

TAGS FOR ENHANCED EXPRESSION OF RECOMBINANT PROTEINS

FIELD OF THE INVENTION

The present invention is in the field of recombinant biotechnology, in particular in the field of protein expression. The invention generally relates to peptide tags capable of increasing the yield and/or titer and/or solubility of at least one protein of interest in a host cell, in particular difficult-to-express recombinant proteins. More specifically, the invention relates fusion proteins comprising such peptide tags and to nucleotide sequences encoding such peptide tags and to host cells, expression vectors and expression cassettes comprising such nucleotide sequences. The invention further relates to a method of producing at least one protein of interest in a recombinant host cell using said peptide tags.

BACKGROUND OF THE INVENTION

Escherichia coli is one of the most commonly used host organisms for the production of biopharmaceuticals, as it allows for cost-efficient and fast recombinant protein expression. Today, about 30% of approved biopharmaceuticals are produced by *E. coli*, and despite the recent advance of other production systems such as mammalian cell culture, new product approvals, like Beovu® and Cablivi® (2019), are constantly being granted, allowing *E. coli* to compete with both mammalian and non-mammalian hosts (McElawain et al. 2022, Baeshen et al. 2015). However, while high yield of several proteins can often be achieved, the production efficiency of numerous proteins is still hampered, resulting in reduced production output and non-profitable manufacturing conditions.

There have been numerous attempts to increase yields and production efficiency of a protein of interest (POI) by targeting different steps of the gene expression cascade within cells. These include methods targeting transcriptional and translational efficiency and regulation as well as post-transcriptional and post-translational modifications increasing mRNA and protein stability and solubility.

For instance, several approaches target the problem of insufficient protein yields at an early stage of the gene expression cascade by e.g. the rise of transcript abundance utilizing strong promoters or increased copy numbers of the heterologous gene. However, it has been shown that these measures clearly reach a plateau,

providing evidence for the existence of further bottlenecks downstream of transcription such as translation efficiency for example.

The initial coding nucleotide sequence, especially the downstream box (DB), has been proposed to contribute to translation efficiency and fidelity. While it was initially believed that the base pair complementary of the DB with the 16S ribosomal RNA stabilizes the interaction of the target transcript with the ribosome, it has been later shown that base pairing between the DB and the 16S ribosomal RNA is dispensable for the DB-associated enhancer activity (Moll et al. 2001; O'Connor et al. 1999, Sprengart et al. 1996). As the precise mechanism of the function of the DB remains elusive, the *a priori* design of effective DB sequences for a given gene is still unattainable. Moreover, the effects of the DB are target-gene dependent, necessitating empirical POI-specific DB optimization (Richter et al., 2018).

In conclusion, there is still a need for methods to improve the production of recombinant POIs in host cells. The technical problem underlying the present invention is to comply with this need.

Here, we provide a solution to the technical problem of unsatisfying recombinant protein production conditions by introducing novel artificial peptide tags that boost recombinant protein yields.

Accordingly, the present invention provides new tools to enhance the protein yield, solubility and/or titer of recombinant proteins in host cells that are also suitable for industrial production. Besides novel peptide tags for the production of recombinant proteins, the invention includes recombinant proteins containing said peptide tags and methods for the production and expression thereof, including expression vectors.

BRIEF SUMMARY OF THE INVENTION

Recombinant protein production is, while being of highest importance for the biopharmaceutical industry, still hampered by several limitations. Among other features, it has been shown that translation efficiency is a key aspect for protein production efficiency. Besides the importance of recombinant protein production, currently available expression systems and tools still suffer from low translation efficiency and expression titers, consequently resulting in inefficient recombinant protein production conditions.

It is the objective of the present invention to overcome current limitations in protein biosynthesis by introducing a novel peptide tag for the high-titer expression of proteins.

According to the invention, there is provided a peptide tag for expression of a recombinant protein of interest (POI) in a host cell, comprising a first peptide sequence consisting of an amino acid sequence selected from SEQ ID NO:1 or SEQ ID NO:2.

According to a specific embodiment the peptide tag comprises a second peptide sequence comprising an amino acid sequence selected from SEQ ID NO:5, SEQ ID NO:10 or SEQ ID NO:21, preferably wherein said second peptide sequence is C-terminal of the first peptide sequence. Preferably, the second peptide sequence comprises an amino acid sequence selected from SEQ ID NO:10 and SEQ ID NO:21.

In a specific example the peptide tag comprises a second peptide sequence selected from SEQ ID NO:107 or SEQ ID:140, preferably wherein said second peptide sequence is C-terminal of the first peptide sequence.

In a specific example the peptide tag comprises at least two first peptide sequences, selected from the group consisting of SEQ ID NO:1 or SEQ ID NO:2, preferably wherein said at least two first peptide sequences are directly linked.

Specifically, the first peptide sequence of the peptide tag is directly linked to the second peptide sequence. Specifically, the first peptide sequence is directly linked to the N-terminus of the second peptide sequence.

Specifically, the peptide tag further comprises one or more linker sequences.

The peptide tag described herein may further comprise a solubility and/or expression enhancement tag, a monitoring tag and/or an affinity tag.

Specifically, the peptide tag further comprises a protease recognition and/or cleavage site, preferably at its C-terminus.

Further specifically provided herein is a host cell comprising the peptide tag described herein, said host cell is a eukaryotic or prokaryotic host cell, preferably a yeast cell, a mammalian cell or a bacterial cell.

Specifically provided herein is a peptide tag for expression of a recombinant protein of interest (POI) in a host cell, comprising an amino acid sequence of SEQ ID NO:84, wherein

- a. X at position 2 is L or H;
- b. X at position 3 is V or S;
- c. X at position 16 is E or Q; and
- d. X at position 18 is E or Q.

5 Further specifically described herein is a peptide tag comprising an amino acid sequence selected from the group consisting of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:32, SEQ ID NO:35, SEQ ID NO:42 and SEQ ID NO:47, or a functionally active variant thereof. Specifically, such peptide and functionally active variants thereof, when used as an expression tag, facilitate
10 the expression of a POI.

Further specifically described herein is a peptide tag comprising an amino acid sequence selected from the group consisting of SEQ ID NO:150 and SEQ ID NO:155, or a functionally active variant thereof. Specifically, such peptide and functionally active variants thereof, when used as an expression tag, facilitate the ex-
15 pression of a POI.

Specifically described herein is a peptide tag comprising an amino acid sequence selected from the group consisting of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:32, SEQ ID NO:35, SEQ ID NO:42 and SEQ ID NO:47, or a functionally active variant thereof comprising an amino acid se-
20 quence having at least 95% sequence identity to SEQ ID NO:10, SEQ ID NO:21, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87 or SEQ ID NO:88 and comprising SEQ ID NO:1 or SEQ ID NO:2, preferably at its N-terminus.

Specifically described herein is a peptide tag comprising an amino acid sequence selected from the group consisting of SEQ ID NO:150 and SEQ ID NO:155, or a functionally active variant thereof comprising an amino acid sequence having
25 at least 95% sequence identity to SEQ ID NO:160 and comprising SEQ ID NO:1 or SEQ ID NO:2, preferably at its N-terminus.

Specifically, described herein is a functionally active variant of a peptide tag comprising an amino acid sequence selected from the group consisting of SEQ ID
30 NO:11, SEQ ID NO:12, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:32, SEQ ID NO:35, SEQ ID NO:42 and SEQ ID NO:47, said functionally active variant comprising SEQ ID NO:1 or SEQ ID NO:2 and comprising at least 95% sequence identity

to SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:32, SEQ ID NO:35, SEQ ID NO:42 or SEQ ID NO:47.

Specifically, described herein is a functionally active variant of a peptide tag comprising an amino acid sequence selected from the group consisting of SEQ ID NO:150 and SEQ ID NO:155, said functionally active variant comprising SEQ ID NO:150 or SEQ ID NO:155.

In yet a further specific example, the peptide tag described herein comprises SEQ ID NO:1 or SEQ ID NO:2 directly linked to the N-terminus of an amino acid sequence selected from the group consisting of SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87 and SEQ ID NO:88.

In yet a further specific example, the peptide tag described herein comprises SEQ ID NO:1 or SEQ ID NO:2 directly linked to the N-terminus of an amino acid sequence selected from the group consisting of SEQ ID NO:160 and SEQ ID: 140.

Further provided herein are nucleic acid sequences encoding a peptide tag of the invention.

Specifically, provided herein is a nucleic acid sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:32, SEQ ID NO:35, SEQ ID NO:42 and SEQ ID NO:47, or a functionally active variant thereof comprising an amino acid sequence having at least 95% sequence identity to SEQ ID NO:10, SEQ ID NO:21, , SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87 or SEQ ID NO:88 and comprising SEQ ID NO:1 or SEQ ID NO:2, preferably at its N-terminus.

Specifically, provided herein is a nucleic acid sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NO:150 and SEQ ID NO:155, or a functionally active variant thereof comprising an amino acid sequence having at least 95% sequence identity to SEQ ID NO:160 and comprising SEQ ID NO:1 or SEQ ID NO:2, preferably at its N-terminus.

Specifically, provided herein is a nucleic acid sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:32, SEQ ID NO:35, SEQ ID NO:42 and SEQ ID NO:47, or a functionally active variant thereof comprising SEQ ID NO:1 or SEQ ID NO:2 and comprising at least 95% sequence

identity to SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:32, SEQ ID NO:35, SEQ ID NO:42 or SEQ ID NO:47.

Specifically, provided herein is a nucleic acid sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NO:150 and SEQ ID NO:155, or a functionally active variant thereof comprising SEQ ID NO:1 or SEQ ID NO:2 and comprising at least 95% sequence identity to SEQ ID NO:150 or SEQ ID NO:155.

In a specific example, the peptide tag comprises the amino acid sequence SEQ ID NO:11 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16.

In a further specific example, the peptide tag comprises the amino acid sequence SEQ ID NO:12 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 and SEQ ID NO:20.

In a further specific example, the peptide tag comprises the amino acid sequence SEQ ID NO:22 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26 and SEQ ID NO:27.

In a further specific example, the peptide tag comprises the amino acid sequence SEQ ID NO:23 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30 and SEQ ID NO:31.

In a further specific example, the peptide tag comprises the amino acid sequence SEQ ID NO:32 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:38, and SEQ ID NO:39.

In a further specific example, the peptide tag comprises the amino acid sequence SEQ ID NO:35 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:40, and SEQ ID NO:41.

In a further specific example, the peptide tag comprises the amino acid sequence SEQ ID NO:42 and is encoded by a nucleotide sequence selected from the

group consisting of SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, and SEQ ID NO:46.

In a further specific example, the peptide tag comprises the amino acid sequence SEQ ID NO:47 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, and SEQ ID NO:51.

In a further specific example, the peptide tag comprises the amino acid sequence SEQ ID NO:150 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, and SEQ ID NO:154.

In a further specific example, the peptide tag comprises the amino acid sequence SEQ ID NO:155 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:158, and SEQ ID NO:159.

Specifically described herein is an expression vector for expression of a recombinant POI as a fusion protein comprising any of the peptide tags of the invention fused to said POI, preferably fused directly to the N- or C-terminus of said POI, wherein said expression vector comprises a nucleic acid sequence encoding said peptide tag and a nucleic acid sequence encoding said POI.

According to a specific example, the expression vector described herein comprises a nucleic acid sequence encoding any one of the peptide tags described herein, and said nucleic acid sequence is selected from the group consisting of SEQ ID NO:13-20, SEQ ID NO:24-31, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:36-41, SEQ ID NO:43-46 and SEQ ID NO:48-51.

According to a specific example, the expression vector described herein comprises a nucleic acid sequence encoding any one of the peptide tags described herein, and said nucleic acid sequence is selected from the group consisting of, SEQ ID NO:151-154 and SEQ ID NO:156-159.

Specifically described herein is an expression vector for expression of a recombinant protein of interest (POI) as a fusion protein comprising any of the peptide tags of the invention, wherein the one terminus of the POI is the N-terminus.

Specifically, the expression vector described herein comprises:

- a. a promoter;

- b. a ribosome binding site (RBS);
- c. a start codon;
- d. the sequence encoding any one of the peptide tags described herein, preferably wherein said sequence encoding the peptide tag is placed immediately after the start codon; and
- e. the sequence encoding the POI, preferably wherein said sequence encoding the POI follows immediately after the 3' end of the sequence encoding the peptide tag.

Further specifically described herein is an expression vector for expression of a recombinant protein of interest (POI) as a fusion protein, comprising:

- a. a promoter;
- b. a ribosome binding site (RBS);
- c. a start codon;
- d. a nucleic acid sequence encoding any one of the peptide tags described herein, preferably wherein said sequence encoding the peptide tag is placed immediately after the start codon; and
- e. a cloning site, preferably placed immediately downstream of the sequence encoding the peptide tag.

Further provided herein is an expression cassette for expression of a recombinant protein of interest (POI) as a fusion protein comprising the peptide tag of the invention fused to one terminus of said POI, wherein said expression cassette comprises a nucleic acid sequence encoding said peptide tag of the invention and a nucleic acid sequence encoding said POI.

According to a specific example, the expression cassette described herein comprises a nucleic acid sequence encoding the peptide tag of the invention, which nucleic acid sequence is selected from the group consisting of SEQ ID NO:13-20, SEQ ID NO:24-31, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:36-41, SEQ ID NO:43-46 and SEQ ID NO:48-51.

According to a specific example, the expression cassette described herein comprises a nucleic acid sequence encoding the peptide tag of the invention, which nucleic acid sequence is selected from the group consisting of SEQ ID NO:151-154 and SEQ ID NO:156-159.

Specifically described herein is an expression cassette for expression of a recombinant protein of interest (POI) as a fusion protein comprising any of the peptide tags of the invention, wherein the one terminus of the POI is the N-terminus.

Specifically, the expression cassette described herein comprises:

- 5 a. a promoter;
- b. a ribosome binding site (RBS);
- c. a start codon;
- d. a nucleic acid sequence encoding the peptide tag of the invention,
preferably wherein said sequence encoding the peptide tag is
10 placed immediately after the start codon; and
- e. a nucleic acid sequence encoding the POI, preferably wherein
said sequence encoding the POI is placed immediately after the
sequence encoding the peptide tag.

Further specifically described herein is an expression cassette for expression of
15 a recombinant protein of interest (POI) as a fusion protein, comprising:

- a. a promoter;
- b. a ribosome binding site (RBS);
- c. a start codon;
- d. a nucleic acid sequence encoding the peptide tag of the invention,
20 preferably wherein said sequence encoding the peptide tag is
placed immediately after the start codon; and
- e. a cloning site, preferably placed immediately downstream of the
sequence encoding the peptide tag.

Further provided herein is an expression vector comprising the expression cas-
25 sette described herein.

Further specifically provided herein is a recombinant protein to which any one of the peptide tags described herein is linked, preferably at the N-terminus.

Specifically, a fusion protein is provided herein comprising a recombinant POI and the peptide tag of the invention, wherein the peptide tag of the invention is fused
30 to the recombinant POI, preferably at the N-terminus of the POI.

Further provided herein are nucleic acid sequences encoding a fusion protein described herein. Preferably, said nucleic acid sequence encoding a fusion protein

as described herein comprises a start codon. Specifically, said start codon encodes a methionine.

Further specifically described herein is a method of producing a recombinant protein of interest (POI) comprising expressing the POI in the form of a fusion protein comprising any one of the peptide tags described herein fused to one terminus of said POI, by a bacterial, mammalian or a yeast expression system.

Specifically, wherein the bacterial expression system is an *E. coli* expression system, for example using a T7 promoter or an A1 promoter or a functionally active variant thereof.

Specifically, wherein the yeast expression system is an expression system using yeast cells of the genus *Komagataella* or *Saccharomyces*.

Specifically, wherein the mammalian expression system is an expression system using CHO cells.

Specifically, in the method of producing a recombinant protein of interest (POI) comprising expressing the POI in the form of a fusion protein described herein, said fusion protein is expressed from an expression vector transiently inserted into a host cell or from an expression cassette stably integrated into the genome of a host cell.

Specifically, in the method of producing a recombinant protein of interest (POI) comprising expressing the POI in the form of a fusion protein described herein, said fusion protein is expressed from an expression vector, preferably a plasmid, within a host cell or from an expression cassette stably integrated into the genome or chromosome of a host cell.

According to a further specific example, the method of producing a recombinant POI described herein comprises culturing the host cell for a period of time under conditions permitting expression of the POI in the form of a fusion protein and comprising any one or more of the steps of:

- a. recovering the fusion protein and/or POI;
- b. purifying the fusion protein and/or POI;
- c. removing the peptide tag;
- d. further purifying the fusion protein and/or POI;
- e. formulating the fusion protein and/or POI;
- f. modifying the fusion protein and/or POI.

Further provided herein is a host cell comprising the peptide tag of the invention and/or a nucleic acid sequence encoding said peptide tag.

Further provided herein is a host cell comprising a fusion protein comprising the peptide tag of the invention and a protein of interest (POI) and/or a nucleic acid
5 sequence encoding said fusion protein.

Further provided herein is a host cell comprising the expression vector described herein and/or the expression cassette described herein.

BRIEF SUMMARY OF THE DRAWINGS

10 **Figure 1:** Amino acid and nucleotide sequences referred to herein.

Figure 2: Schematic overview of tested constructs. N-terminal variants of the tags and their minimal free folding energies of nucleotides -4 to +37.

Figure 3: Recombinant protein production kinetics for the PTH fusion proteins (A and B) and the hFGF2 fusion proteins (D and E) and biomass formation for the
15 PTH fusion proteins (C) and the hFGF2 fusion proteins (F) over the time course of the fermentations. For amino acid sequences and nucleotide sequences of the fusion constructs see Table 4.

Figure 4: Direct comparison of specific end-of-fermentation titres for all tested constructs. For amino acid sequences and nucleotide sequences of the fusion con-
20 structs see Table 4.

Figure 5: Consensus Sequence

Figure 6: Soluble product formation kinetics for TNF α fusion proteins: Volumetric titer [g/L] of fermentations of TNF α fusion proteins comprising LED, PLV or PHS in the T7AC – tag.

25 **Figure 7:** Soluble product formation kinetics for hFGF-2 fusion proteins: Volumetric titer [g/L] of fermentations of hFGF-2 fusion proteins comprising LED, PLV or PHS N-terminal to the PERNKERKE – tag.

Figure 8: Soluble product formation kinetics for PTH fusion proteins: Volumetric titer [g/L] of fermentations of PTH fusion proteins comprising LED, PLV or PHS
30 N-terminal to the PERNKERKE – tag.

Figure 9: Additional Amino acid and nucleotide sequences referred to herein.

DETAILED DESCRIPTION OF THE INVENTION

A primary objective of the present invention is to enhance expression, solubility and/or proper folding of recombinant proteins of interest expressed in host cells. Accordingly, the present invention enables the production of biologically active non-
5 prokaryotic and prokaryotic proteins within host cells in sufficient quantities, e.g. for industrial and/or medical applications. As demonstrated in the Examples provided herein, prokaryotic cells (e.g. *E. coli*) represent an important example of host cells, however, low titres, solubility problems and the formation of inclusion bodies are well-known issues faced in eukaryotic host cells as well.

10 The present invention is based on the surprising finding that a peptide tag comprising the amino acid sequence SEQ ID NO:1 or SEQ ID NO:2 significantly enhance expression of target proteins fused to such peptide tag.

Use of the novel peptide tags described herein significantly increased expression of several pharmaceutically relevant proteins as is shown in the Examples provided herein. The present invention thus provides means to efficiently improve existing strategies of recombinant protein production and to obtain a target protein at
15 a high yield, such that active protein required for biotechnology and pharmaceuticals may be supplied in mass-production.

20 ***Peptide tags of the invention***

Provided herein is a peptide tag comprising a first peptide sequence consisting of an amino acid sequence selected from SEQ ID NO:1 or SEQ ID NO:2, which peptide tag is useful for expression of a recombinant protein of interest (POI) in a host cell. Specifically, the peptide tag described herein increases expression of at
25 least one POI.

Further provided herein are nucleic acid sequences encoding the peptide tag of the invention. For example, the nucleic acid sequence encoding a peptide tag of the invention may comprise or consist of SEQ ID NO:3 or SEQ ID NO:4.

The peptide tag of the invention may be longer than the first peptide sequence
30 of SEQ ID NO:1 or SEQ ID NO:2. Specifically, the peptide tag comprises at least about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 amino acids. Specifically, the peptide tag comprises at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,

21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 or more amino acids.

Specifically, such additional amino acids include, but are not limited to, tag elements, linkers, protease recognition sites and/or protease cleavage sites as described herein.

In a specific embodiment, the peptide tag comprises an amino acid elongation C-terminal to the first peptide sequence, so that the first peptide sequence is located at the N-terminus of the peptide tag, preferably only preceded by a starting methionine. Specifically, the peptide tag comprises additional amino acids at the C-terminus of the first peptide sequence. Specifically, the peptide tag comprises at least about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 amino acids in addition to the first peptide sequence. Specifically, the peptide tag comprises at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 or more amino acids in addition to the first peptide sequence.

For example, the peptide tag may comprise a second peptide sequence comprising SEQ ID NO:5, SEQ ID NO:10 or SEQ ID NO:21, or functionally active variants thereof.

In another example, the peptide tag may comprise a second peptide sequence comprising SEQ ID NO:107 or functionally active variants thereof.

Specifically, the first peptide sequence and the second peptide sequence are directly linked or they are connected via a linker as further described herein. Preferably, the first peptide sequence and the second peptide sequence are directly linked.

Specifically, the first peptide sequence is N-terminal of the second peptide sequence.

In a specific example, the peptide tag of the invention comprises a first peptide sequence of SEQ ID NO:1 directly linked to a second peptide sequence of SEQ ID NO:10. The peptide tag of the invention thus may comprise or consist of the amino acid sequence of SEQ ID NO:11. Specifically, encoded by a nucleic acid sequence comprising or consisting of SEQ ID NO:13.

In another specific example, the peptide tag of the invention comprises a first peptide sequence of SEQ ID NO:1 directly linked to a second peptide sequence of

SEQ ID NO:21. The peptide tag of the invention thus may comprise or consist of the amino acid sequence of SEQ ID NO:22. Specifically, encoded by a nucleic acid sequence comprising or consisting of SEQ ID NO:24.

5 In a specific example, the peptide tag of the invention comprises a first peptide sequence of SEQ ID NO:2 directly linked to a second peptide sequence of SEQ ID NO:10. The peptide tag of the invention thus may comprise or consist of the amino acid sequence of SEQ ID NO:12. Specifically, encoded by a nucleic acid sequence comprising or consisting of SEQ ID NO:17.

10 In a specific example, the peptide tag of the invention comprises a first peptide sequence of SEQ ID NO:2 directly linked to a second peptide sequence of SEQ ID NO:21. The peptide tag of the invention thus may comprise or consist of the amino acid sequence of SEQ ID NO:23. Specifically, encoded by a nucleic acid sequence comprising or consisting of SEQ ID NO:28.

15 In a specific example, the peptide tag of the invention comprises a first peptide sequence of SEQ ID NO:1 directly linked to a second peptide sequence of SEQ ID NO:107. The peptide tag of the invention thus may comprise or consist of the amino acid sequence of SEQ ID NO:111. Specifically, encoded by a nucleic acid sequence comprising or consisting of SEQ ID NO:112.

20 In a specific example, the peptide tag of the invention comprises a first peptide sequence of SEQ ID NO:2 directly linked to a second peptide sequence of SEQ ID NO:107. The peptide tag of the invention thus may comprise or consist of the amino acid sequence of SEQ ID NO:113. Specifically, encoded by a nucleic acid sequence comprising or consisting of SEQ ID NO:114.

25 In a specific example, the peptide tag of the invention comprises a first peptide sequence of SEQ ID NO:1 directly linked to a second peptide sequence of SEQ ID NO:140.

In a specific example, the peptide tag of the invention comprises a first peptide sequence of SEQ ID NO:2 directly linked to a second peptide sequence of SEQ ID NO:140.

30 Specifically, the nucleic acid sequence encoding the peptide tag described herein may comprise one or more silent mutation(s), preferably including the introduction of rare codons. Specifically, the nucleic acid sequence encoding the peptide tag described herein comprises at least 1, 2, 3, 4, 5, 6, 7, or 8 codons that have

been modified to introduce silent mutations. Specifically, such modifications may be achieved by single or multiple nucleotide substitutions, deletions and/or insertions.

Preferably, the nucleic acid sequence encoding the peptide tag described herein comprises at least 1 rare codon. Preferably, more than one of the silent mutations
5 result in rare codons. Rare codons typically are codons with a codon usage of less than 1% in the host cell, or less than 10 in 1000 codons. Specifically, the nucleic acid sequence encoding the peptide tag described herein comprises at least 2, 3, 4, or 5 rare codons.

In a specific example, the peptide tag described herein comprises or consists of
10 SEQ ID NO:11 and is encoded by a nucleic acid sequence comprising SEQ ID NO:13 or by a nucleic acid sequence comprising SEQ ID NO:13 comprising at least 1, 2, 3, 4 or 5 silent mutations. Specific examples of such nucleic acid sequence include but are not limited to SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16.

In a specific example, the peptide tag described herein comprises or consists of
15 SEQ ID NO:12 and is encoded by a nucleic acid sequence comprising SEQ ID NO:17 or by a nucleic acid sequence comprising SEQ ID NO:17 comprising at least 1, 2, 3, 4 or 5 silent mutations. Specific examples of such nucleic acid sequence include but are not limited to SEQ ID NO:18, SEQ ID NO:19 and SEQ ID NO:20.

In a specific example, the peptide tag described herein comprises or consists of
20 SEQ ID NO:22 and is encoded by a nucleic acid sequence comprising SEQ ID NO:24 or by a nucleic acid sequence comprising SEQ ID NO:22 comprising at least 1, 2, 3, 4 or 5 silent mutations. Specific examples of such nucleic acid sequence include but are not limited to SEQ ID NO:25, SEQ ID NO:26 and SEQ ID NO:27.

In a specific example, the peptide tag described herein comprises or consists of
25 SEQ ID NO:23 and is encoded by a nucleic acid sequence comprising SEQ ID NO:28 or by a nucleic acid sequence comprising SEQ ID NO:23 comprising at least 1, 2, 3, 4 or 5 silent mutations. Specific examples of such nucleic acid sequence include but are not limited to SEQ ID NO:29, SEQ ID NO:30 and SEQ ID NO:31.

In a specific example, the peptide tag described herein comprises or consists of
30 SEQ ID NO:111 and is encoded by a nucleic acid sequence comprising SEQ ID NO:112 or by a nucleic acid sequence comprising SEQ ID NO:112 comprising at least 1, 2, 3, 4 or 5 silent mutations. Specific examples of such nucleic acid

sequence include but are not limited to SEQ ID NO:144, SEQ ID NO:145 and SEQ ID NO:146.

In a specific example, the peptide tag described herein comprises or consists of SEQ ID NO:113 and is encoded by a nucleic acid sequence comprising SEQ ID
5 NO:114 or by a nucleic acid sequence comprising SEQ ID NO:114 comprising at least 1, 2, 3, 4 or 5 silent mutations. Specific examples of such nucleic acid sequence include but are not limited to SEQ ID NO:147, SEQ ID NO:148 and SEQ ID NO:149.

The peptide tag described herein may comprise the first peptide sequence
10 and/or the sequence peptide sequence more than once. Such multiples of the first and/or second peptide sequence may be adjacent to each other or may be separated by one or more linkers or tag elements.

Specifically, the fusion protein described herein may comprise multiple parts, and may comprise duplicates of such parts. For example, the fusion protein herein
15 may comprise a first part comprising a peptide tag as described herein, a second part comprising a protease recognition site, another first part comprising the same or different peptide tag, another second part comprising the same or different protease recognition site and a third part comprising a POI.

Specifically, the peptide tag described herein comprises a starting methionine.

20 Specifically, the peptide tag described herein comprises one or more linkers. Such linker may be placed between any individual part or tag element of the peptide tag.

Specifically, the peptide tag described herein comprises one or more further tag elements such as solubility enhancement tags, monitoring tags and/or affinity tags.

25 Specifically, the peptide tag described herein may comprise one or more tag elements more than once.

Specifically, the peptide tag described herein comprises at least one protease recognition site. Specifically, the peptide tag described herein comprises at least one protease cleavage site. Specifically, the peptide tag described herein comprises
30 at least one protease recognition and cleavage site. Specifically, the peptide tag described herein comprises a protease cleavage site at its C-terminus.

In a specific example, the peptide tag described herein additionally comprises C-terminal to the first or second peptide sequence, and preferably in that order:

- a. an affinity tag, preferably a His tag, even more preferably a 6-His-tag;
- b. a linker, preferably GSG, and
- c. a caspase-2 recognition and cleavage site, preferably VDVAD.

Specifically, the peptide tag described herein comprises or consists of SEQ ID NO:42. Specifically, the peptide tag of SEQ ID NO:42 is encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45 and SEQ ID NO:46.

Specifically, the peptide tag described herein comprises or consists of SEQ ID NO:47. Specifically, the peptide tag of SEQ ID NO:47 is encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50 and SEQ ID NO:51.

Specifically, the peptide tag described herein comprises or consists of SEQ ID NO:150. Specifically, the peptide tag of SEQ ID NO:150 is encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153 and SEQ ID NO:154.

Specifically, the peptide tag described herein comprises or consists of SEQ ID NO:155. Specifically, the peptide tag of SEQ ID NO:155 is encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:158 and SEQ ID NO:159.

In another specific example, the peptide tag described herein additionally comprises C-terminal to the first or second peptide sequence, and preferably in the following order,

- a. a first affinity tag, preferably a His tag, even more preferably a 6-His-tag;
- b. a linker, preferably an SA linker;
- c. a second affinity tag different from the first affinity tag, preferably a Strep tag, even more preferably a Strep II tag;
- d. a linker, preferably a GSG linker, and
- e. a caspase-2 recognition and cleavage site, preferably VDVAD.

Specifically, the peptide tag described herein comprises or consists of SEQ ID NO:32. Specifically, the peptide tag of SEQ ID NO:32 is encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:38 and SEQ ID NO:39.

Specifically, the peptide tag described herein comprises or consists of SEQ ID NO:35. Specifically, the peptide tag of SEQ ID NO:35 is encoded by a nucleic acid sequence selected from the group consisting of SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:40 and SEQ ID NO:41.

5

Used Terms and Definitions

Unless indicated or defined otherwise, all terms used herein have their usual meaning in the art, which will be clear to the skilled person. Reference is for example made to standard handbooks, such as Sambrook et al, "Molecular Cloning: A Laboratory Manual" (2nd Ed.), Vols. 1–3, Cold Spring Harbor Laboratory Press (1989);
10 Lewin, "Genes IV", Oxford University Press, New York, (1990), and Janeway et al, "Immunobiology" (5th Ed., or more recent editions, Garland Science, New York, 2001).

The subject matter of the claims specifically refers to artificial products or methods employing or producing such artificial products, which may be variants of native
15 (wild-type) products. Though there can be a certain degree of sequence identity to the native structure, it is well understood that the materials, methods and uses of the invention, e.g., specifically referring to isolated nucleic acid sequences, amino acid sequences, fusion constructs, expression constructs, transformed host cells
20 and modified proteins, are "man-made" or synthetic, and are therefore not considered as a result of "laws of nature".

It must be noted that as used herein, the singular forms "a", "an" and "the" include plural references and vice versa unless the context clearly indicates otherwise. Thus, for example, a reference to "a host cell" or "a method" includes one or
25 more of such host cells or methods, respectively, and a reference to "the method" includes equivalent steps and methods that could be modified or substituted known to those of ordinary skill in the art. Similarly, for example, a reference to "methods" or "host cells" includes "a host cell" or "a method", respectively.

Unless otherwise indicated, the term "at least" preceding a series of elements
30 is to be understood to refer to every element in the series. Those skilled in the art will recognize or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the present invention.

The term "and/or" wherever used herein includes the meaning of "and", "or" and "all or any other combination of the elements connected by said term". For example, A, B and/or C means A, B, C, A+B, A+C, B+C and A+B+C.

The term "about" or "approximately" as used herein means within 20%, preferably within 10%, and more preferably within 5% of a given value or range. It includes also the concrete number, e.g., about 20 includes 20.

The term "less than", "more than" or "larger than" includes the concrete number. For example, less than 20 means ≤ 20 and more than 20 means ≥ 20 .

Throughout this specification and the claims, unless the context requires otherwise, the word "comprise" and variations such as "comprises" and "comprising" will be understood to imply the inclusion of a stated integer (or step) or group of integers (or steps). It does not exclude any other integer (or step) or group of integers (or steps). When used herein, the term "comprising" can be substituted with "containing", "composed of", "including", "having" or "carrying" and vice versa, by way of example the term "having" can be substituted with the term "comprising". When used herein, "consisting of" excludes any integer or step not specified in the claim/item.

It should be understood that this invention is not limited to the particular methodology, protocols, material, reagents, and substances, *etc.*, described herein. The terminologies used herein are for the purpose of describing particular embodiments only and are not intended to limit the scope of the present invention, which is defined solely by the claims.

All publications and patents cited throughout the text of this specification (including all patents, patent applications, scientific publications, manufacturer's specifications, instructions, *etc.*), whether *supra* or *infra*, are hereby incorporated by reference in their entirety. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention. To the extent the material incorporated by reference contradicts or is inconsistent with this specification, the specification will supersede any such material.

The term "nucleic acid" refers to deoxyribonucleic acid (DNA, *e.g.* a cDNA or genomic DNA), ribonucleic acid (RNA, *e.g.* a mRNA), or a DNA or RNA analog and polymers thereof, in either single- or double-stranded form, but preferably is double-stranded DNA, made of monomers (nucleotides) containing a sugar, phosphate and a base that is either a purine or pyrimidine.

A "polynucleotide" as used herein, refers to nucleotides, either ribonucleotides (RNA, e.g., an mRNA) or deoxyribonucleotides (DNA, e.g., a cDNA) or a combination of both, in a polymeric unbranched form of any length. Preferably, a polynucleotide refers to deoxyribonucleotides in a polymeric unbranched form of any length.

5 Here, nucleotides consist of a pentose sugar (deoxyribose), a nitrogenous base (adenine, guanine, cytosine or thymine/uracil) and a phosphate group. The terms "polynucleotide(s)", "nucleic acid sequence(s)" are used interchangeably herein. Unless otherwise indicated, a particular nucleic acid sequence also encompasses complementary sequences.

10 The term "isolated," as used herein, refers to material that is removed from its original or native environment (e.g. the natural environment if it is naturally occurring). For example, a naturally-occurring nucleic acid molecule or polypeptide present in a living animal is not isolated, but the same nucleic acid molecule or polypeptide, separated by human intervention from some or all of the co-existing materials in the natural system, is isolated. Such nucleic acid molecules could be part of
15 a vector and/or such nucleic acid molecules or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of the environment in which the nucleic acid molecule or the polypeptide is found in nature. For example, a nucleic acid molecule or polypeptide is considered to be "(in) isolated
20 (form)" when, compared to its native biological source and/or the reaction medium or cultivation medium from which it has been obtained, it has been separated from at least one other component with which it is usually associated in said source or medium, such as another nucleic acid molecule, another polypeptide, another biological component or macromolecule or at least one contaminant, impurity or minor
25 component.

Amino acid residues will be indicated according to the standard three-letter or one-letter amino acid code, as generally known and agreed upon in the art. There are twenty known naturally occurring amino acids encoded by sixty-one triplet codons. These 20 amino acids can be split into those that have neutral charges, positive charges, and negative charges:
30

The "neutral" amino acids are shown below along with their respective three-letter and single-letter code and polarity:

Alanine: (Ala, A) nonpolar, neutral;

Asparagine: (Asn, N) polar, neutral;
Cysteine: (Cys, C) polar, neutral;
Glutamine: (Gln, Q) polar, neutral;
Glycine: (Gly, G) nonpolar, neutral;
5 Isoleucine: (Ile, I) nonpolar, neutral;
Leucine: (Leu, L) nonpolar, neutral;
Methionine: (Met, M) nonpolar, neutral;
Phenylalanine: (Phe, F) nonpolar, neutral;
Proline: (Pro, P) nonpolar, neutral;
10 Serine: (Ser, S) polar, neutral;
Threonine: (Thr, T) polar, neutral;
Tryptophan: (Trp, W) nonpolar, neutral;
Tyrosine: (Tyr, Y) polar, neutral;
Valine: (Val, V) nonpolar, neutral; and
15 Histidine: (His, H) polar, positive (10%) neutral (90%).
The "positively" charged amino acids are:
Arginine: (Arg, R) polar, positive; and
Lysine: (Lys, K) polar, positive.
The "negatively" charged amino acids are:
20 Aspartic acid: (Asp, D) polar, negative; and
Glutamic acid: (Glu, E) polar, negative.

The terms "peptide", "polypeptide" and "protein" are used herein interchangeably and refer to the arrangement of amino acid residues in a polymer (polypeptide chain). A peptide, polypeptide or protein can be composed of the standard 20 naturally occurring amino acids, and may include rare amino acids and synthetic amino acid analogs (e.g. non-canonical amino acids). A protein can be a monomer, a dimer
25 or a higher -mer comprising one, two or more polypeptide chains. The protein can be a homo-di- or higher -mer comprising two or more identical polypeptide chains or a hetero-di- or higher -mer comprising two or more different polypeptide chains or a mixture thereof, such as a heteromer of a homomer or a homomer of a heteromer.
30 The protein and the polypeptide chains can be any chain of amino acids, regardless of length or post-translational modification (e.g. glycosylation or phosphorylation). The peptide, polypeptide, protein or polypeptide chain(s) described herein includes

recombinantly or synthetically produced peptide tags described herein, either individually or as part of a fusion with e.g. a POI.

The genetic code translates mRNA nucleotide sequences to amino acid sequences. Genetic information is coded using this process with groups of three nucleotides along the mRNA which are commonly known as "codons". The set of three nucleotides almost always produce the same amino acid, with a few exceptions like UGA which typically serves as the stop codon but can also encode tryptophan in mammalian mitochondria. Most amino acids are specified by multiple codons demonstrating that the genetic code is degenerate—different codons result in the same amino acid. Codons that code for the same amino acid are termed synonyms or synonymous codons. Translation is initiated at the start codon, which typically is AUG. The start codon is the first codon of an mRNA transcript translated by a ribosome and typically codes for a methionine. The start codon is preferably preceded by a 5' untranslated region (5' UTR), which in prokaryotes preferably includes the ribosome binding site.

A point mutation is particularly understood as the engineering of a polynucleotide that results in the expression of an amino acid sequence that differs from the non-engineered amino acid sequence in the substitution or exchange, deletion or insertion of one or more single (non-consecutive) or doublets of amino acids for different amino acids.

As used herein, the term "silent mutation(s)" refers to base substitutions that result in no change of the amino acid or amino acid functionality when the altered messenger RNA (mRNA) is translated. Specifically, it refers to a substitution of at least one nucleotide in a codon that does not result in a change of the respective amino acid. For example, if the codon AAA is altered to become AAG, the same amino acid – lysine - will be incorporated into the peptide chain.

The term "recombinant" as used herein shall mean "being prepared by or the result of genetic engineering". A "recombinant cell" or "recombinant host cell" refers to a cell or host cell that has been genetically altered to comprise a nucleic acid sequence which was not native to said cell. A recombinant host specifically comprises a recombinant expression vector or cloning vector, or it has been genetically engineered to contain a recombinant nucleic acid sequence in its genome, such as its chromosome.

As used herein, the term "homologous" means derived from the same cell or organism with the same genomic background. As used herein, the term "heterologous" means derived from a cell or organism with a different genomic background, or an artificial sequence not found in nature. Thus, a heterologous nucleotide sequence, e.g. a DNA, can be a DNA encoding an artificial protein not found in nature.

As used herein the term "endogenous" means originating inside the organism or cell. Specifically, use of an endogenous DNA, e.g. for expression of a protein, means that the DNA originating inside the organism or cell is used for expression. As used herein the term "exogenous" means originating outside the organism or cell. Specifically, use of an exogenous sequence, such as DNA, for expression of a protein, means that the sequence is introduced into the cell, where expression from this sequence takes place.

A heterologous or a homologous gene for expression of a heterologous or homologous protein can be introduced into the cell by integrating a vector carrying the gene (there are one or more copies of the vector and thus one or more copies of the gene in the host cell) and being expressed from the vector or one or more copies of the gene can be integrated into the genome and/or chromosome of the host cell, from where the gene is expressed. A homologous protein can also be expressed or overexpressed in a host cell from an endogenous gene. In this case, the tag sequence described herein and optionally a promoter, is integrated into the genome and/or chromosome of the host cell such that it is operably linked to the endogenous gene.

A recombinant protein also may be a homologous protein. In this case, one or more copies of the polynucleotide encoding the homologous protein are introduced into the host cell by genetic manipulation.

Specifically, heterologous nucleotide sequences are those not found in the same relationship to a host cell in nature (i.e., "not natively associated"). Any recombinant or artificial nucleotide sequence is understood to be heterologous.

The term "increasing the expression of at least one protein of interest" or "improving the expression of at least one protein of interest" means that the yield and/or the titer and/or secretion and/or the soluble yield and/or soluble titer of a POI which is expressed by a host cell is increased, when said POI is expressed from a fusion construct comprising a nucleotide sequence encoding a peptide tag of the invention

compared to expression by the same host cell but from an expression construct comprising a nucleotide sequence encoding the POI without a peptide tag or encoding the POI in fusion to a peptide tag but wherein said peptide tag does not comprise SEQ ID NO:1 or SEQ ID NO:2. Specifically, the peptide tags of the invention improve
5 the expression of at least one POI, such as e.g. PTH or hFGF2, as compared to T7AC (SEQ ID NO:79) and/or as compared to T7A3 (SEQ ID NO:78). Specifically, expression is increased at least by about 1.1 fold, 1.2 fold, 1.3 fold, 1.4 fold, 1.5 fold, 1.6 fold, 1.7 fold, 1.8 fold, 1.9 fold, 2 fold, 2.1 fold, 2.2 fold, 2.3 fold, 2.4 fold, 2.5 fold, 2.6 fold, 2.7 fold, 2.8 fold, 2.9 fold, 3 fold, 3.1 fold, 3.2 fold, 3.3 fold, 3.4 fold, 3.5 fold,
10 3.6 fold, 3.7 fold, 3.8 fold, 3.9 fold, 4 fold or even about 5 fold.

The term “yield”, synonymously used herein with the terms “specific titer” or “spec. titer” or “POI [mg/g CDM]”, refers to the amount of POI(s) as described herein per cell or biomass or product per biomass in mg/g biomass measured as dry cell weight (CDM) and may be presented by mg POI/g biomass (biomass being meas-
15 ured as dry cell weight or cell dry mass (CDM) or wet cell weight (WCW), preferably measured as CDM so that “yield” is represented by mg POI/g (CDM) of a host cell.

As used herein, the term “titer” or “titre”, synonymously used herein with the terms “volumetric titer” or “vol. titer”, refers similarly to the amount of produced POI or model protein(s) as described herein per volume and may be presented as prod-
20 uct per volume in mg/L or mg POI/L culture supernatant or whole cell broth.

The expressed POI(s) can be located within the cytosol and/or periplasm (in soluble form or in insoluble form, specifically as Inclusion Bodies (“IB”)) of the host cell and/or the supernatant of the cell culture or cell broth. Thus, the terms “titer” (volumetric titer) and “yield” (specific titer) can be used to refer to “soluble titer” /
25 “soluble volumetric titer” or “soluble yield” / “soluble specific titer” for the part of the POI or fusion protein which, is expressed solubly and “total titer” / “total volumetric titer” or “total yield” / “total specific titer” for the whole amount of the POI or fusion protein including soluble and insoluble fractions, and “IB titer” / “IB volumetric titer” or “IB yield” / “IB specific titer” for the part of the POI or fusion protein which is in the
30 form of IBs.

With regard to the protein or polypeptide of interest (POI) there are no limitations. More specifically, the protein may either be a polypeptide not naturally occurring in the host cell, *i.e.* a heterologous protein. The POI may further be a

homologous protein to the host cell, but is produced, for example, upon integration by recombinant techniques of one or more copies of the nucleic acid sequence encoding the homologous POI into a host cell. The POI may be produced upon integration of the nucleic acid sequence encoding the POI into the genome or chromosome of the host cell. According to a further example, the POI can also be expressed in a host using a vector, more specifically a plasmid. Thus, the POI can be expressed from a plasmid or from the chromosome or genome of the host cell. In case the POI is homologous to the host cell, a nucleic acid sequence comprising the tag sequence of the invention, and, optionally, a promoter different to the native promoter of the endogenous gene, may be integrated into the host cell so that it is operably linked to the endogenous gene encoding the homologous POI.

The POI can be a monomer, dimer or multimer, it can be a homomer or heteromer or a mixture thereof. The POI can be any peptide, protein or polypeptide, whether naturally occurring or not. The POI can also be a synthetic peptide, protein or polypeptide. The POI may also comprise one or more rare amino acids and synthetic amino acid analogs (e.g. non-canonical amino acids). Examples for proteins that can be produced by the method of the invention are, without limitation, enzymes, regulatory proteins, receptors, growth factors, hormones, peptides, e.g. peptide hormones, cytokines, membrane or transport proteins. The POIs may also be antigens as used for vaccination, vaccines, antigen-binding proteins, immune stimulatory proteins, interleukins, interferons, allergens, full-length antibodies or antibody fragments or derivatives thereof or affinity scaffolds. Antibody derivatives may be for example, but not limited to single chain variable fragments (scFv), Fab fragments or single domain antibodies or camelid antibodies or heavy chain antibodies or derivatives thereof such as V_{HH} fragments or the like. The POI can be an artificial protein comprising polypeptide chains comprising natural as well as artificial amino acid sequences and/or comprising non-canonical amino acids. The POI can also exclusively comprise artificial sequences or can comprise more than one artificial polypeptide chains.

The DNA molecule encoding the protein of interest is also termed "gene of interest" or "GOI". The gene of interest encoding the POI can be a naturally existing DNA sequence or a non-natural DNA sequence. One or more GOI(s) can be under the control of one promoter. Alternatively, each gene of interest is under the control

of its own promoter. If the POI is a heteromer, all different polypeptide chains may be under the control of the same promoter, or each polypeptide chain may be under the control of its own promoter or a mixture thereof. The genes of interest resp. the genes encoding the polypeptide chains of a heteromer may all be on the same expression cassette or on multiple expression cassettes. The POI can be modified in any way. Non-limiting examples for modifications can be insertion or deletion of post-translational modification sites, insertion or deletion of targeting signals (e.g.: leader peptides), fusion to tags (e.g. improving expression and/or solubility of the POI, affinity tags such as 6His-tag for IMAC purification, detection tags such as GFP and/or any tag having other functions), protease recognition and/or cleavage sites, proteins or protein fragments facilitating purification or detection, mutations affecting changes in stability or changes in solubility, or any other modification known in the art. In certain embodiments of the invention the recombinant protein is a biopharmaceutical product, which can be any protein suitable for therapeutic or prophylactic purposes in mammals.

As used herein the term "host cell" refers to a cell which is capable of protein expression and optionally protein secretion. Such host cell is applied in the methods of the present invention. For that purpose, for the host cell to produce a recombinant POI, a nucleotide sequence encoding a peptide tag as described herein is introduced or present in the cell. Typically, the term refers to viable cells, capable of growing in a cell culture and/or capable of expressing genes such as the GOI. Common cells that serve as hosts for expression of recombinant genes are prokaryotic and eukaryotic cells, such as bacteria, yeast, or mammalian cells. Examples of eukaryotic cells include, but are not limited to, vertebrate cells, mammalian cells, human cells, animal cells, invertebrate cells, plant cells, nematodal cells, insect cells, stem cells, fungal cells, or yeast cells.

Examples of bacterial cells include, but are not limited to *Escherichia coli*, *Bacillus species*, *Corynebacterium species*, *Pseudomonas species*, *Salmonella species* and *Streptomyces species*. Specific examples of *E. coli* strains include but are not limited to B strains and K strains, such as BL21, BL21 (DE3), HMS 174, HMS 174 (DE3).

The eukaryotic host cell may be a fungal cell. More preferred is a yeast host cell. Examples of yeast cells include but are not limited to the *Saccharomyces* genus

(e.g. *Saccharomyces cerevisiae*, *Saccharomyces kluyveri*, *Saccharomyces uvarum*), the *Komagataella* genus (*Komagataella pastoris*, *Komagataella pseudo-pastoris* or *Komagataella phaffii*), *Kluyveromyces* genus (e.g. *Kluyveromyces lactis*, *Kluyveromyces marxianus*), the *Candida* genus (e.g. *Candida utilis*, *Candida caoi*), as well as *Hansenula polymorpha*.

In a preferred embodiment, the genus *Pichia* is of particular interest. *Pichia* comprises a number of species, including the species *Pichia pastoris*, *Pichia methanolica*, *Pichia kluyveri*, and *Pichia angusta*. Most preferred is the species *Pichia pastoris*. The former species *Pichia pastoris* has been divided and renamed to *Komagataella pastoris*, *Komagataella phaffii* and *Komagataella pseudopastoris*. Therefore, *Pichia pastoris* is a synonymous for both *Komagataella pastoris*, *Komagataella phaffii* and *Komagataella pseudopastoris*.

Non-limiting examples of mammalian cells include human, mice, rat, monkey and rodent cells lines. Specific mammalian cell lines available as host cells for expression are well known in the art and include, inter alia, Chinese hamster ovary (CHO) cells, NS0, SP2/0 cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human carcinoma cells (e.g., Hep G2 and A-549 cells), 3T3 cells or the derivatives/progenies of any such cell line.

Appropriate culture media and conditions for the above-described host cells are known in the art.

Alternatively, the gene can be translated into protein using cell free translation systems, possibly coupled to an *in vitro* transcription system. These systems provide all steps necessary to obtain protein from DNA by supplying the necessary enzymes and substrates in an *in vitro* reaction. In principle, any living cell or organism can provide the necessary enzymes for this process and extraction protocols for obtaining such enzyme systems are known in the art. Common systems used for *in vitro* transcription/translation are extracts or lysates from reticulocytes, wheat germ or *Escherichia coli*.

The term "expression" or "expressing" is understood in the following way. Nucleic acid molecules containing a desired coding sequence of an expression product such as e.g., a peptide tag, a POI or a fusion protein as described herein, may be used for expression purposes. Hosts transformed or transfected with or containing these sequences are capable of producing the encoded proteins. In order to effect

transformation, the expression system may be included in a vector; however, the relevant DNA may also be integrated into the host chromosome or genome. Specifically, the term refers to a host cell and compatible vector under suitable conditions, *e.g.*, for the expression of a protein coded for by foreign DNA carried by the vector and introduced to the host cell.

A POI or a fusion protein can be expressed as soluble protein within the cytoplasm or as soluble protein within the periplasm (*e.g.* upon fusion of a leader or signal sequence to the POI or fusion protein). A POI or a fusion protein can also be expressed in soluble form and released to the culture supernatant by active transport of the POI or fusion protein from the cytosol, periplasm, or endoplasmic reticulum to the supernatant. For transport of the POI or fusion protein through the cell membrane(s) or membrane(s) of the involved cell organelles, a leader or signal sequence (specifically in bacterial expression systems) or a pre and/or pro sequence (specifically in yeast expression systems) may be fused to the N-terminus of the POI or fusion protein. The POI or fusion protein can also be released from the periplasm to the supernatant upon cell lysis or leakiness of the cell membrane. This may be achieved by addition of enzymes inducing lysis of the cells or by inducing expression of a cell lytic enzyme within the host cell during or after fermentation. Leakiness of the cell membrane(s) can be achieved by mutagenesis of the cell and selection of leaky cells, genetic engineering of cells to make them leaky or by inducing the overexpression and/or underexpression/knock-down of proteins of the cell leading to leakiness of the cell. A POI or fusion protein can also be expressed as inclusion bodies within the cytosol or, for example when an N-terminal signal or leader peptide is used, within the periplasm of the host cell.

The term "operably linked" as used herein refers to the association of nucleotide sequences on a single nucleic acid molecule, *i.e.* the vector, in a way such that the function of one or more nucleotide sequences is affected by at least one other nucleotide sequence present on said nucleic acid molecule. For example, a promoter is operably linked with a coding sequence encoding the protein of interest or a fusion protein comprising the POI, when it is capable of effecting the expression of that coding sequence. Specifically, such nucleic acids operably linked to each other may be immediately linked, *i.e.* directly linked without further elements or nucleic acid sequences in between or may be indirectly linked with spacer sequences or other

sequences in between. Specifically, in the context of a lac operator being operably linked to a promoter refers to the ability of the lac operator to regulate the ability of the promoter to control expression of the coding sequence under specific conditions. Such as the ability of the lac operator to inhibit promoter-dependent expression of the gene of interest when lac repressor protein is bound thereto.

The term "vector" as used herein includes autonomously replicating nucleotide sequences as well as genome integrating nucleotide sequences. A common type of vector is a "plasmid", which generally is a molecule of double-stranded DNA that can readily accept additional (foreign) DNA and which can readily be introduced into a suitable host cell. A plasmid vector often contains coding DNA and promoter DNA and has one or more restriction sites suitable for inserting foreign DNA. Specifically, the term "vector" or "plasmid" refers to a vehicle by which a DNA or RNA sequence (e.g., a foreign gene) can be introduced into a host cell, so as to transform the host and promote expression (e.g., transcription and translation) of the introduced sequence. The plasmid or expression vector can also be the 2 μ plasmid for use in yeasts from which the POI or fusion protein is expressed.

"Expression vectors" or "vectors" as used herein are defined as DNA sequences that are required for the transcription of cloned recombinant nucleotide sequences, i.e. of recombinant genes and the translation of their mRNA in a suitable host organism. To obtain expression, a sequence encoding a desired expression product, such as e.g. the fusion protein described herein, is typically cloned into an expression vector that contains a promoter to direct transcription. Suitable bacterial and eukaryotic promoters are well known in the art. The promoter used to direct expression of a nucleic acid depends on the particular application. For example, an inducible or regulatable promoter is typically used for expression and purification of proteins such as e.g. fusion proteins. Expression vectors comprise the expression cassette and additionally usually comprise an origin for autonomous replication in the host cells or a genome or chromosome integration site, one or more selectable markers (e.g., an amino acid synthesis gene or a gene conferring resistance to antibiotics such as zeocin, kanamycin, G418 or hygromycin), a number of restriction enzyme cleavage sites, a suitable promoter sequence, a cloning site or multiple cloning site and a transcription terminator, which components are operably linked together.

Examples for inducible promoters for use in bacteria, include but are not limited to the T7, A1 (also termed "T7A1" or "T7_{A1}") and T5 promoter systems. For example, being inducible in connection with e. g. one or more lacO repressor binding sites, the Pm promoter/operator or the pBAD promoter/operator. Examples for inducible
5 promoters for use in yeast, include the AOX 1 promoter or derivatives thereof, such as those published in WO2006/089329A1.

The promoter can also be a constitutive promoter, such as but not limited to the GAP or ADH1 promoter for yeast, or a promoter which is repressible by a metabolite or carbon source and is activated upon depletion of the metabolite or carbon source.
10 Examples include the FMD promoter or the modified AOX 1 promoters, which are modified in that they are repressible (see e.g. WO2006/089329A1). The promoter can be any promoter with and without additional regulatory sequences (such as e.g. repressor binding sites) facilitating translation of the GOI.

An "expression cassette" refers to a DNA coding sequence or segment of DNA
15 coding for an expression product that can be inserted into a vector at defined restriction sites. The cassette restriction sites are designed to ensure insertion of the cassette in the proper reading frame. Generally, DNA is inserted at one or more restriction sites of the vector DNA, and then is carried by the vector into a host cell along with the transmissible vector DNA. A segment or sequence of DNA having
20 inserted or added DNA, such as an expression vector, can also be called a "DNA construct".

Preferably, an expression cassette is a DNA construct comprising essentially a promoter, a gene of interest, and upstream of the gene of interest a Shine-Dalgarno (SO) sequence, also termed ribosome binding site (RBS).

25 The term "expression cassette" or "integration cassette" may also refer to a linear or circular DNA construct to be integrated, specifically "stably integrated", into the host genome, such as the bacterial chromosome. As a result of integration, the expression host cell has an integrated expression cassette. If the expression cassette is to be inserted into the genome of a host cell, the expression cassette preferably also comprises two terminally flanking regions which are homologous to a
30 genomic region and which enable homologous recombination. In addition, the cassette may contain other sequences such as for example sequences coding for anti-biotic selection markers, prototrophic selection markers or fluorescent markers,

markers coding for a metabolic gene, genes which improve protein expression or two flippase recognition target sites (FRT) which enable the removal of certain sequences (e.g. antibiotic resistance genes) after integration. The use of a linear expression cassette provides the advantage that the genomic integration site can be
5 freely chosen by the respective design of the flanking homologous regions of the cassette. Thereby, integration of the linear expression cassette allows for greater variability with regard to the genomic region. Expression products, such as the fusion protein described herein, can be expressed from an autonomously replicating nucleotide sequence, or from nucleotide sequences stably integrated into the ge-
10 nome or chromosome of a host cell.

Any of the known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, nucleofection, liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and integrative, and
15 any of the other well-known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al.).

Specifically, "functionally active variants" or "functionally active derivatives" of a peptide tag of the invention may be obtained by substituting, deleting, adding,
20 inserting and/or modifying one or more amino acids of/from/to/of the peptide, whose substitution(s), deletion(s), additions(s), insertion(s) and/or modification(s) preserve the function of the peptide according to the present invention. Specifically, the peptides of the invention or functionally active variants thereof, when used as an expression tag, increase the expression of a protein of interest, optionally, as compared to T7AC (SEQ ID NO:79) and/or as compared to T7A3 (SEQ ID NO:78) and/or
25 facilitate expression of a protein of interest. Specifically, the yield and/or titer is increased. Specifically, expression is increased at least by about 1.1 fold, 1.2 fold, 1.3 fold, 1.4 fold, 1.5 fold, 1.6 fold, 1.7 fold, 1.8 fold, 1.9 fold, 2 fold, 2.1 fold, 2.2 fold, 2.3 fold, 2.4 fold, 2.5 fold, 2.6 fold, 2.7 fold, 2.8 fold, 2.9 fold, 3 fold, 3.1 fold, 3.2 fold,
30 3.3 fold, 3.4 fold, 3.5 fold, 3.6 fold, 3.7 fold, 3.8 fold, 3.9 fold, 4 fold or even about 5 fold.

In the case of nucleotide sequences, a functionally active variant preferably means a nucleotide sequence having a sequence different from the original

nucleotide sequence, but which still codes for the same amino acid sequence, due to the use of the degenerated genetic code. Functional variants of a protein, in particular the peptide tags described herein, may be obtained by substituting one or more amino acids of the protein or peptide, which substitution(s) preserve(s) the function of the protein or peptide. Preferably, such substitutions are conservative amino acid substitutions. Specifically, functionally active variants of the peptide tags described herein comprise the first peptide sequence of SEQ ID NO:1 or SEQ ID NO:2, preferably at the same position as the peptide tag sequence from which the functionally active variant is derived.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:11 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:11 and comprises SEQ ID NO:1.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:12 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:12 and comprises SEQ ID NO:2.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:22 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:22 and comprises SEQ ID NO:1.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:23 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:23 and comprises SEQ ID NO:2.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:111 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:111 and comprises SEQ ID NO:1.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:113 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:113 and comprises SEQ ID NO:2.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:32 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:32 and comprises SEQ ID NO:1.

5 Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:35 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:35 and comprises SEQ ID NO:2.

10 Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:42 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:42 and comprises SEQ ID NO:1.

15 Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:47 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:47 and comprises SEQ ID NO:2.

20 Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:150 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:150 and comprises SEQ ID NO:1.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:155 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:155 and comprises SEQ ID NO:2.

25 Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:11 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:10 and in addition comprises SEQ ID NO:1, preferably at the N-terminus.

30 Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:12 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence

as shown in SEQ ID NO:10 and in addition comprises SEQ ID NO:2, preferably at the N-terminus.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:22 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%,
5 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:21 and in addition comprises SEQ ID NO:1, preferably at the N-terminus.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:23 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%,
10 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:21 and in addition comprises SEQ ID NO:2, preferably at the N-terminus.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:111 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%,
15 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:107 and in addition comprises SEQ ID NO:1, preferably at the N-terminus.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:113 has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%,
20 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:107 and in addition comprises SEQ ID NO:2, preferably at the N-terminus.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:32 has at least about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%,
25 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:85 and in addition comprises SEQ ID NO:1, preferably at the N-terminus.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:35 has at least about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%,
30 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:85 and in addition comprises SEQ ID NO:2, preferably at the N-terminus.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:42 has at least about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:87 and in addition comprises SEQ ID NO:1, preferably at the N-terminus.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:47 has at least about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:87 and in addition comprises SEQ ID NO:2, preferably at the N-terminus.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:150 has at least about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:160 and in addition comprises SEQ ID NO:1, preferably at the N-terminus.

Specifically, a functionally active variant of the amino acid sequence as shown in SEQ ID NO:155 has at least about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or even 100% amino acid sequence identity to the amino acid sequence as shown in SEQ ID NO:160 and in addition comprises SEQ ID NO:2, preferably at the N-terminus.

The term "sequence identity" as used herein is understood as the relatedness between two amino acid sequences or between two nucleotide sequences and described by the degree of sequence identity or sequence complementarity. The sequence identity of a variant, homologue or orthologue as compared to a parent nucleotide or amino acid sequence indicates the degree of identity of two or more sequences. Two or more amino acid sequences may have the same or conserved amino acid residues at a corresponding position, to a certain degree, up to 100%. Two or more nucleotide sequences may have the same or conserved base pairs at a corresponding position, to a certain degree, up to 100%.

Sequence similarity searching is an effective and reliable strategy for identifying homologs with excess (e.g., at least 50%) sequence identity. Sequence similarity search tools frequently used are e.g., BLAST, FASTA, and HMMER.

Sequence similarity searches can identify such homologous proteins or polynucleotides by detecting excess similarity, and statistically significant similarity that reflects common ancestry. Homologues may encompass orthologues, which are herein understood as the same protein in different organisms, *e.g.*, variants of such protein in different organisms or species.

To determine the % complementarity of two complementary sequences, one of the two sequences needs to be converted to its complementary sequence before the % complementarity can then be calculated as the % identity between the first sequence and the second converted sequences using the above-mentioned algorithm.

"Percent (%) identity" with respect to an amino acid sequence, homologs and orthologues described herein is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in the specific polypeptide sequence, after aligning the sequence and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

For purposes described herein, the sequence identity between two amino acid sequences can be determined using NCBI BLAST, specifically NCBI BLAST + 2.9.0 program version (Apr-02-2019).

"Percent (%) identity" with respect to a nucleotide sequence *e.g.*, of a nucleic acid molecule or a part thereof, in particular a coding DNA sequence, is defined as the percentage of nucleotides in a candidate DNA sequence that is identical with the nucleotides in the DNA sequence, after aligning the sequence and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent nucleotide sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting examples of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, 5 Novoalign (Novocraft Technologies; available at novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomies.org.cn), and Maq (available at maq.sourceforge.net).

As used herein the term "fusion protein" refers to a POI comprising, preferably at its N-terminus, an engineered fusion sequence comprising the peptide tag(s) de- 10 scribed herein. Such fusion sequence, specifically the peptide tag described herein, is "fused" or "linked" to a POI, thus making up the fusion protein. Specifically, such fusion sequence and POI linked to each other may be immediately linked, *i.e.* without further elements or nucleic acid sequences in between or may be indirectly linked with linkers or other sequences in between.

Such engineered fusion sequence may comprise a protease recognition and/or 15 cleavage site, *e.g.* a caspase recognition and/or cleavage site or a TEV-protease recognition and/or cleavage site, to facilitate separation of the POI and the fusion sequence. The fusion protein may be cleaved at the protease cleavage site by addition of a protease such as caspase or TEV-protease, *e.g.* to the cell culture, to 20 remove the peptide tag from the POI. In a specific aspect, the engineered fusion sequence described herein comprises one or more peptide tags as described herein, at least one caspase recognition site, and optionally one or more linkers. According to a specific example, the fusion protein comprises one or more peptide tags, optionally linked via linker sequences, one or more caspase recognition sites 25 and one or more POIs.

According to a specific embodiment, the fusion protein provided herein comprises a first part, comprising one or more peptide tags, optionally linked via linker sequences, a second part, comprising a recognition site for target-specific proteolytic cleavage, *e.g.* by caspases, cp-caspase, caspase-2 or cp caspase-2, or vari- 30 ants thereof having one or more modifications of the amino acid sequence and a third part, comprising a POI, whereas "cp" refers to a "circularly permuted" or reversed caspase. Specifically, the second part may be part of the peptide tag. Specifically, the fusion protein described herein may comprise each part more than once

and in different order. For example, the fusion protein provided herein may comprise a first part comprising a tag, a second part comprising a caspase recognition site, another first part comprising the same or a different tag sequence, another second part comprising the same or a different recognition site and a third part comprising a POI. According to a further example, the fusion protein described herein may comprise more than one POI separated by one or more fusion sequences comprising one or more recognition sites. The fusion protein described herein is encoded by a heterologous gene which is engineered in such a way that it is translated into protein by a host organism. Specifically, said heterologous gene comprises a nucleotide sequence encoding the peptide tag of the invention.

As described herein, the peptide tag of the invention comprises a first peptide sequence consisting of an amino acid sequence selected from SEQ ID NO:1 or SEQ ID NO:2, that can be used to facilitate expression of a POI. The peptide tag may comprise additional elements N-terminal or C-terminal of SEQ ID NO:1 and/or SEQ ID NO:2.

In a specific example, the peptide tag comprises further amino acids C-terminal of SEQ ID NO:1 and/or SEQ ID NO:2. Specifically, the peptide tag comprises at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 additional amino acids C-terminal of SEQ ID NO:1 and/or SEQ ID NO:2.

In a specific example, the peptide tag comprises a second peptide sequence of 8 additional amino acids, for example SEQ ID NO:5.

In another specific example, the peptide tag comprises a second peptide sequence of 19 additional amino acids, for example SEQ ID NO:10 or SEQ ID NO:21.

In another specific example, the peptide tag comprises a second peptide sequence of 9 additional amino acids, for example SEQ ID NO:107.

In addition, the peptide tag described herein may comprise further tag elements, such as those selected from the group consisting of one or more affinity tags, one or more solubility enhancement tags, one or more monitoring tags, and may comprise one or more linkers, and one or more protease recognition and/or cleavage sites.

As used herein, the term "tag element" refers to the first and second peptide sequences as described herein. The term "tag element" as used herein further refers to elements such as solubility enhancement tags, monitoring tags and affinity tags.

Specifically, the protease recognition and/or cleavage site refers to those recognized and/or cleaved by TEV protease or a caspase, cp-caspase or reversed caspase, preferably caspase-2, cp caspase-2 or reversed caspase-2.

Affinity tags are amino acid sequences that can be used for example for the purification of proteins where they are attached to (fusion proteins with affinity tag e.g. at its N-terminus). These affinity tags have high affinity to appropriate ligands of a solid support, like chromatography resins or directly to the resins. By selectively binding of the fusion protein having the affinity tag to the particular resin, the fusion protein can be purified highly effective by only one chromatography step. According to a specific embodiment, affinity tag sequences used herein are selected from histidine (His) tag, specifically a poly-histidine tag, arginine-tag, specifically a poly-arginine tag, peptide substrate for antibodies, chitin binding domain, RNase S peptide, protein A, β -galactosidase, FLAG tag, Strep II tag, streptavidin binding peptide (SBP) tag, calmodulin-binding peptide (CBP), glutathione S-transferase (GST), maltose-binding protein (MBP), S-tag, HA tag, or c-Myc tag or any other tag known to be useful for the efficient purification of a protein it is fused to. Preferably, the affinity tag is a His tag comprising one or more H, specifically a hexahistidine tag. Specifically, fusion proteins comprising a poly-, or hexa-histidine tag (His-tag) can be captured and purified by IMAC, preferably using a Ni-NTA chromatography material.

Solubility and/or expression enhancement tags used herein may be calmodulin-binding peptide (CBP), poly Arg, poly Lys, G B1 domain, protein D, Z domain of Staphylococcal protein A, and thioredoxin or any other tag known to improve the solubility and/or expression of the protein it is fused to, e.g. during expression in a host cell. Preferably the solubility tag is based on highly charged peptides of bacteriophage genes, for example such as those listed in US 8,535, 908 B2. Specifically, the solubility enhancement tag is selected from the group consisting of T7C, T7B, T7B1, T7B2, T7B3, T7B3, T7B4, T7B5, T7B6, T7B6, T7B7, T7B8, T7B9, T7B10, T7B11, T7B12, T7B13, T7A, T7A1, T7A2, T7A3, T7A4, T7A5, T7AC T3, N1, N2, N3, N4, N5, N6, N7, calmodulin-binding peptide (CBP), poly Arg, poly Lys, G B1

domain, protein D, Z domain of Staphylococcal protein A, DsbA, DsbC and thioredoxin.

According to a further specific embodiment, the monitoring tag sequence used herein is m-Cherry, GFP or f-Actin or any other tag useful for detection or quantification of the POI and/or the fusion protein during production of the fusion protein including fermentation, isolation and purification by simple in-situ, inline online or atline detectors, like UV, IR, Raman, Fluorescence and the like.

According to a specific embodiment, the peptide tag described herein comprises more than one additional tag elements, specifically it comprises an affinity tag and a solubility enhancement tag. Specifically, it comprises an affinity tag, a solubility enhancement tag and a monitoring tag. Specifically, it comprises more than one tag element of the same functionality, specifically it comprises more than one affinity tag, more than one solubility enhancement tag and/or more than one monitoring tag, and any combination thereof.

For example, the first peptide sequence described herein may be linked to an amino acid sequence comprising the second peptide sequence and further tag elements, which amino acid sequence may be selected from the group consisting of SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87 and SEQ ID NO:88.

Specifically, such tag elements may be immediately linked, *i.e.* without further elements or nucleic acid sequences in between or may be indirectly linked with linkers in between.

In another example, the first peptide sequence described herein may be linked to an amino acid sequence comprising the second peptide sequence and further tag elements, which amino acid sequence may be SEQ ID NO:160.

Specifically, such tag elements may be immediately linked, *i.e.* without further elements or nucleic acid sequences in between or may be indirectly linked with linkers in between.

In another example, the first peptide sequence described herein may be linked to an amino acid sequence comprising the second peptide sequence and further tag elements, which amino acid sequence may be SEQ ID NO:140.

Specifically, such tag elements may be immediately linked, *i.e.* without further elements or nucleic acid sequences in between or may be indirectly linked with linkers in between.

The term "linker" as used herein refers to any amino acid sequence that does not interfere with the function of elements being linked. Linkers may connect e.g., nucleotide sequences, or amino acid sequences. Linkers can be used between the tag and the POI or between tag sequences. The linkers may be used to engineer appropriate amounts of flexibility. Preferably, the linkers are short, e.g., 1–20 nucleotides or amino acids or even more and are typically flexible. Amino acid linkers commonly used consist of a number of glycine, serine, and optionally alanine, in any order. Such linkers usually have a length of at least any one of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or 20 amino acids, as required. Preferably, the linker comprises 1 to 12 amino acid residues, preferably it is a short linker. Preferably the linker is a GS, GSG, GGSGG (SEQ ID NO:80), GSAGSAAGSG (SEQ ID NO:81), (GS)_n, GSGSGSG (SEQ ID NO:82), GSG or GGGGS (SEQ ID NO:83) linker or any combination thereof. In some embodiments, the linker comprises one or more units, repeats or copies of a motif, such as for example GS, GSG or G₄S. The linker may also consist of a series of the same or different amino acids.

Further provided herein is a method of producing a POI using the peptide tags described herein. Specifically, the fusion protein described herein, comprising at least one POI and the peptide tag of the invention, is cloned into an expression vector under operable linkage to a promoter. Said expression vector is integrated into a host cell and the host cell is cultured under conditions allowing expression of the fusion protein, optionally following a growth phase for the accumulation of biomass before the recombinant protein is expressed. Integration may be stable integration into the genome or chromosome of the host cell or may be transient integration using plasmid(s).

The POI may be produced employing a fed-batch process as described herein, comprising an expression phase as described herein and optionally a growth phase as described herein.

According to a specific embodiment of the method of producing a POI as described herein, the fusion protein, comprising a protease recognition/ cleavage site, is contacted with a protease after expression, to cleave off the peptide tag and produce a POI comprising the desired N-terminus, *i.e.* the natural or designed N-terminus without any unwanted amino acids attached. Specifically, the fusion protein is contacted with the protease enzyme after isolation of the fusion protein from the

host cell culture. Specifically, the protease recognition site is recognized by a caspase, preferably caspase-2 and the fusion protein is contacted with the corresponding caspase, preferably caspase-2 or a variant thereof such as circularly permuted (cp) caspase-2 or mutated variants thereof having a higher P1'prime tolerance.

After production of the POI according to the method described herein, the POI may be further modified, purified and/or formulated.

In this context, the term "modifying the at least one protein of interest" is meant that the POI may be chemically, physically or enzymatically modified. There are many methods known in the art to modify proteins. Proteins can be coupled to carbohydrates or lipids. The POI may be PEGylated (the POI chemically coupled to polyethyleneglycole) or HESylated (the POI is chemically coupled to hydroxyethyl starch) for half-life extension. The POI may also be coupled with other moieties such as affinity domains for e.g. human serum albumin for half-life extension. The POI also may be treated by a protease, e.g. such as described above, or under hydrolytic conditions for cleavage to form the active ingredient from a pre-sequence or to cleave off a tag such as an affinity tag for purification. The POI may also be coupled to other moieties such as toxins, radioactive moieties or any other moiety. The POI may further be treated under conditions to form dimers, trimers and the like.

Additionally, the term "formulating the at least one protein of interest" refers to bringing the POI to conditions, where the POI can be stored for a longer time and/or for optimized pharmaceutical form and/or pharmaceutical application and/or to better adjust the concentration and/or to provide higher concentrations in liquid formulations. Many different methods known in the art are available to stabilize proteins. By exchanging the buffer in which the POI is existent after purification and / or modification, the POI can be brought under conditions, where it is more stable. Different buffer substances and additives, such as sucrose, mild detergents, stabilizer and the like, known in the art can be used. The POI can also be stabilized by lyophilization. For some POIs formulations can be done by formation of complexes of the POI with lipids or lipoproteins, such as polyplexes, and the like. Some protein may be co-formulated with other proteins.

The methods described herein specifically refer to the production of heterologous compounds. Such term used with respect to a nucleotide or amino acid

sequence or protein, refers to a compound which is either foreign, *i.e.* “exogenous”, such as not found in nature, to a given host cell; or that is naturally found in a given host cell, *e.g.*, is “endogenous”, however, in the context of a heterologous construct, *e.g.*, employing a heterologous nucleic acid, thus “not naturally-occurring”. The heterologous nucleotide sequence as found endogenously may also be produced in an unnatural, *e.g.*, greater than expected or greater than naturally found, amount in the cell. The heterologous nucleotide sequence, or a nucleic acid comprising the heterologous nucleotide sequence, possibly differs in sequence from the endogenous nucleotide sequence but encodes the same protein as found endogenously. Specifically, heterologous nucleotide sequences are those not found in the same relationship to a host cell in nature (*i.e.*, “not natively associated”). Any recombinant or artificial nucleotide sequence is understood to be heterologous.

Recombinant host cells according to the present invention can be obtained by introducing a vector or plasmid (such as an expression vector as mentioned above) comprising the target polynucleotide sequences, including the tag sequences described herein, into the cells. Techniques for transfecting or transforming eukaryotic cells or transforming prokaryotic cells are well known in the art. These can include lipid vesicle mediated uptake, heat shock mediated uptake, calcium phosphate mediated transfection (calcium phosphate/DNA co-precipitation), viral infection, particularly using modified viruses such as, for example, modified adenoviruses, microinjection and electroporation. For prokaryotic transformation, techniques can include heat shock mediated uptake, bacterial protoplast fusion with intact cells, microinjection and electroporation. The DNA can be single or double stranded, linear or circular, relaxed or supercoiled DNA. For various techniques for transfecting mammalian cells, see, for example, Keown et al. (1990) *Processes in Enzymology* 185:527-537. Procedures used to manipulate polynucleotide sequences, *e.g.* coding for the proteins of the present invention and/or the POI, the promoters, enhancers, leaders, etc., are well known to persons skilled in the art, *e.g.* described by J. Sambrook et al., *Molecular Cloning: A Laboratory Manual* (3rd edition), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, New York (2001).

Recombinant host cells according to the present invention may also comprise tag sequences described herein stably integrated into their genome. A foreign or target polynucleotide such as the nucleic acid sequence encoding a peptide tag as

described herein can be inserted into the chromosome by various means, e.g., by homologous recombination or by using a hybrid recombinase that specifically targets sequences at the integration sites. The foreign or target polynucleotide comprising a tag sequence as described herein and/or a GOI is typically present in a vector ("inserting/integration vector"), preferably within an expression cassette. These vectors are typically circular and linearized before they are used for homologous recombination. As an alternative, the foreign or target polynucleotides may be DNA fragments joined by fusion PCR or synthetically constructed DNA fragments which are then recombined into the host cell. In addition to the homology arms, the vectors may also contain markers suitable for selection or screening, an origin of replication, and other elements. It is also possible to use heterologous recombination which results in random or non-targeted integration. Heterologous recombination refers to recombination between DNA molecules with significantly different sequences. Methods of recombinations are known in the art and for example described in Boer et al., Appl Microbiol Biotechnol (2007) 77:513–523. One may also refer to Principles of Gene Manipulation and Genomics by Primrose and Twyman (7th edition, Blackwell Publishing 2006) for genetic manipulation of yeast cells.

The phrase "culturing said host cell under conditions permitting" expression of a target protein refers to maintaining and/or growing host cells under conditions (e.g., but not limited to temperature, aeration, pressure, pH, induction, growth rate, culture medium, nutrients duration of the cultivation, mode of nutrient feed(s) etc.) appropriate or sufficient to obtain production of the desired compound (fusion protein, POI).

A host cell may preferably first be cultivated at conditions to grow efficiently to a large cell number without the burden of expressing a protein. When the cells are prepared for POI expression, suitable cultivation conditions are selected and optimized to produce the POI.

An inducible promoter may be used that becomes activated as soon as an inductive stimulus is applied to direct transcription of the gene under its control. An inductive stimulus is preferably the addition of an appropriate agents (e.g. methanol for the AOX-promoter or IPTG or lactose for promoters under the control of lac operators (binding sites for the lac repressor) such as the lac, T7, or A1 promoters) or the depletion of an appropriate nutrient (e.g., methionine for the MET3-promoter or

phosphate for the *phoA* promoter). Also, the addition of ethanol, methylamine, cadmium or copper as well as heat or an osmotic pressure increasing agent can induce the expression depending on the promoters operably linked to the proteins of the invention and the POI(s). Some inducible or de-repressible promoters can be induced by depletion of a certain medium component or metabolite, such as phosphate for the *phoA* promoter. An inductive stimulus can also be a pH, temperature or osmotic pressure change. Some promoters are activated by a combination of an inductive stimulus and de-repression by depletion or limitation of certain metabolites.

It is preferred to cultivate the host cell(s) according to the invention in a bioreactor under optimized growth conditions to obtain a cell density of at least 1 g/L, preferably at least 10 g/L cell dry weight, more preferably at least 50 g/L cell dry weight. It is advantageous to achieve such yields of biomolecule production not only on a laboratory scale, but also on a pilot or industrial scale.

Preferably, the host cells are cultivated in a minimal medium with a suitable carbon source, thereby further simplifying the isolation process significantly. By way of example, the minimal medium contains a utilizable carbon source (e.g. glucose, glycerol, ethanol or methanol), salts containing the macro elements (potassium, magnesium, calcium, ammonium, chloride, sulphate, phosphate) and trace elements (copper, iodide, manganese, molybdate, cobalt, zinc, and iron salts, and boric acid).

In the case of yeast cells, the cells may be transformed with one or more of the above-described expression vector(s) and cultured in conventional nutrient media, optionally modified as appropriate for inducing promoters, selecting transformants or amplifying the genes encoding the desired sequences. A number of minimal media suitable for the growth of yeast are known in the art. Any of these media may be supplemented as necessary with salts (such as sodium chloride, calcium, magnesium, and phosphate), buffers (such as HEPES, citric acid and phosphate buffer), nucleosides (such as adenosine and thymidine), antibiotics, trace elements, vitamins, and glucose or an equivalent energy source. Any other necessary supplements may also be included at appropriate concentrations that would be known to those skilled in the art. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression and are known

to the ordinarily skilled artisan. Cell culture conditions for other type of host cells are also known and can be readily determined by the artisan. Descriptions of culture media for various microorganisms are for example contained in the handbook "Manual of Methods for General Bacteriology" of the American Society for Bacteriology
5 (Washington D.C, USA, 1981).

Host cells can be cultured (*e.g.*, maintained and/or grown) in liquid media and preferably are cultured, either continuously or intermittently, by conventional culturing methods such as standing culture, test tube culture, shaking culture (*e.g.*, rotary shaking culture, shake flask culture, etc.), aeration spinner culture, or fermentation.
10 In some embodiments, cells are cultured in shake flasks or deep well plates. In yet other embodiments, cells are cultured in a bioreactor (*e.g.*, in a bioreactor cultivation process). Cultivation processes include, but are not limited to, batch, fed-batch and continuous methods of cultivation. The terms "batch process" and "batch cultivation" refer to a closed system in which the composition of media, nutrients, supplemental
15 additives and the like is set at the beginning of the cultivation and not subject to alteration during the cultivation; however, attempts may be made to control such factors as pH and oxygen concentration to prevent excess media acidification and/or cell death. The terms "fed-batch process" and "fed-batch cultivation" refer to a batch cultivation with the exception that one or more substrates and/or supplements
20 and/or inducers are added (*e.g.*, added in increments or continuously) as the cultivation progresses in one or more feed solutions/suspensions. The terms "continuous process" and "continuous cultivation" refer to a system in which a defined cultivation media is added continuously to a bioreactor and an equal amount of used or "conditioned" media is simultaneously removed, for example, for recovery of the
25 desired product. A variety of such processes has been developed and is well-known in the art.

In some embodiments, host cells are cultured for about 12 to 24 hours, in other embodiments, host cells are cultured for about 24 to 36 hours, about 36 to 48 hours, about 48 to 72 hours, about 72 to 96 hours, about 96 to 120 hours, about 120 to 144
30 hours, or for a duration greater than 144 hours. In yet other embodiments, culturing is continued for a time sufficient to reach desirable production yields of POI.

The above-mentioned methods may further comprise a step of isolating the expressed at least one target protein, *e.g.* the POI or fusion protein described herein,

from the cell culture and optionally followed by a step of purifying the at least one POI or fusion protein. If the target protein is secreted from the cells, it can be isolated and then purified from the culture medium using state of the art techniques. Secretion of the target protein from the cells is generally preferred, since the products are recovered from the culture supernatant rather than from the complex mixture of proteins that results when cells are disrupted to release intracellular proteins. A protease inhibitor, such as phenyl methyl sulfonyl fluoride (PMSF) or any other means may be useful to inhibit proteolytic degradation during purification, and antibiotics may be included to prevent the growth of adventitious contaminants. The composition may be concentrated, filtered, dialyzed, etc., using methods known in the art. The cell culture after fermentation / cultivation can be centrifuged using a separator or a tube centrifuge to separate the cells from the culture supernatant or the cells can be separated from the supernatant by filtration, e.g. depth or tangential flow filtration. The supernatant can then be filtered and/or concentrated by using a tangential flow filtration. Alternatively, cultured host cells may also be ruptured sonically or mechanically (e.g. high pressure homogenisation), enzymatically or chemically or by heat to obtain a cell extract containing the desired POI, from which the POI may be isolated and purified.

The step of "removing the peptide tag", may be effected by cleavage of the tag using a protease as described herein, such as for examples a cp caspase-2.

Isolation and purification methods, also referred to as methods of recovery, for obtaining the POI, including "recovering" and/or "purifying" the POI, may be based on methods utilizing difference in solubility, such as salting out, solvent precipitation, heat precipitation, methods utilizing difference in molecular weight, such as size exclusion chromatography, ultrafiltration and gel electrophoresis, methods utilizing difference in electric charge, such as ion-exchange chromatography, methods utilizing specific affinity, such as affinity chromatography, methods utilizing difference in hydrophobicity, such as hydrophobic interaction chromatography and reverse phase high performance liquid chromatography, methods utilizing difference in isoelectric point, such as isoelectric focusing may be used and methods utilizing certain amino acids, such as IMAC (immobilized metal ion affinity chromatography). If the POI is expressed as inactive and soluble Inclusion Bodies the solubilized Inclusion Bodies need to be refolded and may be purified. When the POI is fused to a His-tag such

as e.g. the 6 His-tag, which can be included in the peptide tag described herein, IMAC may be used for capturing and purification of the fusion protein. After cleavage of the fusion protein to release the POI, the POI can be separated by IMAC in the flow through mode from the cleaved tag, un-cleaved fusion protein and the protease, if the protease also comprised a His-tag.

The isolated and purified POI can be identified by conventional methods such as Western Blotting or specific assays for POI activity. The structure of the purified POI can be determined by amino acid analysis, amino-terminal peptide sequencing, primary structure analysis for example by mass spectrometry, RP-HPLC, ion exchange-HPLC, ELISA and the like. It is preferred that the POI is obtainable in large amounts and in a high purity level, thus meeting the necessary requirements for being used as an active ingredient in pharmaceutical compositions or as feed or food additive.

Further described herein are the following items:

1. A peptide tag for expression of a recombinant protein of interest (POI) in a host cell, comprising a first peptide sequence consisting of an amino acid sequence selected from SEQ ID NO:1 or SEQ ID NO:2.
2. The peptide tag of item 1, comprising a second peptide sequence comprising an amino acid sequence selected from SEQ ID NO:5, SEQ ID NO:10 or SEQ ID NO:21, preferably wherein said second peptide sequence is C-terminal of the first peptide sequence.
3. The peptide tag of item 1 or 2, comprising at least two first peptide sequences, each consisting of an amino acid sequence selected from SEQ ID NO:1 or SEQ ID NO:2, preferably wherein said at least two first peptide sequences are directly linked.
4. The peptide tag of any one of items 1 to 3, wherein the first peptide sequence is directly linked to the second peptide sequence.
5. The peptide tag of any one of items 1 to 4, further comprising one or more linker sequences.
6. The peptide tag of any one of items 1 to 5, further comprising a solubility and/or expression enhancement tag, a monitoring tag and/or an affinity tag.

7. The peptide tag of any one of items 1 to 6, further comprising a protease recognition and/or cleavage site, preferably at its C-terminus.

8. The peptide tag of any one of items 1 to 7, wherein the host cell is a eukaryotic or prokaryotic host cell, preferably a yeast cell, a mammalian cell or a bacterial cell.

9. The peptide tag of item 8, wherein the host cell is a mammalian host cell, preferably a CHO cell.

10. The peptide tag of any one of items 1 to 8, comprising an amino acid sequence selected from the group consisting of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:32, SEQ ID NO:35, SEQ ID NO:42 and SEQ ID NO:47, or a functionally active variant thereof comprising an amino acid sequence having at least 95% sequence identity to SEQ ID NO:10, SEQ ID NO:21, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87 or SEQ ID NO:88 and comprising SEQ ID NO:1 or SEQ ID NO:2, preferably at its N-terminus.

11. The peptide tag of item 9, wherein the peptide tag comprises the amino acid sequence SEQ ID NO:11 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16.

12. The peptide tag of item 9, wherein the peptide tag comprises the amino acid sequence SEQ ID NO:12 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 and SEQ ID NO:20.

13. The peptide tag of item 9, wherein the peptide tag comprises the amino acid sequence SEQ ID NO:22 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26 and SEQ ID NO:27.

14. The peptide tag of item 9, wherein the peptide tag comprises the amino acid sequence SEQ ID NO:23 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30 and SEQ ID NO:31.

15. The peptide tag of item 9, wherein the peptide tag comprises the amino acid sequence SEQ ID NO:32 and is encoded by a nucleotide sequence

selected from the group consisting of SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:38, and SEQ ID NO:39.

16. The peptide tag of item 9, wherein the peptide tag comprises the amino acid sequence SEQ ID NO:35 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:40, and SEQ ID NO:41.

17. The peptide tag of item 9, wherein the peptide tag comprises the amino acid sequence SEQ ID NO:42 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, and SEQ ID NO:46.

18. The peptide tag of item 9, wherein the peptide tag comprises the amino acid sequence SEQ ID NO:47 and is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, and SEQ ID NO:51.

19. An expression vector for expression of a recombinant protein of interest (POI) as a fusion protein comprising the peptide tag of any one of items 1 to 18 fused to one terminus of said POI, wherein said expression vector comprises a nucleic acid sequence encoding said peptide tag of any one of items 1 to 18 and a nucleic acid sequence encoding said POI.

20. The expression vector of item 19, wherein the nucleic acid sequence encoding the peptide tag of any one of items 1 to 18 is selected from the group consisting of SEQ ID NO:13-20, SEQ ID NO:24-31, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:36-41, SEQ ID NO:43-46 and SEQ ID NO:48-51.

21. The expression vector of item 19 or 20, wherein the one terminus of the POI is the N-terminus.

22. The expression vector of any one of items 19 to 21, comprising:

- a promoter;
- a ribosome binding site (RBS);
- a start codon;
- the sequence encoding the peptide tag of any one of items 1 to 18, preferably wherein said sequence encoding the peptide tag is placed immediately after the start codon; and

e. the sequence encoding the POI, preferably wherein said sequence encoding the POI is placed immediately after the sequence encoding the peptide tag.

23. An expression vector for expression of a recombinant protein of interest (POI) as a fusion protein, comprising:

- a. a promoter;
- b. a ribosome binding site (RBS);
- c. a start codon;
- d. a nucleic acid sequence encoding the peptide tag of any one of items 1 to 18, preferably wherein said sequence encoding the peptide tag is placed immediately after the start codon; and
- e. a cloning site, preferably placed immediately downstream of the sequence encoding the peptide tag.

24. An expression cassette for expression of a recombinant protein of interest (POI) as a fusion protein comprising the peptide tag of any one of items 1 to 18 fused to one terminus of said POI, wherein said expression cassette comprises a nucleic acid sequence encoding said peptide tag of any one of items 1 to 18 and a nucleic acid sequence encoding said POI.

25. The expression cassette of item 24, wherein the nucleic acid sequence encoding the peptide tag of any one of items 1 to 18 is selected from the group consisting of SEQ ID NO:13-20, SEQ ID NO:24-31, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:36-41, SEQ ID NO:43-46 and SEQ ID NO:48-51.

26. The expression cassette of item 24 or 25, wherein the one terminus of the POI is the N-terminus.

27. The expression cassette of any one of items 24 to 26, comprising:

- a. a promoter;
- b. a ribosome binding site (RBS);
- c. a start codon;
- d. the sequence encoding the peptide tag of any one of items 1 to 18, preferably wherein said sequence encoding the peptide tag is placed immediately after the start codon; and

e. the sequence encoding the POI, preferably wherein said sequence encoding the POI is placed immediately after the sequence encoding the peptide tag.

28. An expression cassette for expression of a recombinant protein of interest (POI) as a fusion protein, comprising:

- a. a promoter;
- b. a ribosome binding site (RBS);
- c. a start codon;
- d. a nucleic acid sequence encoding the peptide tag of any one of items 1 to 18, preferably wherein said sequence encoding the peptide tag is placed immediately after the start codon; and
- e. a cloning site, preferably placed immediately downstream of the sequence encoding the peptide tag.

29. A fusion protein comprising a recombinant protein of interest (POI) and the peptide tag of any one of items 1 to 18, wherein the peptide tag of any one of items 1 to 18 is fused to the recombinant POI, preferably at the N-terminus of the recombinant protein.

30. A method of producing a recombinant protein of interest (POI) comprising expressing the POI in the form of a fusion protein comprising the peptide tag of any one of items 1 to 18 fused to one terminus of said POI, by a prokaryotic or eukaryotic expression system, preferably by a bacterial, yeast or mammalian cell expression system.

31. The method of item 30, wherein the bacterial expression system is an *E. coli* expression system.

32. The method of item 30, wherein the yeast expression system is an expression system using yeast cells of the genus *Komagataella* or *Saccharomyces*.

33. The method of item 30, wherein the mammalian expression system is an expression system using CHO cells.

34. The method of any one of items 30 to 33, wherein said fusion protein is expressed from an expression vector, preferably a plasmid, within a host cell or from an expression cassette stably integrated into the genome or chromosome of a host cell.

35. The method of any one of items 30 to 34, comprising culturing the host cell for a period of time under conditions permitting expression of the POI in the form of a fusion protein and comprising any one or more of the steps of:

- a. recovering the fusion protein and/or POI;
- b. purifying the fusion protein and/or POI;
- c. removing the peptide tag;
- d. further purifying the fusion protein and/or POI;
- e. formulating the fusion protein and/or POI;
- f. modifying the fusion protein and/or POI.

36. A nucleic acid sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:32, SEQ ID NO:35, SEQ ID NO:42 and SEQ ID NO:47, or a functionally active variant thereof comprising an amino acid sequence having at least 95% sequence identity to SEQ ID NO:10, SEQ ID NO:21, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87 or SEQ ID NO:88 and comprising SEQ ID NO:1 or SEQ ID NO:2, preferably at its N-terminus.

37. A nucleic acid sequence encoding the fusion protein of item 29, preferably comprising a start codon.

38. A peptide tag for expression of a recombinant protein of interest (POI) in a host cell, comprising an amino acid sequence of SEQ ID NO:84, wherein

- a. X at position 2 is L or H;
- b. X at position 3 is V or S;
- c. X at position 16 is E or Q; and
- d. X at position 18 is E or Q.

39. A host cell comprising the peptide tag of any one of items 1 to 18 and/or a nucleic acid sequence encoding said peptide tag.

40. A host cell comprising a fusion protein comprising the peptide tag of any one of items 1 to 18 and a protein of interest (POI) and/or a nucleic acid sequence encoding said fusion protein.

41. A host cell comprising the expression vector of any one of items 19 to 23 or the expression cassette of any one of items 24 to 29.

42. The peptide tag of item 9, wherein the peptide tag comprises the amino acid sequence SEQ ID NO:111 and is encoded by a nucleotide sequence

selected from the group consisting of SEQ ID NO:112, SEQ ID NO:144, SEQ ID NO:145 and SEQ ID NO:146.

43. The peptide tag of item 9, wherein the peptide tag comprises the amino acid sequence SEQ ID NO:113 and is encoded by a nucleotide sequence
5 selected from the group consisting of SEQ ID NO:114, SEQ ID NO:147, SEQ ID NO:148 and SEQ ID NO:149.

44. The peptide tag of item 9, wherein the peptide tag comprises the amino acid sequence SEQ ID NO:150 and is encoded by a nucleotide sequence
10 selected from the group consisting of SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, and SEQ ID NO:154.

45. The peptide tag of item 9, wherein the peptide tag comprises the amino acid sequence SEQ ID NO:155 and is encoded by a nucleotide sequence
selected from the group consisting of SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:158, and SEQ ID NO:159.

15 46. An expression vector for expression of a recombinant protein of interest (POI) as a fusion protein comprising the peptide tag of any one of items 42 to 45 fused to one terminus of said POI, wherein said expression vector comprises a nucleic acid sequence encoding said peptide tag of any one of items 42 to 45 and a nucleic acid sequence encoding said POI.

20 47. The expression vector of item 46, wherein the nucleic acid sequence encoding the peptide tag of any one of items 42 to 45 is selected from the group consisting of SEQ ID NO:112 and 144-146, SEQ ID NO: 113 and 147-149, SEQ ID NO:151-154 and SEQ ID NO:156-159.

48. An expression cassette for expression of a recombinant protein of interest (POI) as a fusion protein comprising the peptide tag of any one of items 42 to
25 45 fused to one terminus of said POI, wherein said expression cassette comprises a nucleic acid sequence encoding said peptide tag of any one of items 42 to 45 and a nucleic acid sequence encoding said POI.

49. The expression cassette of item 48, wherein the nucleic acid sequence encoding the peptide tag of any one of items 42 to 45 is selected from the
30 group consisting of SEQ ID NO:112 and 144-146, SEQ ID NO: 114 and 147-149, SEQ ID NO:151-154 and SEQ ID NO:156-159.

50. A nucleic acid sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NO:111, SEQ ID NO:114, SEQ ID NO:150 and SEQ ID NO:155, or a functionally active variant thereof comprising an amino acid sequence having at least 95% sequence identity to SEQ ID NO:107 or SEQ ID NO:160 and comprising SEQ ID NO:1 or SEQ ID NO:2, preferably at its N-terminus.

The following examples are put forth to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the subject invention and are not intended to limit the scope of what is regarded as the invention and defined in the claims. Efforts have been made to ensure accuracy with respect to the numbers used (e.g. amounts, temperature, concentrations, etc.) but some experimental errors and deviations should be allowed for.

EXAMPLES

The examples below will demonstrate that the newly identified peptide tag(s) increase(s) the titer (product per volume in mg/L, in the examples called also called "volumetric titer") and the yield (product per biomass in mg/g biomass measured as dry cell weight and in the examples also called "specific titer"), respectively, of recombinant proteins upon their expression. As an example, the volumetric titer and the yield of the recombinant proteins PTH and hFGF2 in the bacteria *E. coli* are increased.

Example 1: Materials and Methods

1.1. *Escherichia coli* strains

Strains used for recombinant protein expression as well as for cloning purposes were acquired from New England Biolabs (NEB, Ipswich, MA, USA). Chemically competent *E. coli* BL21(DE3) cells were used for all protein expression experiments, while chemically competent NEB-5 α cells were applied for cloning purposes. Transformations and subsequent cultivations for cloning purposes were performed according to the manufacturer's instructions.

1.2. Generation of expression systems

Expression vectors were created using backbones from existing pET30acer plasmids from a previous study by Köppl et al. 2022. Q5® High-Fidelity DNA Polymerase, Bsal-HF®v2, DpnI and T4 DNA Ligase were purchased from NEB. Construction of the vector plasmids followed a standard cloning protocol. In brief, site-directed mutagenesis was carried out utilizing primers carrying the desired mutations. A polymerase chain reaction (PCR) was performed employing Q5® High-Fidelity DNA Polymerase to amplify the original plasmid backbone and simultaneously insert the wanted mutations. After amplification, the linear DNA fragment was purified using the Monarch® PCR & DNA Cleanup Kit (NEB) followed by restriction digest with Bsal-HF®v2 and DpnI (37°C, overnight) to create specific overhangs for the following ligation step and cut the template plasmid used for PCR amplification. After preparative agarose gel (2% agarose) purification (90 V, 90 min), the correctly sized band was excised and dissolved using the Monarch® DNA Gel Extraction Kit (NEB). Ligation to generate circular plasmid DNA was performed at 16 °C overnight using T4 DNA Ligase (NEB) and the ligation mix was directly transformed into chemically competent cells according to the manufacturer's instructions. A comprehensive list of all primers and DNA fragments used can be found in Table 1. All oligonucleotides labelled with the letter P (being the primers for site directed mutagenesis – P1 to P7 - and subsequent screening and sequencing – P8 and P9) were bought from Sigma Aldrich (Taufkirchen, Germany). Primers for site directed mutagenesis were used as follows: P1 and P2 were used to create the DBv1-T7AC construct, primers P3 and P4 were used to create the DBv2-T7AC constructs, P5 and P7 were used to create the DBv1-T7AC_{new} construct and P6 and P7 were used to create the DBv2-T7AC_{new} constructs. After cloning, all constructs were sequenced to confirm their correctness.

Table 1. Primers used in the Examples. Lower case letters denote BsaI cleavage sites with accompanying inserted bases for specific overhang creation.

Primer number	Name	Sequence (5'-3')
P1	DBv1-rev	ggtctctagggCATATGTATATCTCCTTCTTAAAGTTA (SEQ ID NO:89)
P2	DBv1-for	ggtctcacccttggtcCCGGAACGCAACAAAG (SEQ ID NO:90)
P3	DBv2-rev	ggtctctgtggCATATGTATATCTCCTTCTTAAAGTTA (SEQ ID NO:91)
P4	DBv2-for	ggtctcaccacacagcCCGGAACGCAACAAAG (SEQ ID NO:92)
P5	DBv1-NEW-rev	ggtctcttttcgGGGACCAAGGGCATATG (SEQ ID NO:93)
P6	DBv2-NEW-rev	ggtctcttttcgGGGCTGTGTGGCATAT (SEQ ID NO:94)
P7	T7AC-NEW-for	ggtctcaGAAAGGAATAAAGAAAGGAAAGAAGCTGAGTT-GCAAG (SEQ ID NO:95)
P8	Seq-MCS-for	GAAGCAGCCCAGTAGTAG (SEQ ID NO:96)
P9	Seq-MCS-rev	CGGATATAGTTCCTCCTTCAG (SEQ ID NO:97)

All GOIs were expressed under the T7 promoter/operator system.

5 The resulting fusion proteins are shown in Table 4.

1.1. Laboratory scale fed-batch fermentations

All fermentation procedures, as well as media preparation and composition were performed according to Fink et al. 2019, 2021.

10 In each experimental series, the same fermentation process was repeated in an identical fashion for each construct. Briefly, cells were cultivated in the DASGIP parallel bioreactor system (Eppendorf SE, Hamburg, Germany) using a vessel with 2.1 L volume which allows for a maximum working volume of 1.8 L. To control relevant process parameters, the bioreactors were equipped with pH probes (Hamilton, Bonaduz, GR, Switzerland), optical dissolved O₂ probes (Hamilton, Bonaduz, GR, Switzerland) and temperature probes. The dissolved oxygen concentration in the media was kept constantly at >30% and the pH was sustained at 7.0 ± 0.2 by the addition of 12.5% ammonia solution. Temperature of the cultivation media was maintained at 37 ± 0.2 °C during the batch phase and shifted to 30 ± 0.2 °C at the beginning of the feed. Induction was facilitated by the introduction of Isopropyl-β-D-thiogalactopyranoside (IPTG). All fermentation parameters not mentioned here are listed in Table 2.

20

Table 2. Fermentation strategy for all DB-T7AC and DB-T7AC_{new} experiments using PTH and hFGF2 as model proteins (production during the last 1.2 generations).

	<i>Temp. [°C]</i>	<i>CDM [g/L]</i>	<i>Vol. [mL]</i>	<i>Duration [h]</i>	<i>μ [h⁻¹]</i>	<i>Generations</i>
Batch	37	6.7	500	7.5	-	-
Fed- Batch	30	40.4	1014	21.0	0.1	3.6

5 Induction of fermentations according to Table 2 was carried out at feed hour 17 with an IPTG pulse added directly into the bioreactor corresponding to 2 μ mol IPTG per gram of biomass at the end of fermentation.

1.2. Cell Lysis

10 Cell lysis and separation of the soluble und insoluble intracellular protein fractions was carried out according to Fink et al. (2019) [7]. Additionally, the cell lysis buffer contained 4 mmol L⁻¹ of NuPAGE Sample Reducing Agent (10x) (Invitrogen, Waltham, MA, USA).

1.3. Recombinant protein quantification

15 To estimate the recombinant protein concentration present in the cell lysate, reducing SDS-polyacrylamide gel electrophoresis (PAGE) was performed as described by Stargardt et al. (2020) and Cserjan-Puschmann et al. (2020) [2, 8]. Bovine Serum Albumin heat shock fraction (Sigma Aldrich, St. Louis, MO, USA) in the concentrations of 25, 50 and 75 μ g/mL was used as quantification standard. A brief description of the exact analytical procedure has been published by Köppl et al. (2022) [4].

Example 2: Novel peptide tags enhance protein expression

2.1 Generation of Fusion constructs

25 Recently, the development of a universally applicable, generic fusion tag, the CASPON™ tag, has been reported. This tag does not only facilitate high soluble titres but also enhances recombinant protein production in *E. coli* (Köppl et al. 2022). The CASPON™ tag comprises SEQ ID NO:52 (here shown with a starting methionine), including a T7AC tag at its N-terminus.

To further improve the efficiency of recombinant protein expression, two novel peptide sequences, named DBv1 (SEQ ID NO:1) and DBv2 (SEQ ID NO:2), were incorporated at the N-terminus of the T7AC tag at amino acid positions +2 to +4 (see Figure 1).

Furthermore, performance of the novel peptide tags was assessed. The resulting amino acid and nucleotide sequences of the tags are shown in Table 3.

Table 3. Tag sequences (T7AC-tag means the T7AC-tag encoded by SEQ ID NO:53; T7ACnew-tag means the T7AC-tag encoded by SEQ ID NO:54)

Tag	Amino acid sequence	Nucleotide sequence
T7AC	LEDPERNKERKEAEL- QAQTAEQ (SEQ ID NO:79)	CTG GAG GAT CCG GAA CGC AAC AAA GAG CGA AAG GAA GCT GAG TTG CAA GCT CAA ACC GCT GAG CAA (SEQ ID NO:53)
T7ACnew	LEDPERNKERKEAEL- QAQTAEQ (SEQ ID NO:79)	CTA GAA GAT CCC GAA AGG AAT AAA GAA AGG AAA GAA GCT GAG TTG CAA GCT CAA ACC GCT GAG CAA (SEQ ID NO:54)
DBv1- T7AC	PLVPERNKERKEAEL- QAQTAEQ (SEQ ID NO:11)	CCC TTG GTC CCG GAA CGC AAC AAA GAG CGA AAG GAA GCT GAG TTG CAA GCT CAA ACC GCT GAG CAA (SEQ ID NO:13)
DBv1- T7ACnew	PLVPERNKERKEAEL- QAQTAEQ (SEQ ID NO:11)	CCC TTG GTC CCC GAA AGG AAT AAA GAA AGG AAA GAA GCT GAG TTG CAA GCT CAA ACC GCT GAG CAA (SEQ ID NO:14)
DBv2- T7AC	PHSPERNKERKEAEL- QAQTAEQ (SEQ ID NO:12)	CCA CAC AGC CCG GAA CGC AAC AAA GAG CGA AAG GAA GCT GAG TTG CAA GCT CAA ACC GCT GAG CAA (SEQ ID NO:17)
DBv2- T7ACnew	PHSPERNKERKEAEL- QAQTAEQ (SEQ ID NO:12)	CCA CAC AGC CCC GAA AGG AAT AAA GAA AGG AAA GAA GCT GAG TTG CAA GCT CAA ACC GCT GAG CAA (SEQ ID NO:18)

The free folding energy at the mRNA nucleotide positions -4 to +37 of the tag sequences was calculated using the method developed by Zuker et al. 1980, resulting in e.g. a $\Delta G_{\min}$ of -7.50 kcal/mol for T7AC and -1.50 kcal/mol for T7ACnew (see Fig. 2).

The different tag sequences were then introduced into the CASPON™ tag, fused to two different pharmaceutically relevant model proteins: parathyroid hormone (PTH) and human fibroblast growth factor 2 (hFGF2) tested in a series of carbon limited fed-batch fermentations.

For a general structure of the expression constructs see Fig. 2. For sequences of specific constructs tested in this Example see Table 4 and the respective sequences in Fig. 1.

5 Table 4. Sequences of fusion constructs

Construct name	Amino acid sequence	Nucleotide Sequence
T7AC-PTH	SEQ ID NO:57	SEQ ID NO:58
T7ACnew-PTH	SEQ ID NO:57	SEQ ID NO:59
DBv1-T7AC-PTH	SEQ ID NO:60	SEQ ID NO:61
DBv1-T7ACnew-PTH	SEQ ID NO:60	SEQ ID NO:62
DBv2-T7AC-PTH	SEQ ID NO:63	SEQ ID NO:64
DBv2-T7ACnew-PTH	SEQ ID NO:63	SEQ ID NO:65
T7AC-hFGF2	SEQ ID NO:68	SEQ ID NO:69
T7ACnew-hFGF2	SEQ ID NO:68	SEQ ID NO:70
DBv2-T7AC-hFGF2	SEQ ID NO:71	SEQ ID NO:72
DBv2-T7ACnew-hFGF2	SEQ ID NO:71	SEQ ID NO:73

2.2 Cellular fitness of fusion constructs

The above constructs were cloned into the plasmid pET30acer to drive the expression of the fusion proteins, and the resulting plasmids were transformed into the host strain BL21(DE3). In identical laboratory-scale carbon-limited fed-batch culti-
 10 vations (according to Table 2) all constructs were investigated regarding their influence on production strain growth (Figure 3 C and F) and expression of fusion protein (Figure 3 A, B, D and E) immediately before and at four time points after induction. The fitness of all production hosts during the fermentations was comparable, as all
 15 variants almost reached the theoretical values of the calculated cell dry mass (CDM).

Cellular growth of all tag variants, regardless of the expressed model protein was highly similar in all fermentations. Biomass formation followed the predicted values reasonably well with a trend to reduced cellular growth at the end of fermenta-
 20 tions (Figure 3 C and F). The theoretical CDM at fermentation end is 40.4 g/L. Measured CDM concentrations are summarized in Table 5 below.

Table 5. Measured CDM concentrations, soluble recombinant protein titer (volumetric (vol.) titer) and yield (specific (spec.) recombinant protein titer) at the end of fermentations.

	CDM [g/L]	spec. titer [mg/g CDM]	vol. titer [g/L]
T7AC-PTH	38.8	46.2	1.8
T7ACnew-PTH	38.7	97.2	3.6
DBv1-T7AC-PTH	35.9	58.0	2.1
DBv1-T7ACnew-PTH	36.0	232.0	8.6
DBv2-T7AC-PTH	41.1	145.2	5.9
DBv2-T7ACnew-PTH	39.6	195.3	7.7
T7AC-hFGF2	36.0	56.4	2.0
T7ACnew-hFGF2	37.5	72.7	2.7
DBv2-T7AC-hFGF2	37.1	102.4	3.8
DBv2-T7ACnew-hFGF2	34.9	127.5	4.7

2.3 Increased POI production

Recombinant protein formation was higher with the DBv1-T7AC-tag and DBv2-T7AC-tag compared to the T7AC-tag for every tested model protein. The same has been observed with the DBv1-T7ACnew and DBv2-T7ACnew variants compared to the T7ACnew variant (see Table 5). The comparative increase in recombinant soluble protein titre at the end of fermentation is dependent on the fusion tag variant used. With PTH increases in specific titre of 1.26-fold and 3.14-fold have been observed for DBv1-T7AC and DBv2-T7AC, respectively, compared to the T7AC-tag. For PTH, the DB-T7ACnew variants increased expression 2.39-fold and 2.01-fold with constructs DBv1-T7ACnew and DBv2-T7ACnew compared to T7ACnew. hFGF2 showed an increase in titre of 1.82-fold and 1.75-fold comparing DBv2-T7AC with T7AC and DBv2-T7ACnew with T7ACnew, respectively. Measured specific and volumetric recombinant protein titres are summarized in Table 5 above.

ΔG is the free folding energy. Lower ΔG values (lower free folding energy) mean more tightly folded mRNA, whereas higher ΔG values mean less folding. At a value of 0 kcal/mol, there is no folding present on the mRNA. Previously, it was described that a high free folding energy ($\Delta G_{\min}$) would aid in protein production and lead to higher titers, because of better accessibility of the mRNA for the ribosome. Tuller et al., for example, describe that there is a global selection for non-folding structures at the beginning of *E. coli* and *S. cerevisiae* genes (compared with the other parts

of the coding sequences) and suggested that this selection could act to allow faster binding of ribosomes to the transcript so as to initiate translation (Tuller et al., 2009). However, the construct which provided the highest titre for PTH, DBv1-T7ACnew, has a low, and thus poor, $\Delta G_{\min}$ value, showing that, contrary to previously published literature, the free folding energy is actually no predictor for tag performance.

The calculated $\Delta G_{\min}$ of DBv1-T7ACnew is -4.10 kcal/mol, while $\Delta G_{\min}$ for T7ACnew is -1.50 kcal/mol (see Example 2.1. and Fig. 2). Surprisingly, even though T7ACnew has a significantly higher free folding energy than DBv1-T7ACnew, the latter performed significantly better and provided a 2.39-fold higher titre for PTH than T7ACnew (see Fig. 3A and 3B and Fig. 4).

Example 3: Novel peptide tags enhance production of hFGF-2, PTH and TNF α

The POI production using peptide tags comprising the newly identified peptide sequences DBv1 (SEQ ID NO: 1) or DBv2 (SEQ ID NO: 2) of the present invention is compared to the POI production using the T7AC - tag (SEQ ID NO: 79) or the LED-PERNKERKE – tag (SEQ ID NO: 109) in high cell density fermentations.

3.1 Material and Methods

3.1.1 Tag sequences used for example 3

The T7AC tag and the N-terminal part of the T7AC-tag, LED-PERNKERKE (SEQ ID NO: 109), comprise the amino acids LED (SEQ ID NO: 139) at their N-termini. The influence of the peptides, DBv1 (PLV; SEQ ID NO: 1) or DBv2 (PHS; SEQ ID NO: 2) of the present invention on the titer of the POIs is tested by substituting LED (SEQ ID NO: 139) by the peptides DBv1 (PLV; SEQ ID NO: 1) or DBv2 (PHS; SEQ ID NO: 2) in the T7AC - and the LED-PERNKERKE – tags (SEQ ID NO: 79 and SEQ ID NO: 109, respectively), resulting in the following tags:

Table 6: amino acid sequences of the tags used for example 3 and nucleic acids encoding them.
See Fig. 1 and Fig. 9 for the respective sequences.

Tag	Amino acid sequence	Nucleic acid sequence
T7AC	SEQ ID NO: 79	SEQ ID NO: 53
DBv1-T7AC	SEQ ID NO: 11	SEQ ID NO: 13
DBv2-T7AC	SEQ ID NO: 12	SEQ ID NO: 17
PERNKERKE	SEQ ID NO: 107	SEQ ID NO: 108
PERNKERKE _{new}	SEQ ID NO: 107	SEQ ID NO: 135
LED-PERNKERKE	SEQ ID NO: 109	SEQ ID NO: 110
LED-PERNKERKE _{new}	SEQ ID NO: 109	SEQ ID NO: 136
PLV-PERNKERKE	SEQ ID NO: 111	SEQ ID NO: 112
PLV-PERNKERKE _{new}	SEQ ID NO: 111	SEQ ID NO: 137
PHS-PERNKERKE	SEQ ID NO: 113	SEQ ID NO: 114
PHS-PERNKERKE _{new}	SEQ ID NO: 113	SEQ ID NO: 138

3.1.2 POI fusion proteins

5 The different tag sequences were then fused to a 6-His-tag, GSG linker and a caspase 2 recognition and cleavage site (SEQ ID NO: 140), and further fused to three pharmaceutically relevant model proteins: TNF α , human fibroblast growth factor 2 (hFGF2) or parathyroid hormone (PTH).

10 3.1.2.1 POIs

Table 7: Amino acid and nucleic acid sequences of the tested POIs. See Fig. 1 and Fig. 9 for the respective sequences.

POI	Amino acid sequence	Nucleic acid sequence
hFGF2	SEQ ID NO: 66	SEQ ID NO: 67
PTH	SEQ ID NO: 55	SEQ ID NO: 56
TNF α	SEQ ID NO: 115	SEQ ID NO: 116

3.1.2.2 POI fusion proteins

Table 8: Amino acid and nucleic acid sequences of the tested POI fusion proteins. See Fig. 9 for the respective sequences.

Name of POI fusion protein	Amino acid sequence	Nucleic acid sequence
T7AC - TNF α	SEQ ID NO: 117	SEQ ID NO: 118
DBv1-T7AC - TNF α	SEQ ID NO: 119	SEQ ID NO: 120
DBv2-T7AC - TNF α	SEQ ID NO: 121	SEQ ID NO: 122
LED-PERNKERKE – hFGF-2	SEQ ID NO: 123	SEQ ID NO: 124
PLV-PERNKERKE – hFGF-2	SEQ ID NO: 125	SEQ ID NO: 126
PHS-PERNKERKE – hFGF-2	SEQ ID NO: 127	SEQ ID NO: 128
LED-PERNKERKE - PTH	SEQ ID NO: 129	SEQ ID NO: 130
PLV-PERNKERKE - PTH	SEQ ID NO: 131	SEQ ID NO: 132
PHS-PERNKERKE - PTH	SEQ ID NO: 133	SEQ ID NO: 134

5

3.1.3 Cell bank generation

The nucleic acids encoding the POI fusion proteins were synthesized and cloned into the respective expression vectors (pET30a-cer derivative). The correct sequences of all expression constructs were confirmed by sequencing. The expression vectors were transformed into *E. coli* BL21(DE3). A research cell bank was produced. The expression of the POI-fusions was under the control of the T7 promoter system and the tZenit terminator (Mairhofer 2014).

10

3.1.4 Fed batch fermentations (1 L)

All production strains were cultivated in a 1 L Multifors bioreactor with a working volume of 0.7 L and a starting batch volume of 0.4 L. The preculture was inoculated with research cell bank and performed in shake flask containing preculture medium at 37 °C. The batch medium in the 1 L bioreactor was inoculated with preculture in a ratio of 1:100, when the OD₅₅₀ in the preculture reached a value of 2. Temperature, pH (by addition of an ammonia solution), and dissolved oxygen (pO₂; by stirrer speed and increasing the concentration of O₂ in the in-air) in the bioreactor were controlled. The pO₂ was kept constant at 20 %, the aeration at 1 vvm and pressure at 1 bar. After an increase of pO₂ in the batch phase the glucose feed was started

20

(exponential feed for 9.5 hours and constant feed for 12 hours). Expression of the POI fusion protein was started 11.5 h after start of the exponential feed by the addition of 0.179 g IPTG (Isopropyl- β -D-thiogalactopyranoside). The end of fermentation was reached at 21.5 h after start of glucose feed. The production phase was 10 h.
5 Temperature was 37 °C in growth phase and 30 °C in production phase. pH was kept at 7.0 +/- 0.2.

3.1.5 Analytics

Samples of the fermentation broth were taken at 5 time points. The WCW [g/L]
10 (wet cell weight) was measured by weighing the pellet of 1 mL of culture broth after centrifugation.

3.1.5.1 Sample Preparation

The pellets from the fermentation broth samples were resuspended in Bug-
15 Buster™ from Novagen (including 1 μ l Lysonase™ from Novagen) by vortexing until the pellet has been completely resuspended and centrifuged after incubation for 15 minutes. The supernatant was used for the determination of the titer of the intracellular soluble POI fusion protein.

3.1.5.2 Determination of the volumetric titer of the POI fusion proteins

20 The titer was determined using SDS-Page. scFvM (a single chain antibody) was used as a standard for the calibration curve for quantification.

3.2 Results

25 Recombinant production of POI fusions (cytosolic, soluble titer) comprising the peptides PLV (DBv1; SEQ ID NO: 1) or PHS (DBv2; SEQ ID NO: 2) instead of LED (SEQ ID NO: 139) in high cell density fermentations was increased for all three POIs. For TNF α see Figure 6 and Table 9, for hFGF-2 see Figure 7 and Table 10 and for PTH see Figure 8 and Table 11.

30

Table 9: Expression of TNF α fusion proteins comprising PLV (DBv1) or PHS (DBv2) compared to the same comprising LED instead of PLV or PHS.

Time after induction [h]	T7AC - TNF α	DBv1-T7AC - TNF α	DBv2-T7AC - TNF α
0	n.d.	0	0
2.5	n.d.	3.8	3.6
5.5	n.d.	11.4	11.5
7.5	n.d.	18.9	19.2
10.0	12.9	26.0	28.1

n.d.: not detected

- 5 Table 10: Expression of hFGF-2 fusion proteins comprising PLV (DBv1) or PHS (DBv2) compared to the same comprising LED instead of PLV or PHS.

Time after induction [h]	LED-PERN-KERKE – hFGF-2	PLV-PERN-KERKE – hFGF-2	PHS-PERN-KERKE – hFGF-2
0	0	0	0
2.5	1.4	2.0	2.5
5.5	1.9	2.5	3.3
7.5	4.5	5.0	7.9
10.0	6.1	8.3	10.2

Table 11: Expression of PTH fusion proteins comprising PLV (DBv1) or PHS (DBv2) compared to the same comprising LED instead of PLV or PHS.

Time after induction [h]	LED-PERN-KERKE – PTH	PLV-PERN-KERKE – PTH	PHS-PERN-KERKE – PTH
0	0	0	0
2.5	1.5	2.0	3.6
5.5	4.8	6.1	10.7
7.5	11.6	12.8	18.8
10.0	12.4	14.6	18.7

REFERENCES

- Baeshen MN, Al-Hejin AM, Bora RS, Ahmed MM, Ramadan HA, Saini KS, Baeshen NA, Redwan EM: Production of Biopharmaceuticals in *E. coli*: Current Scenario and Future Perspectives. *J Microbiol Biotechnol* 2015, 25:953-962.
- 5 Cserjan-Puschmann, M., et al., Production of Circularly Permuted Caspase-2 for Affinity Fusion-Tag Removal: Cloning, Expression in *Escherichia coli*, Purification, and Characterization. *Biomolecules*, 2020. **10**(12).
- Fink, M., et al., High-throughput microbioreactor provides a capable tool for early stage bioprocess development. *Scientific Reports*, 2021. **11**(1): p. 2056.
- 10 Fink, M., et al., Integrated process development: The key to improve Fab production in *E. coli*. *Biotechnology Journal*, 2021. **16**(6): p. 2000562.
- Fink, M., et al., Microbioreactor Cultivations of Fab-Producing *Escherichia coli* Reveal Genome-Integrated Systems as Suitable for Prospective Studies on Direct Fab Expression Effects. *Biotechnology Journal*, 2019. **14**(11): p. 1800637.
- 15 Köppl, C., et al., Fusion Tag Design Influences Soluble Recombinant Protein Production in *Escherichia coli*. *Int J Mol Sci*, 2022. **23**(14).
- Kröß C, Engele P, Sprenger B, Fischer A, Lingg N, Baier M, Öhlknecht C, Lier B, Oostenbrink C, Cserjan-Puschmann M, et al: PROFICS: A bacterial selection system for directed evolution of proteases. *Journal of Biological Chemistry* 2021. **297**: p.101095.
- 20 Lingg N, Kröß C, Engele P, Öhlknecht C, Köppl C, Fischer A, Lier B, Loibl J, Sprenger B, Liu J, et al: CASPON platform technology: Ultrafast circularly permuted caspase-2 cleaves tagged fusion proteins before all 20 natural amino acids at the N-terminus. *New Biotechnology* 2022. **71**:37-46.
- 25 Mairhofer Juergen et al.: Preventing T7 RNA Polymerase Read-through Transcription - A Synthetic Termination Signal Capable of Improving Bioprocess Stability. *ACS Synthetic Biology*. 2014, 4/3: 256-273
- 30 McElwain L, Phair K, Kealey C, Brady D. Current trends in biopharmaceuticals production in *Escherichia coli*. *Biotechnol Lett*. 2022;44(8):917-931. doi:10.1007/s10529-022-03276-5

- Moll I, Huber M, Grill S, Sairafi P, Mueller F, Brimacombe R, Londei P, Bläsi U. Evidence against an Interaction between the mRNA downstream box and 16S rRNA in translation initiation. *J Bacteriol.* 2001. **183**(11): 3499-505.
- O'Connor M, Asai T, Squires CL, Dahlberg AE. Enhancement of translation by the downstream box does not involve base pairing of mRNA with the penultimate stem sequence of 16S rRNA. *Proc Natl Acad Sci U S A.* 1999. **96**(16):8973-8.
- Richter LV, Yang H, Yazdani M, Hanson MR, Ahner BA. A downstream box fusion allows stable accumulation of a bacterial cellulase in *Chlamydomonas reinhardtii* chloroplasts. *Biotechnol Biofuels.* 2018. **11**: p. 133.
- Sprengart, M.L., E. Fuchs, and A.G. Porter, The downstream box: an efficient and independent translation initiation signal in *Escherichia coli*. *Embo j*, 1996. **15**(3): p. 665-74.
- Stargardt, P., et al., Bacteriophage Inspired Growth-Decoupled Recombinant Protein Production in *Escherichia coli*. *ACS Synthetic Biology*, 2020. **9**(6):1336-1348.
- Tuller T. et al., Translation Efficiency is Determined by Both Codon Bias and Folding Energy. *PNAS.* 2010. **107**(8): 3645-3650.

CLAIMS

1. A peptide tag for expression of a recombinant protein of interest (POI) in a host cell, comprising a first peptide sequence consisting of an amino acid sequence selected from SEQ ID NO:1 or SEQ ID NO:2.
2. The peptide tag of claim 1, comprising a second peptide sequence comprising an amino acid sequence selected from SEQ ID NO:5, SEQ ID NO:10 or SEQ ID NO:21, preferably wherein said second peptide sequence is C-terminal of the first peptide sequence.
3. The peptide tag of claim 2, wherein the tag comprises an amino acid sequence of SEQ ID NO:84, wherein
 - a. X at position 2 is L or H;
 - b. X at position 3 is V or S;
 - c. X at position 16 is E or Q; and
 - d. X at position 18 is E or Q.
4. The peptide tag of claim 1 or 2, wherein the first peptide sequence is directly linked to the second peptide sequence.
5. The peptide tag of any one of claims 1 to 4, further comprising one or more linker sequences.
6. The peptide tag of any one of claims 1 to 5, further comprising a solubility and/or expression enhancement tag, a monitoring tag and/or an affinity tag.
7. The peptide tag of any one of claims 1 to 6, further comprising a protease recognition and/or cleavage site, preferably at its C-terminus.
8. The peptide tag of any one of claims 1 to 7, comprising an amino acid sequence selected from the group consisting of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:32, SEQ ID NO:35, SEQ ID NO:42 and SEQ ID NO:47, or a functionally active variant thereof comprising an amino acid sequence having at least 95% sequence identity to SEQ ID NO:10, SEQ ID NO:21, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87 or SEQ ID NO:88 and comprising SEQ ID NO:1 or SEQ ID NO:2, preferably at its N-terminus.
9. An expression cassette for expression of a recombinant protein of interest (POI) as a fusion protein comprising the peptide tag of any one of claims 1 to

8 fused to one terminus of said POI, wherein said expression cassette comprises:

- a. a promoter;
- b. a ribosome binding site (RBS);
- 5 c. a start codon;
- d. a sequence encoding the peptide tag of any one of claims 1 to 8, preferably wherein said sequence encoding the peptide tag is placed immediately after the start codon; and
- 10 e. a sequence encoding the POI, preferably wherein said sequence encoding the POI is placed immediately after the sequence encoding the peptide tag.

10. An expression vector comprising the expression cassette of claim 9.

11. A fusion protein comprising a recombinant protein of interest (POI) and the peptide tag of any one of claims 1 to 8, wherein the peptide tag of any one of claims 1 to 8 is fused to the recombinant POI, preferably at the N-terminus of the recombinant protein.

12. A method of producing a recombinant protein of interest (POI) comprising expressing the POI in the form of a fusion protein comprising the peptide tag of any one of claims 1 to 8 by a prokaryotic or eukaryotic expression system, wherein said fusion protein is expressed from an expression vector within a host cell or from an expression cassette stably integrated into the genome or chromosome of a host cell.

13. The method of claim 12, comprising culturing the host cell for a period of time under conditions permitting expression of the POI in the form of a fusion protein and comprising any one or more of the steps of:

- a. recovering the fusion protein and/or POI;
- b. purifying the fusion protein and/or POI;
- c. removing the peptide tag;
- d. further purifying the fusion protein and/or POI;
- 30 e. formulating the fusion protein and/or POI;
- f. modifying the fusion protein and/or POI.

14. A nucleic acid sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:11, SEQ ID

NO:12, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:32, SEQ ID NO:35, SEQ ID NO:42 and SEQ ID NO:47, or a functionally active variant thereof comprising an amino acid sequence having at least 95% sequence identity to SEQ ID NO:10, SEQ ID NO:21, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87 or SEQ ID NO:88 and comprising SEQ ID NO:1 or SEQ ID NO:2, preferably at its N-terminus.

15. A host cell comprising the peptide tag of any one of claims 1 to 8 and/or a nucleic acid sequence encoding said peptide tag.

10