID 12.

10. Expression cassette comprising a nucleic acid encoding a leader operably linked to a nucleic acid sequence encoding a POI, characterized in that the leader is selected from the group consisting of

a) a leader peptide with the amino acid sequence of SEQ ID 10 or a functional variant thereof with one or two point mutations,

b) a leader peptide with the amino acid sequence selected from the group consisting of SEQ ID 11, 12, 13 and 14, preferably which is a nucleotide sequence encoding a leader peptide, selected from the group consisting of SEQ ID 18, 19 and 20, or a codon optimized variant of any of SEQ ID 18, 19, or 20,

c) a signal peptide with the amino acid sequence of SEQ ID 1 or a functional variant thereof with one or two point mutations, and

d) a signal peptide with the amino acid sequence selected from the group consisting of SEQ ID 2, 3, 4 and 5, preferably which is a nucleotide sequence encoding a signal peptide, selected from the group consisting of SEQ ID 15, 16 and 17, or a codon optimized variant of any of SEQ ID 15, 16, or 17.

11. The expression cassette according to claim 10, wherein said POI comprises an amino acid sequence with a native N-terminal amino acid sequence.

12. The expression cassette according to claim 10 or 11, wherein the POI is selected from therapeutic proteins, including antibodies or fragments thereof, enzymes and peptides, protein antibiotics, toxin fusion proteins, carbohydrate - protein conjugates, structural proteins, regulatory proteins, vaccines and vaccine-like proteins or particles, process enzymes, growth factors, hormones and cytokines, or wherein said POI mediates the production of a host cell metabolite, preferably selected from the group consisting of antibodies or fragments thereof, growth factors, hormones and cytokines.

13. The expression cassette according to claim 12, wherein the POI is selected from the group consisting of growth factors, hormones, cytokines, antibody and antibody fragments, preferably selected from the group consisting of a full-length antibody, an scFv, minibody, diabody, triabody, tetrabody, Fab and a Fc-fusion protein.

14. The expression cassette according to any of claims 10 to 13, which further comprises a promoter operably linked to the nucleic acid encoding the leader.

15. Recombinant yeast host cell or a vector comprising the expression cassette according to any of claims 10 to 14.

16. The host cell according to claim 15, wherein the yeast is selected from the genus group consisting of Pichia, Candida, Torulopsis, Arxula, Hansenula, Ogatea, Yarrowia, Kluyveromyces, Saccharomyces and Komagataella.

17. A method of producing a POI in a yeast host cell, comprising:
- providing the host cell according to claim 15 or 16,
- cultivating said host cell to express said POI, and
- purifying the POI to obtain a preparation of a purified POI.

18. The method according to claim 17, wherein said POI comprises an amino acid sequence with a native N-terminal amino acid sequence.

19. The method according to claim 17 or 18, wherein said POI comprises an amino acid sequence that does not comprise an additional Alanine as N-terminal amino acid residue.

20. The method according to any of claims 17 to 18, wherein the POI is selected from therapeutic proteins, including antibodies or fragments thereof, enzymes and peptides, protein antibiotics, toxin fusion proteins, carbohydrate - protein conjugates, structural proteins, regulatory proteins, vaccines and vaccine-like proteins or particles, process enzymes, growth factors, hormones and cytokines, or wherein said POI mediates the production of a host cell metabolite.

21. The method according to claim 20, wherein the POI is selected from the group consisting of growth factors, hormones, cytokines, antibody and antibody fragments, preferably selected from the group consisting of a full-length antibody, an scFv, minibody, diabody, triabody, tetrabody, Fab and a Fc-fusion protein.

22. The use of the nucleic acid according to claim 1 or 2, or the expression cassette according to any of claims 3 to 5 or 7, or the leader according to claim 8 or 9, for the secretion of a POI from a host cell and/or to increase the secretion of a POI from a host cell.

23. Use according to claim 22, wherein at least 60, 65, 70, 75, 80, 85, 90, 95, 98, or 100% of the secreted POI comprises a native N-terminal amino acid sequence.