to be degraded or to induce toxicity of the expression system by accumulating in the intracellular environment of the expression system. Thus, an antibody may be expressed as a fusion protein comprising a preferred N-terminal sequence fused to a mature sequence of the heavy chain or the light chain of an antibody.

[0006] To effectively use this technique, the signal peptides must be identified. However, the amino acid sequence of most recombinant therapeutic antibodies disclosed in publicly available databases does not include the sequences of the signal peptides. Furthermore, the signal peptide information required for production in mammalian cell culture is also absent in the case of monoclonal antibodies isolated through methods such as phage display.

[0007] Rituxan is the only antibody for which heavy and light chain signal peptide sequence information is available in public database. However, the availability of the information does not entail that the original signal peptide is optimized for the secretion of the antibody in the desired host cell selected for production.

[0008] Another issue associated with production of recombinant antibodies comprising a signal peptide is the cleavage heterogeneity. Cleavage heterogeneity may arise from nonspecific cleavage of a signal peptide by a signal peptidase. As cleavage of the signal peptide occurs within the variable region in the N-terminus of both the heavy and light chains, non-specific cleavage may affect antigen recognition.

[0009] Recombinant IgG production may also be affected by glycan heterogeneity that is present at the *N*-glycosylation site of the C_H2 constant domain of the heavy chains. One recent study explored an LC/MS with a column-switching system to rapidly evaluate glycan heterogeneity.

[0010] Thus, there is a need to provide further signal peptides to create recombinant antibodies that are efficiently secreted and produced in expression systems.

SUMMARY OF THE INVENTION

[0011] Described below are antibodies of interest with a light chain and heavy chain fused to signal peptides to enhance secretion of the antibody during production and methods of determining a signal peptide for enhancing secretion of an antibody of interest.

Thus, in one aspect, there is provided an amino acid sequence of interest comprising a polypeptide that corresponds to a heavy chain of an antibody of interest, and a

heavy chain signal peptide, wherein the heavy chain signal peptide comprises a N-domain, a H-domain and a C-domain, wherein the heavy chain signal peptide is bound to the polypeptide via the C-terminal end of the C-domain of the heavy chain signal peptide; wherein the heavy chain signal peptide comprises an amino acid sequence selected from the group consisting of MELGLSWIFLLAILKGVQC (H1, SEQ ID NO: 5), MELGLRWVFLVAILEGVQC (H2, SEQ ID NO: 6), MDWTWRFLFVVAAATGVQS (H5, SEQ ID NO: 1), MEFGLSWLFLVAILKGVQC (H6, SEQ ID NO: 9), MEFGLSWVFLVALFRGVQC (H7, SEQ ID NO:3) and MDLLHKNMKHLWFFLLLVAAPRWVLS (H8, SEQ ID NO: 11), and wherein the antibody of interest is selected from the group consisting of trastuzumab (Herceptin, CAS number 180288-69-1), bevacizumab (Avastin, CAS number 216974-75-3), infliximab (Remicade, CAS number 170277-31-3), rituximab (Rituxan, CAS number 174722-31-7) and adalimumab (Humira, CAS number 331731-18-1).

[0012] In one example, there is provided an antibody of interest comprising a heavy chain signal peptide, wherein the heavy chain signal peptide comprises a N-domain, a H-domain and a C-domain, wherein the heavy chain signal peptide is bound to the antibody via

the C-terminal end of the C-domain; wherein the H-domain comprises an amino acid sequence of formula (I):

$$X_{aa6}\text{-}X_{aa7}\text{-}X_{aa8}\text{-}X_{aa9}\text{-}X_{aa10}\text{-}X_{aa11}\text{-}X_{aa12}\text{-}X_{aa13},$$

wherein X_{aa6} is any one of Trp, Phe or Ile; X_{aa7} is any one of Val, Ile or Leu; X_{aa8} is Phe or Leu; X_{aa9} is Leu or Val; X_{aa10} is Val or Leu; X_{aa11} is Ala; X_{aa12} is any one of Leu, Ile or Ala; X_{aa13} is any one of Phe, Leu, Pro or Ala; and wherein the amino acid sequence of formula (I) has optionally one or two mutations; provided that an antibody showing the lowest secretion is excluded, wherein the lowest secretion is determined by a) fusing polynucleotides encoding the heavy chain of the antibody of interest to a polynucleotide encoding the heavy chain signal peptide as defined herein, to obtain different combinations of heavy chain signal peptides and antibody of interest; b) creating at least one expression vector comprising the polynucleotides defined under a); and c) quantifying secretion of the antibody of interest encoded by the vector referred to under b) after transfection into an expression system to determine the antibody with the lowest secretion.

[0013] In another example, there is provided an antibody of interest comprising a light chain signal peptide, wherein the light chain signal peptide comprises an amino acid sequence of formula (IV):

$$X_{bb6}\text{-}X_{bb7}\text{-}X_{bb8}\text{-}X_{bb9}\text{-}X_{bb10}\text{-}X_{bb11}\text{-}X_{bb12}\text{-}X_{bb13}\text{-}X_{bb14},$$

wherein X_{bb6} is Ala or Thr; X_{bb7} is Gln or Ala; X_{bb8} is Leu or Ala; X_{bb9} is Leu or Ala; X_{bb10} is Gly; X_{bb11} to X_{bb14} is Leu; or wherein the amino acid sequence of the formula (IV) has one or two mutations; provided that an antibody showing the lowest secretion is excluded, wherein the lowest secretion is determined by a) fusing polynucleotides encoding the light chain of the antibody of interest to a polynucleotide encoding the light chain signal peptide as defined herein to obtain different combinations of light chain signal peptides and antibody of interest; b) creating at least one expression vector comprising the polynucleotides defined under a); and c) quantifying secretion of the antibody of interest encoded by the vector referred to under b) after transfection into an expression system to determine the antibody with the lowest secretion.

[0014] In another aspect, there is provided a polynucleotide encoding the amino acid sequence as described herein.

[0015] In a further aspect, there is provided a vector comprising the polynucleotide as defined above.

[0024] In one aspect, there is provided a polynucleotide sequence encoding the amino acid sequence as described herein.

[0025]

In another aspect, there is provided a vector comprising a polynucleotide sequence as described herein. In a further aspect, there is provided a method of determining a signal peptide for enhancing secretion of an antibody of interest, wherein the method comprises:

(i) fusing

(a) a polynucleotide encoding the heavy chain of the antibody of interest to a polynucleotide encoding the amino acid sequence selected from the group consisting of MELGLSWIFLLAILKGVQC (H1, SEQ ID NO: 5), MELGLRWVFLVAILEGVQC (H2, SEQ ID NO: 6), MKHLWFFLLLVAAPRWVLS (H3, SEQ ID NO: 7), MDWTWRFLFVVAAATGVQS (H5, SEQ ID NO: 1), MEFGLSWLFLVAILKGVQC (H6, SEQ ID NO: 9), MEFGLSWVFLVALFRGVQC (H7, SEQ ID NO:3) and MDLLHKNMKHLWFFLLLVAAPRWVLS (H8, SEQ ID NO: 11), and

(b) a polynucleotide encoding the light chain of the antibody of interest to a polynucleotide encoding the amino acid sequence MDMRVPAQLLGLLLLWLSGARC (L1, SEQ ID NO: 2) or MKYLLPTAAAGLLLLAAQPAMA (L2, SEQ ID NO: 4),

(ii) creating (a) at least one vector comprising a polynucleotide resulted from (i)(a) and (b) at least one vector comprising a polynucleotide resulted from (i)(b); and

(iii) quantifying and comparing secretion of the antibody after transfection of the at least one vector from (ii)(a) and the at least one vector from (ii)(b) into an expression system to determine the combination of heavy and light chain signal peptides that enhance expression and secretion of the antibody of interest the most.

BRIEF DESCRIPTION OF THE DRAWINGS

[0026] The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings, in which:

[0027] **Fig. 1** is a hierarchical clustering dendogram of signal peptides using agglomerative nesting to search for sequence similarity in (A) heavy chain signal peptides from a database of 173 human Ig heavy chains and (B) kappa (κ) light chain signal peptides from a database of 62

Claims

1. An amino acid sequence of interest comprising a polypeptide that corresponds to a heavy chain of an antibody of interest, and a heavy chain signal peptide, wherein the heavy chain signal peptide comprises a N-domain, a H-domain and a C-domain, wherein the heavy chain signal peptide is bound to the polypeptide via the C-terminal end of the C-domain of the heavy chain signal peptide; wherein the heavy chain signal peptide comprises an amino acid sequence selected from the group consisting of MELGLSWIFLLAILKGVQC (H1, SEQ ID NO: 5), MELGLRWVFLVAILEGVQC (H2, SEQ ID NO: 6), MDWTWRFLFVVAAATGVQS (H5, SEQ ID NO: 1), MEFGLSWLFLVAILKGVQC (H6, SEQ ID NO: 9), MEFGLSWVFLVALFRGVQC (H7, SEQ ID NO:3) and MDLLHKNMKHLWFFLLLVAAPRWVLS (H8, SEQ ID NO: 11), and wherein the antibody of interest is selected from the group consisting of trastuzumab (Herceptin, CAS number 180288-69-1), bevacizumab (Avastin, CAS number 216974-75-3), infliximab (Remicade, CAS number 170277-31-3), rituximab (Rituxan, CAS number 174722-31-7) and adalimumab (Humira, CAS number 331731-18-1).

2. The amino acid sequence according to claim 1, wherein the heavy chain signal peptide is selected from the group consisting of: MDWTWRFLFVVAAATGVQS (H5, SEQ ID NO: 1), and MEFGLSWVFLVALFRGVQC (H7, SEQ ID NO: 3).

3. A polynucleotide encoding the amino acid sequence of claim 1 or 2.

4. A vector comprising the polynucleotide according to claim 3.

5. A host cell comprising the polynucleotide of claim 3 and/or the vector of claim 4.

6. An amino acid sequence comprising

a) the amino acid sequence for trastuzumab (Herceptin, CAS number 180288-69-1);
b) a heavy chain signal peptide of SEQ ID NO: 1 fused to the heavy chain amino acid sequence of trastuzumab; and

c) a light chain signal peptide of SEQ ID NO: 2 fused to the light chain amino acid sequence of trastuzumab.

7. An amino acid sequence comprising

a) the amino acid sequence for bevacizumab (Avastin, CAS number 216974-75-3);

b) a heavy chain signal peptide of SEQ ID NO: 3 fused to the heavy chain amino acid sequence of bevacizumab; and

c) a light chain signal peptide of SEQ ID NO: 2 fused to the light chain amino acid sequence of bevacizumab.

8. An amino acid sequence comprising

a) the amino acid sequence for infliximab (Remicade, CAS number 170277-31-3);

b) a heavy chain signal peptide of SEQ ID NO: 3 fused to the heavy chain amino acid sequence of infliximab; and

c) a light chain signal peptide of SEQ ID NO: 4 fused to the light chain amino acid sequence of infliximab.

9. An amino acid sequence comprising

a) the amino acid sequence for rituximab (Rituxan, CAS number 174722-31-7);

b) a heavy chain signal peptide of SEQ ID NO: 3 fused to the heavy chain amino acid sequence of rituximab; and

c) a light chain signal peptide of SEQ ID NO: 4 fused to the light chain amino acid sequence of rituximab.

10. An amino acid sequence comprising

a) the amino acid sequence for adalimumab (Humira, CAS number 331731-18-1);

b) a heavy chain signal peptide of SEQ ID NO: 3 fused to the heavy chain amino acid sequence of adalimumab; and

c) a light chain signal peptide of SEQ ID NO: 2 fused to the light chain amino acid sequence of adalimumab.

11. A polynucleotide sequence encoding the amino acid sequence of any one of claims 6 to 10.

12. A vector comprising a polynucleotide sequence according to claim 11.

13. The amino acid sequence of claim 1 or 2, wherein the expression system is a host cell.

14. The amino acid sequence of claim 1 or 2, wherein the host cell is a mammalian or bacterial cell.

15. The amino acid sequence of claim 14, wherein the expression system is selected from the group consisting of Chinese Hamster Ovary Cell (CHO cell), baculovirus expression vector system, *Leishmania tarentolae* expression system, galline cell, yeast expression system, and plant expression system.

16. A method of determining a signal peptide for enhancing secretion of an antibody of interest, wherein the method comprises:

(i) fusing

(a) a polynucleotide encoding the heavy chain of the antibody of interest to a polynucleotide encoding the amino acid sequence selected from the group consisting of MELGLSWIFLLAILKGVQC (H1, SEQ ID NO: 5), MELGLRWVFLVAILEGVQC (H2, SEQ ID NO: 6), MKHLWFFLLLVAAPRWVLS (H3, SEQ ID NO: 7), MDWTWRFLFVVAAATGVQS (H5, SEQ ID NO: 1), MEFGLSWLFLVAILKGVQC (H6, SEQ ID NO: 9), MEFGLSWVFLVALFRGVQC (H7, SEQ ID NO:3) and MDLLHKNMKHLWFFLLLVAAPRWVLS (H8, SEQ ID NO: 11), and

(b) a polynucleotide encoding the light chain of the antibody of interest to a polynucleotide encoding the amino acid sequence MDMRVPAQLLGLLLLWLSGARC (L1, SEQ ID NO: 2) or MKYLLPTAAAGLLLLAAQPAMA (L2, SEQ ID NO: 4),

(ii) creating (a) at least one vector comprising a polynucleotide resulted from (i)(a) and (b) at least one vector comprising a polynucleotide resulted from (i)(b); and

(iii) quantifying and comparing secretion of the antibody after transfection of the at least one vector from (ii)(a) and the at least one vector from (ii)(b) into an expression system to determine the combination of heavy and light chain signal peptides that enhance expression and secretion of the antibody of interest the most.

17. The method of claim 16, wherein the expression system is a host cell.

18. The method of claim 17, wherein the host cell is a eukaryotic cell or a prokaryotic cell.

19. The method of claim 18, wherein the eukaryotic cell is selected from the group consisting of a mammalian cell, an avian cell, an insect cell, a fungal cell, a yeast cell or a plant cell.

20. The method of claim 16, wherein the expression system is selected from the group consisting of Chinese Hamster Ovary Cell (CHO cell), baculovirus expression vector system, *Leishmania tarentolae* expression system, galline cell, yeast expression system, and plant expression system.

21. The method of any one of claims 16 to 20, wherein the antibody is a monoclonal antibody.

22. The method of any one of claims 16 to 21, wherein the antibody has an isotype selected from the group consisting of mammalian IgG, IgM, IgD, IgA and IgE.