OPTIMISED HEAVY CHAIN AND LIGHT CHAIN SIGNAL PEPTIDES FOR THE PRODUCTION OF RECOMBINANT ANTIBODY THERAPEUTICS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority of SG provisional application No. 201207629-5, filed October 12, 2012, the contents of it being hereby incorporated by reference in its entirety for all purposes.

FIELD OF THE INVENTION

[0002] The invention relates to the field of biotechnology and molecular biology. The invention relates to an antibody of interest comprising an optimised selection of signal peptides and their use for enhanced antibody secretion from cells. In particular, embodiments of the present invention relate to methods of determining signal peptide for enhancing secretion of an antibody as well as other related antibodies, compositions and methods.

BACKGROUND OF THE INVENTION

[0003] Recombinant protein expression represents a large portion of the production of proteins used in molecular biology, agronomy, veterinary science or medicine. For example, recombinant therapeutic antibodies represent a large percentage of biopharmaceuticals (i.e. medical drugs produced using biotechnology) produced.

[0004] Protein signal sequences, also called topogenic signals or signal peptides, play a central role in the targeting and translocation of nearly all secreted proteins and many integral membrane proteins in both prokaryotes and eukaryotes. The signal peptides from various proteins generally consist of three structurally, and possibly functionally distinct, regions: (1) an amino terminal (N-terminal) positively charged N-region (or N-domain), (2) a central hydrophobic H-region (or H-domain), and (3) a neutral but polar carboxy terminal (C-region or C-domain).

[0005] The determination of protein signal sequences is important to improve excretion and secretion of antibodies produced using an expression system such as bacteria, plants, and animals to produce effective drugs (especially therapeutic proteins). By adding a specific tag to an antibody of interest, it is possible to improve the antibody excretion or secretion in the extracellular environment. In this manner, the antibody is easier to harvest and is less likely

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.

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, a heavy chain signal peptide, a polypeptide that corresponds to a light chain of the antibody of interest, and a light 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), wherein the light chain signal peptide is bound to the polypeptide that corresponds to a light chain of the antibody of interest, and wherein the light chain signal peptide comprises an amino acid sequence selected from the group consisting of MDMRVPAQLLGLLLLWLSGARC (L1, SEQ ID NO: 2) and MKYLLPTAAAGLLLLAAQPAMA (L2, SEQ ID NO: 4), 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).

In one example, 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.

[0016] In a further aspect there is provided a host cell comprising the nucleotide as described hereinbefore and/or the vector as described hereinbefore.

[0017] In one example, there is provided a hybridoma cell line capable of producing the antibody as described above.

[0018] In one aspect, there is provided an amino acid sequence comprising the amino acid sequence for trastuzumab (Herceptin, CAS number 180288-69-1); a heavy chain signal peptide of SEQ ID NO: 1 fused to the heavy chain amino acid sequence of trastuzumab; and a light chain signal peptide of SEQ ID NO: 2 fused to the light chain amino acid sequence of trastuzumab.

[0019] In another aspect, there is provided an amino acid sequence comprising the amino acid sequence for bevacizumab (Avastin, CAS number 216974-75-3); a heavy chain signal peptide of SEQ ID NO: 3 fused to the heavy chain amino acid sequence of bevacizumab; and a light chain signal peptide of SEQ ID NO: 5 fused to the light chain amino acid sequence of bevacizumab.

[0020] In yet another aspect, there is provided an amino acid sequence comprising the amino acid sequence for infliximab (Remicade, CAS number 170277-31-3); a heavy chain signal peptide of SEQ ID NO: 3 fused to the heavy chain amino acid sequence of infliximab; and a light chain signal peptide of SEQ ID NO: 4 fused to the light chain amino acid sequence of infliximab.

[0021] In a further aspect, there is provided an amino acid sequence comprising the amino acid sequence for rituximab (Rituxan, CAS number 174722-31-7); a heavy chain signal peptide of SEQ ID NO: 3 fused to the heavy chain amino acid sequence of rituximab; and a light chain signal peptide of SEQ ID NO: 4 fused to the light chain amino acid sequence of rituximab.

[0022] In yet a further aspect, there is provided an amino acid sequence comprising the amino acid sequence for adalimumab (Humira, CAS number 331731-18-1); a heavy chain signal peptide of SEQ ID NO: 3 fused to the heavy chain amino acid sequence of adalimumab; and a light chain signal peptide of SEQ ID NO: 5 fused to the light chain amino acid sequence of adalimumab.

[0023] In one example, there is provided a nucleotide encoding an antibody as described 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), 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

human kappa chains. Fig. 1 C to F is an enlargement of the dendogram obtained in Fig. 1 A showing the amino acid sequence of the signal peptides present in each of the created clusters H1 to H8, representative of the Ig heavy chain signal peptides. The selected amino acid sequence of SEQ ID: 1, 3, and 5 to 10 is highlighted for each one of the clusters H1 to H8. Fig. 1 G is an enlargement of the dendogram shown in Fig. 1 B showing the specific amino acid sequences of each of the 57 human kappa chains that were used to create the database. The specific amino acid sequence of the two clusters, L1 and L2, having the amino acid sequence of SEQ ID NO: 2 and 4, are highlighted.

[0028] **Fig. 2** is a table providing the amino acid sequences and DNA sequences of the heavy and light chain signal peptides that were selected for testing of the antibodies' improved production and secretion. The amino acid sequences correspond to the amino acid sequence already highlighted in Fig. 1 C to G.

Fig. 3 is a series of histogram plots showing the amount of Herceptin heavy chain produced when fused with one of the eight heavy chain signal peptides selected from H1 to H8, as described above, and when co-expressed with the Herceptin kappa light chain fused either to L1 or L2 signal peptide described above. Fig. 3 A and C: Ig ELISA results were plotted showing the monoclonal antibody titer (in ng/ml) with eight different heavy chain signal peptides and two different light chain signal peptides. Fig. 3A shows results when heavy chains were paired with light chain signal peptide 1 (L1), while Fig. 3C shows results when heavy chains were paired with light chain signal peptide 2 (L2). Fig. 3 B and D: The

efficiency of the heavy chain signal peptides in each antibody is shown relative to heavy chain signal peptide 1 (H1), which is set to 1. Fig. 3B shows results when heavy chains were paired with light chain signal peptide 1 (L1), while Fig. 3D shows results when heavy chains were paired with light chain signal peptide 2 (L2). Mean fluorescence intensity of GFP was used to normalize variations due to difference in transfection efficiency. Herceptin is efficiently secreted with several different signal peptides. The combination that had the highest secretion when normalized as indicated herein is H5/L1, whereas H7 exhibited the best titre for Herceptin when paired with L2. The values were obtained following normalization with the GFP values as shown in Figure 3A and C to account for transfection efficiency variations among the different experiments.

[0030] **Fig. 4** is a series of histogram plots obtained as in Fig. 3. Avastin heavy and light chain where used instead of Herceptin. Fig. 4A and C: Ig ELISA results. Fig 4B and D: relative productivity normalized to H1. The best signal peptides combination for Avastin is H7/L1.

[0031] **Fig. 5** is a series of histogram plots showing the amount of Remicade heavy chain produced when different signal peptides were used as in Fig. 3 and 4 above. Light chain with signal peptide 1 was used in Fig. 5 A and B; and light chain with signal peptide 2 was used in Fig. 5 C and D; A and C: Ig ELISA results. Fig. 5 B and D: relative productivity normalized to H1. The heavy chain signal peptide that gave the highest secretion and expression rate when fused to the heavy chain of Remicade is H7 either when used in combination when L1 or L2 light chain signal peptide fused to the light chain of remicade.

[0032] **Fig. 6** is a series of histogram plots showing the amount of Rituxan heavy chain produced when different signal peptides were used as indicated above. A and C: ELISA results. Bottom panel: relative productivity normalized to H1. Fig. 6 A and B, light chain with signal peptide 1 was used. Fig. 6 C and D light chain with signal peptide 2 was used. The best signal peptide combinations for Rituxan are the heavy chain signal peptide H7 and either L1 or L2.

[0033] **Fig. 7** is a series of histogram plots showing the amount of Humira heavy chain produced when different signal peptides were used. Fig. 7A and C: Ig ELISA results. Bottom panel: relative productivity normalized to H1. A and B, light chain with signal peptide 1 was used. Fig. 7C and D, light chain with signal peptide 2 was used. Best signal peptides combination for Humira is H7/L1.

[0034] **Fig. 8** is a series of histogram plots comparing the efficiency of secretions according to the light chain L1 or L2 that was used, for each of the indicated five therapeutic antibodies. Based on the results obtained in Figures 3 to 7, the heavy chain of the antibody of interest with a heavy chain signal peptide combination with the highest secretion was selected. The selected construct was co-expressed in CHO cells with the light chain signal peptide L1 or L2 fused with the light chain of said antibody of interest. The titers of the produced antibody were measured by IgG ELISA and the values thus obtained compared. The comparison allows the selection of the pair of heavy chain/signal peptide and corresponding light chain/signal peptide that gives the best production/secretion for the antibody of interest.

[0035] For Herceptin, the best pair of signal peptides was H5/L1. For Avastin and Humira, the best pair was H7/L1 while for Remicade and Rituxan the best pair was H7/L2. Mean fluorescence intensity of GFP was used to normalize the transfection efficiency.

[0036] **Fig. 9** is a bar graph representative of the efficiency of the optimized signal peptides in the secretion of Rituxan. Fig. 9 is a comparison between the original (that is the signal peptide present in public database) and the optimized signal peptide pairs for Rituxan, relative to the control pair H1/L2, which is set to 1. The relative efficiency of the optimized signal peptides for Rituxan (H7/L2) are more than twofold that of the original signal peptides.

[0037] **Fig. 10** represents a sequence alignment depicting the consensus sequence between the heavy chain signal peptides selected from the clusters H1 to H8, as described herein, (Fig. 10A) and the amino acid substitution of H7 signal peptide of the Avastin heavy chain (Fig. 10B). The N-domain, H-domain and C-domain are indicated below the consensus sequence and the number of amino acid within each domain is given above. Amino acid substitution mutations were performed on the H7 signal peptide to monitor the impact of such mutation on the secretion of Avastin heavy chain. For Avastin heavy chain, H7 is the signal peptide that provides the highest secretion efficiency. (A): A sequence comparison of all heavy chain signal peptide with H7 reveals several highly conserved amino acid residues: M***W**LFLVAA*GVQS. (B): Six hybrid constructs (H7a to H7f) were generated, each with amino acid substitution(s) within the conserved sites to study their impact on Avastin secretion. (C): Fig. 10C is a histogram plot demonstrating the impact of amino acid substitutions in the H7 signal peptide fused to the heavy chain of Avastin on the secretion of the recombinant heavy chain. Only C to S mutation (H7d) did not significantly affect the

expression level of Avastin. All other mutations resulted in decreased secretion of Avastin. H1 was included in the experiment as a control to normalize the results.

[0038] **Fig. 11** depicts a mass spectrum representing the relative abundance versus the mass-to-charge ratio plot of Herceptin produced using recombinant signal peptide-heavy and light chain Herceptin antibodies. MS^2 fragment ions were generated from Herceptin using the LTQ Orbitrap Velos mass spectrometer. The peaks corresponding to the heavy chain and light chain N-terminal fragments, are indicated by the arrows. The amino acid sequences of the fragments are in bold type.

[0039] **Fig. 12** represents a mass spectrum of recombinant Avastin. MS^2 fragment ions were generated from Avastin using the LTQ Orbitrap Velos mass spectrometer. The peaks corresponding to the heavy chain and light chain N-terminal fragments, are indicated by the arrows. The amino acid sequences of the fragments are in bold type. Extracted ion chromatogram (XIC) peaks and MS/MS fragmentation spectra matching to b- and y-ions series of (A) heavy chain N-terminal peptide, (B) erroneously processed heavy chain N-terminal peptide resulting from an alternative cleavage site 5 residues downstream of expected cleavage site, and (C) light chain N-terminal peptide. Mass/charge ratio (m/z) of each N-terminal peptide, peak area (PA) and its percentage representation relative to total N-terminal peptides detected for corresponding heavy or light chain polypeptides are shown in the XIC spectra.

[0040] **Fig. 13** is a representative diagram summarizing the method of determining the optimal signal peptide for enhancing antibody secretion.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

[0041] Before the present methods, antibodies, nucleotides, amino acid sequences and uses thereof are described, it is to be understood that this invention is not limited to particular methods, antibodies, nucleotides, amino acid sequences, uses and experimental conditions described, as such methods, antibodies, nucleotides, amino acid sequences, uses and conditions may vary. It is also to be understood that the terminology used herein is for purposes of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

[0042] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art, to which this

invention belongs. Any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, as it will be understood that modifications and variations are encompassed within the spirit and scope of the instant disclosure.

[0043] The mechanism by which secreted proteins are translocated into the lumen of endoplasmic reticulum (ER, in eukaryotes) or through the plasma membrane (in prokaryotes) is universally conserved. In eukaryotes, this process occurs in the cytosol and involves recognition of the signal peptide in the nascent polypeptide chain by the signal recognition particle (SRP) as it emerges from the ribosome (ribosome-nascent chain, or RNC), causing elongation arrest. This SRP-RNC complex then binds to the membrane-anchored SRP-receptor (SR) in the ER, where GTP-dependent mechanism delivers the RNC to a membrane-bound Sec61 translocon which allows translocation of the growing polypeptide chain into the lumen of the ER. After crossing the translocon, the signal peptide is cleaved off by a signal peptidase.

[0044] Almost all secreted proteins have a signal peptide sequence, generally consisting of 20 to 30 amino acid residues beginning within 10 residues of the N-terminus, with a 10- to 15-residue long hydrophobic core. A hydrophilic stretch precedes this sequence, while a more polar stretch follows it up to the cleavage site. Signal sequence preferentially though weakly, binds in an α-helical confirmation to the hydrophobic groove in SRP in a 4-4 'ridges-into-grooves' packing, with the ribosome reinforcing this complex formation. The hydrophobic residues form a binding site that is critical for its interaction with the SRP, the translocon and the signal peptidase. Therefore, the exact sequence of the signal peptide may affect the efficiency of a protein to cross the ER membrane. Individual signal peptide sequences show little discernable primary sequence similarities and sequence variations can be considerable. As a result, it is challenging to anticipate, with any assurance, which amino acid sequence can serve as a signal peptide to achieve enhanced expression and secretion of recombinant antibodies in either prokaryotes or eukaryotes.

[0045] Recombinant therapeutic antibodies represent a large percentage of biopharmaceuticals produced. Advances in the manufacturing of such antibodies have been documented and are well-known in the art. The heavy chain of human IgG usually has a 19-amino acid signal peptide while the human kappa light chain contains a 22-amino acid signal peptide.

[0046] As described above, the expression of recombinant antibodies using expression systems requires the identification of optimized signal peptides to enhance efficacy of secretion of the heavy and light chain of the antibody.

[0047] Accordingly, it is an object of the present invention to identify and determine secretory signal peptides that enhance secretion and/or excretion of the heavy and light chain of antibodies when fused to their respective sequences. The present inventors have successfully designed a method to determine such optimized signal peptides in a protein expression system, while maintaining the activity and specificity of the antibody by providing homogenous cleavage at the proper cleavage site.

[0048] Therefore, the present invention provides a method of determining a signal peptide for enhancing secretion of an antibody of interest. The method may comprise the step of creating a database using publicly available amino acid sequence information of signal peptides associated with the heavy chain and the light chain of antibodies.

[0049] As defined herein the term "signal peptides" (SPs; also referred to as leader peptide, targeting signal, signal sequence, transit peptide or localization signal) are sequence motifs targeting proteins for translocation across the endoplasmic reticulum membrane. SPs are found at the amino terminus of nascent proteins, and function by prompting the transport mechanism within the cell to bring the proteins to their specific destination within the cell, or outside the cell if the proteins are to be secreted. If secreted in the extracellular environment, it may be specified that the SPs are secretory signal peptides. Thus, antibodies, which are naturally secreted like many other proteins, require a secretory signal peptide to function as sorting signals. In one example, the signal peptide as described herein may comprise 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 amino acids. In a further example, the sequence of the signal peptide as disclosed herein has about 10 to about 40 amino acids. In yet another example, the sequence of the signal peptide, as disclosed herein, has about 15 to about 35 amino acids. In another example, the sequence of the signal peptide as disclosed herein has about 18 to about 25 amino acids.

[0050] As described herein, the term "antibody" is used in the broadest sense and specifically covers monoclonal antibodies (including full length monoclonal antibodies), polyclonal antibodies, monospecific antibodies, multispecific antibodies (e.g., bispecific antibodies), and antibody fragments as long as they still exhibit the desired biological

activity. The term "monoclonal antibody" refers to an antibody composition having a homogenous (essentially identical) antibody population. The term is not limited regarding the species (e.g., human, murine, mouse, or cat) or the source of the antibody, nor is it limited by the manner in which it is made. For example, the term includes monoclonal antibodies produced by a methodology other than hybridoma which results in monoclonal antibodies no matter how it is subcategorized, e.g., hybrid, altered, chimeric, or humanized. Further, the term includes variants that naturally arise during the production of monoclonal antibodies. The term includes whole immunoglobulins. The term "humanized antibody", as used herein, refers to an engineered antibody that typically comprises the variable region of a non-human (e.g., murine) antibody, i.e. a chimeric antibody, or at least the complementarity determining regions (CDRs) thereof, and the remaining immunoglobulin portions derived from a human antibody. Procedures for the production of chimeric antibodies and further engineered monoclonal antibodies include those described in the art. Such humanized antibodies may be prepared by known techniques and offer the advantage of reduced immunogenicity when the antibodies are administered to humans.

[0051] Antibodies of the present invention can include at least one of a heavy chain constant region (H_c), a heavy chain variable region (H_v), a light chain variable region (L_v) and a light chain constant region (L_c), wherein a polyclonal Ab, monoclonal Ab, fragment and/or regions thereof include at least one heavy chain variable region (H_v) or light chain variable region (L_v).

[0052] For example the signal peptides of antibodies may be the signal peptides of the heavy chains of antibodies and the light chain of antibodies. The isotype of the antibody may comprise, but is not limited, to IgG, IgM, IgD, IgA and IgE. Thus, the heavy chain may comprise, but is not limited to, gamma, mu, delta, alpha and epsilon heavy chains, whereas the light chain may be a kappa or a lambda light chain. The antibody may be a mammalian antibody. The antibody may comprise, but is not limited to a human, a murine (rat and mice), an avian, a porcine, an equine, a bovine and a primate antibody.

[0053] Once the database had been created, the present inventors clustered the amino acid sequence based on similarity of the sequences. By “clustering” it is meant the task of grouping a set of sequences in such a way that sequences in the same group (called cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). To achieve clustering a number of algorithm are available depending on the task to

be achieved. In one embodiment, examples of clustering may comprise, but are not limited, to hierarchical clustering, centroid-based clustering, distribution-based clustering, density-based clustering and seed-based clustering. In a further example, the hierarchical clustering used in the disclosure may comprise an agglomerative clustering or a divisive clustering. Clustering methods specifically directed at primary sequence analysis are well known in the art and are commercially or publicly available.

[0054] For example, the present inventors tested a clustering algorithm referred to as Neighbor Joining with uncorrected ("p"), wherein the distance between pairwise sequence alignments is treated as the percentage number of residue positions at which the sequences differ. This Neighbor Joining method is a type of agglomerative hierarchical clustering method, which is a bottom-up clustering method for creating phonetic trees.

[0055] By clustering the signal peptide sequence present in the database, the inventors were able to create two groups of clusters. One group of clusters is related to the amino acid sequence information of signal peptides of the heavy chain of antibodies; the other group is related to the amino acid sequence information of signal peptides of the light chain of antibodies.

[0056] The number of clusters that is created will depend, for example, on the number of signal sequences present in the database, the experimental facilities (that is factors like the number of sequences that may be tested by the skilled artisan, the budget, and time allocated to testing), the similarity between the amino acid sequences of the signal peptides in the database and the number of identical sequences in the database. In case most amino acid sequences are identical then the number of clusters will be reduced. Conversely, the less similarity between the sequences in a database, the higher the number of clusters that will be created.

[0057] Where the number of clusters created is at least three clusters, clusters may comprise at least about 1%, at least about 2%, at least about 3%, at least about 4%, at least about 5%, at least about 6%, at least about 7%, at least about 8%, at least about 9%, at least about 10%, at least about 11%, at least about 12%, at least about 13%, at least about 14%, at least about 15%, at least about 16%, at least about 17%, at least about 18%, at least about 19%, at least about 20%, at least about 22%, at least about 25%, at least about 30%, at least about 40%, at least about 50% or at least about 60% of the total number of signal peptides sequences present in the database created based on publicly available information.

[0058] For example, based on a database comprising amino acid sequence information for 173 Ig heavy chain human signal peptides, the inventors created eight clusters, numbered H1 to H8 comprising between 16 and 30 signal peptide sequences, whereas based on a database comprising the signal peptide sequences from 62 human kappa chains, two clusters were created. These two clusters are represented for example, in Figure 1 of the drawings.

[0059] In a non-limiting example in Fig.1A representing the phylogenetic tree of heavy chain signal peptides from 173 human Ig heavy chain and 1C to 1F representing the amino acid sequences of the signal peptides within each clusters, clusters H1 and H6 for example (Fig. 1E and 1F) comprise 16 signal peptides sequences. In other words, they comprise 9.25% of the total number of signal peptide in the database (or $n = \frac{x(Hi)}{x(Ht)} \times 100$, wherein n is the number of peptide in a given cluster (given as a percentage) and wherein x(Hi) is the total number of signal peptides in a given cluster and x(Ht) is the total number of signal peptides in the database). In the non-limiting example above taking clusters H1 or H6 as an illustration, x(Hi) is 16 and x(Ht) is 173, thus $n = \frac{16}{173} \times 100 = 9.25\%$. Using the above equation, clusters H2 (23 signal peptides) and H3 (18 signal peptides) in Fig. 1 C have about 13.29% and 10.40% of the total number of signal peptides, whereas clusters 4 (25 signal peptides) and H5 (30 signal peptides) in Fig. 1D , H8 (19 signal peptides; Fig 1E), and H7 (26 signal peptides; Fig 1E) have 14.45%, 17.34%, 10.98% and 15.03%, respectively of the total number of signal peptides present in the exemplary database of 173 heavy chain signal peptides.

[0060] The light chain signal peptides database created from the available sequence information of 62 human kappa light chain only generated two clusters. The two clusters were created due partly to the high degree of similarity/identity of the sequences of the light chain signal peptides in the database. Using the non-limiting example above, L1 contains 18 light-chain signal peptides (or 29.03 % of the total number of signal peptides in the database) and L2 contains the remaining 44 light chain signal peptides (or 70.97% of the total number of signal peptides). However, it should be noted that 40 of the light chain signal peptides in cluster L2 have the same amino acid sequence (i.e. 100% sequence identity).

[0061] The resulting clusters may contain amino acid sequences of signal peptides that are the same (identical) or different. By "identical", it is meant that the amino acid sequence is 100% conserved (i.e. all the amino acid within the sequences are identical). By different, it is

meant that the primary amino acid sequence of the signal peptides within a cluster differ by at least one amino acid. The difference may be related to the length of the amino acid sequence (added or deleted amino acid), the identity of the amino acid (point mutation) or the structure of the sequence.

[0062] If the amino acid sequence is different, the amino acid difference may be at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen or at least fifteen amino acids. If the sequences are different, then a consensus sequence may be determined using any multiple alignment computer program available, based on the sequences present in the cluster.

[0063] Thus, in one example, there is provided a method of determining a signal peptide for enhancing secretion of an antibody of interest, wherein the method may comprise (a) creating a database consisting of amino acid sequences of a plurality of signal peptides from the heavy chain and the light chain of a plurality of antibodies; (b) clustering the sequences based on sequence similarity, wherein the number of clusters is determined based on the total number of the plurality of signal peptides such that any one of the clusters comprises at least about 5% of the plurality of signal peptides present in the database; (c) selecting at least one amino acid sequence within each clusters of heavy chain signal peptide; and light chain signal peptide, wherein if the amino acid sequences of the signal peptides within the clusters are: (1) identical, then any amino acid sequence may be selected within the cluster; and (2) different, then the at least one amino acid sequence may be selected based on: (i) the sequence appearing most often in the cluster; and/or (ii) the sequence having the highest similarity to a consensus sequence of the cluster, wherein the consensus sequence is created based on comparison of the amino acid signal peptide sequence within the cluster; (d) fusing polynucleotides encoding the heavy chain of the antibody to a polynucleotide encoding the heavy chain signal peptide selected as defined hereinbefore; and the light chain of the antibody to a polynucleotide encoding the light chain signal peptide selected as defined hereinbefore; wherein selection of different signal peptide amino acid sequences for the antibody results in a library of antibodies fused to different signal sequences selected hereinbefore; (e) creating at least one vector comprising the polynucleotides defined above; and (f) quantifying and comparing the secretion of the antibody after transfection of the at least one vector above into an expression system to determine the combination of heavy and

light chain signal peptides, which enhance expression and secretion of the antibody of interest the most.

[0064] In one limiting example, by applying the method of the invention described herein, step (a) when applied to available human immunoglobulin amino acid sequence information generated a database of 173 immunoglobulin heavy chain signal peptides and 62 kappa light chain signal peptides amino acid sequences. As more information becomes available, more databases may be created. As explained using the non-limiting example herein, when the clustering of step (b) is applied to the database of 173 heavy chain signal peptides, eight clusters were created, named by H1 to H8, as shown in Fig. 1C to 1F.

[0065] The method may further comprise the step of selecting at least one amino acid sequence within at least one cluster of heavy chain signal peptides and light chain signal peptides. For example, the selection may be based on the identity of amino acid sequence within one cluster as indicated above. If all the sequences within at least one cluster are identical as defined above, then any sequence can be selected.

[0066] If the sequences within one cluster are different, then at least one signal peptide is selected as being representative of the cluster. For example, if the signal peptide is represented more than once within the cluster and if the signal peptide is represented more often than any other signal peptide within the same cluster, then the signal peptide is selected as being the most representative of the cluster. In case every one of the signal peptides within a cluster has a different amino acid sequence then the most representative signal peptide can be the signal peptide having the highest degree of primary sequence similarity to the consensus sequence of said cluster.

[0067] In a non-limiting example, cluster H3 (shown in Fig. 1C) that was created using the signal peptide sequences from 173 Ig human heavy chains described above, contains the signal peptide sequences of 18 Ig human heavy chains. Of the 18 signal peptide sequences, 10 are identical (that is have the same sequence MKHLWFFLLLVAAPRWVLS (H3, SEQ ID NO: 7)) and thus, the signal peptide sequence of SEQ ID NO: 7 was selected as being representative of cluster H3. In a further example, applying the selection step (c) of the method described above to the 23 signal peptide sequences of cluster H2 (see e.g. Fig. 1C), since no sequence appears more often than any other, the sequence selected is MELGLRWVFLVAILEGVQC (H2, SEQ ID NO: 6), which shares the highest similarity with the consensus sequence deducted from the list of 23 signal peptide sequences of cluster

H2 as stated in further step 2 (ii) of the method as described herein. Applying the same method to other exemplary clusters H1, H4 to H8, the following approach is taken. For clusters H4, H5 (see e.g. Fig. 1D), H8, H1 (see e.g. Fig 1E), and H6 (see Fig. 1F), the signal peptide sequence is selected according to the step (c).(2).(i), that is according to the fact that these sequence appear most often in aforementioned clusters. Thus, the following amino acid sequences were selected from clusters H4, H5, H8, H1 and H6, respectively: MDWTWRILFLVAAATGAHS (H4, SEQ ID NO: 8), MDWTWRFLFVVAAATGVQS (H5, SEQ ID NO: 1), MDLLHKNMKHLWFFLLLVAAPRWVLS (H8, SEQ ID NO: 10), MELGLSWIFLLAILKGVQC (H1, SEQ ID NO: 5), and MEFGLSWLFLVAILKGVQC (H6, SEQ ID NO: 9). For exemplary cluster H7 (see Fig. 1F), the signal peptide having the amino acid sequence MEFGLSWVFLVALFRGVQC (H7, SEQ ID NO:3) is selected based on the highest similarity to the consensus sequence deducted from the alignment of all 26 sequences within the cluster.

[0068] The signal peptide was further optimised by comparing the sequences of the signal peptide within a cluster. For example, the N-domain, H-domain and C-domain described herein were found to encompass the first to the sixth amino acid, the seventh to the fourteenth amino acid and the fifteenth to the nineteenth amino acid of the Ig human heavy chain signal peptide sequence, respectively.

[0069] The method of determining the signal peptide for enhancing secretion may further comprise the step of fusing at least one polynucleotide (also referred as oligonucleotide or nucleotide) encoding a signal peptide sequence to at least one polynucleotide encoding the light chain of the antibody of interest and/or the heavy chain of the antibody of interest, wherein the signal peptide has been selected as indicated above and wherein at least one signal peptide is selected from the clusters created from the database created from the signal peptides of light chains and at least one signal peptide is selected from the clusters created from the signal peptides of heavy chains. The number of signal peptides selected may depend on the number of clusters created and the number of peptides selected from each cluster. For example, the number of heavy chain signal peptides selected may comprise 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25 or 30 different signal peptides. The number of light chain signal peptides selected may comprise 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30 or 35 different signal peptides.

[0070] Examples of selected signal peptides are given in the experimental section below and a list of exemplary publicly available antibodies signal peptides is given in Fig. 1. For example, the antibody of interest may be modified by a heavy chain signal peptide. In one example, there is provided the antibody as disclosed herein, wherein the N-domain comprises an amino acid sequence of formula (II): M-X_{aa1}-X_{aa2}-X_{aa3}-X_{aa4}-X_{aa5} (SEQ ID NO: 278), wherein X_{aa1} is any one of Asp, Asn or Lys; X_{aa2} is any one of Phe, Leu, His or Trp; X_{aa3} is any one of Gly, Leu or Thr; X_{aa4} is Leu or Trp; X_{aa5} is any one of Ser, Arg or Phe; wherein the amino acid sequence of formula (II) has optionally one or two mutations.

[0071] In another example, there is provided the antibody, as described herein, wherein the C-domain comprises an amino acid sequence of the formula (III), X_{aa14}-X_{aa15}-X_{aa16}-X_{aa17}-X_{aa18} (SEQ ID NO: 279), wherein X_{aa14} is any one of Arg, Lys, Asp or Thr; X_{aa15} is Gly or Trp; X_{aa16} is Val or Ala; X_{aa17} is any one of Glu, Leu, or His; X_{aa18} is Cys or Ser; wherein the amino acid sequence of formula (III) has optionally one or two mutations.

[0072] In an additional example, there is provided the antibody, as disclosed herein, further excluding antibodies whose secretion is no more than 1.5 fold higher when compared to the secretion of the antibody with the lowest secretion. In yet another example, there is provided the antibody, as described herein, further excluding antibodies whose secretion is no more than 2 fold higher when compared to the secretion of the antibody with the lowest secretion.

[0073] In one example, there is provide the antibody as disclosed herein, wherein the amino acid sequence of the N-domain of the heavy chain signal peptide is selected from the group consisting of MELGLS (SEQ ID NO: 293), MELGLR (SEQ ID NO: 294), MKHLWF (SEQ ID NO: 295), MDWTWR (SEQ ID NO: 296) and MEFGLS (SEQ ID NO: 297). In another example, there is provided the antibody, as described herein, wherein the amino acid sequence of the C-domain of the heavy chain signal peptide is selected from the group consisting of KGVQC (SEQ ID NO: 298), EGVQC (SEQ ID NO: 299), RWVLS (SEQ ID NO: 300), TGAHS (SEQ ID NO: 301), TGVQS (SEQ ID NO: 302) and RGVQC (SEQ ID NO: 303). In yet another example, there is provided the antibody, as described herein, wherein the amino acid sequence of the H-domain of the heavy chain signal peptide is selected from the group consisting of WIFLLAIL (SEQ ID NO: 304), WVFLVAIL (SEQ ID NO: 305), FLLLVAAP (SEQ ID NO: 306), ILFLVAAA (SEQ ID NO: 307), FLFVVAAA (SEQ ID NO: 308), WLFLVAIL (SEQ ID NO: 309) and WVFLVALF (SEQ ID NO: 310).

[0074] Accordingly in one example, there is provided the antibody of interest, as disclosed herein, comprising a heavy chain signal peptide, wherein the amino acid sequence of the heavy chain signal peptide comprises but is not limited to MELGLSWIFLLAILKGVQC (H1, SEQ ID NO: 5), MELGLRWVFLVAILEGVQC (H2, SEQ ID NO: 6), MKHLWFFLLLVAAPRWVLS (H3, SEQ ID NO: 7), MDWTWRILFLVAAATGAHS (H4, SEQ ID NO: 8), MDWTWRFLFVVAAATGVQS (H5, SEQ ID NO: 1), MEFGLSWLFLVAILKGVQC (H6, SEQ ID NO: 9), MEFGLSWVFLVALFRGVQC (H7, SEQ ID NO:3), MDLLHKNMKHLWFFLLLVAAPRWVLS (H8, SEQ ID NO: 10).

[0075] In a further example, there is provided an antibody, as disclosed herein, wherein the heavy chain of the antibody of interest is fused to a signal peptide comprising, but not limited to, MDWTWRFLFVVAAATGVQS (H5, SEQ ID NO: 1), and MEFGLSWVFLVALFRGVQC (H7, SEQ ID NO:3).

[0076] In one 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}-X_{bb7}-X_{bb8}-X_{bb9}-X_{bb10}-X_{bb11}-X_{bb12}-X_{bb13}-X_{bb14} (SEQ ID NO: 280), 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.

[0077] In one embodiment, there is provided the antibody as disclosed herein, wherein the light signal peptide is selected from signal peptides of mammalian kappa light chains. In a further example, there is provided the antibody, as defined herein, wherein the antibody comprises an amino acid sequence of the formula (V): M-X_{bb1}-X_{bb2}-X_{bb3}-X_{bb4}-X_{bb5} (SEQ ID NO: 281), wherein X_{bb1} is Asp or Lys; X_{bb2} is Met or Tyr; X_{bb3} is Arg or Leu; X_{bb4} is Val or Leu; X_{bb5} is Pro; and wherein the amino acid sequence of formula (V) has optionally one or two mutations.

[0078] In yet another example, there is provided the antibody, as defined herein, wherein the antibody further comprises an amino acid sequence of the formula (VI): X_{bb15}-X_{bb16}-X_{bb17}-X_{bb18}-X_{bb19}- X_{bb20}-X_{bb21} (SEQ ID NO: 282), wherein X_{bb15} is Ala or Trp; X_{bb16} is Ala or Leu; X_{bb17} is Ser or Glu; X_{bb18} is Gly or Pro; X_{bb19} is Ala; X_{bb20} is Arg or Met; X_{bb21} is Ala or Cys; and wherein the amino acid sequence of formula (VI) has optionally one or two mutations.

[0079] In one example, there is provided the antibody, as defined herein, wherein the amino acid sequence of the light chain signal peptide may comprise MDMRVP (SEQ ID NO: 311) or MKYLLP (SEQ ID NO: 312). In another example, provided is the antibody, as defined herein, wherein the amino acid sequence of the light chain signal peptide may further comprise SGARC (SEQ ID NO: 313) or QPAMA (SEQ ID NO: 314). In a further example, there is provided the antibody, as defined herein, wherein the amino acid sequence of the light chain signal peptide may also comprise AQLLGLLLLWL (SEQ ID NO: 315) or TAAAGLLLLAA (SEQ ID NO: 316).

[0080] Accordingly, in another example, there is provided an antibody of interest comprising a light chain signal peptide, wherein the light chain of the antibody of interest is fused to a signal peptide comprising but not limited to MDMRVPAQLLGLLLLWLSGARC (L1, SEQ ID NO: 2) and MKYLLPTAAAGLLLLAAQPAMA (L2, SEQ ID NO: 4).

[0081] Since the antibody of interest may comprise a heavy chain and a light, the present invention also provides for combination of heavy chain and light chain signal peptides to be fused to the antibody of interest in order to improve expression and secretion of said antibody. Accordingly in one example, there is provided the antibody, as disclosed herein, wherein the antibody of interest comprises combinations of one heavy and one light chain signal peptide as defined above for the heavy chain and for the light chain.

[0082] In one example, provided herein is an antibody of interest comprising a combination of a heavy chain signal peptide and a light 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 heavy chain of the antibody via the C-terminal end of the C-domain and the light chain signal peptide is bound to the light chain of the antibody via the C-terminal end of the C-domain; wherein the H-domain of the heavy chain signal peptide comprises an amino acid sequence of formula (I): X_{aa6}-X_{aa7}-X_{aa8}-X_{aa9}-X_{aa10}-X_{aa11}-X_{aa12}-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 light chain comprises an amino acid sequence of formula (IV): X_{bb6}-X_{bb7}-X_{bb8}-X_{bb9}-X_{bb10}-X_{bb11}-X_{bb12}-X_{bb13}-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; and wherein the amino acid sequence of formula (I) or the formula (IV) has independently 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, and 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 heavy chain signal peptides and antibodies of interest and combinations of light chain signal peptides and antibodies 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.

[0083] The exclusion of the antibody with the lowest secretion ensures that the selected signal peptide sequence to be fused to the antibody of interest is optimised. In other words, the secretion and expression of the antibody of interest when fused to the selected signal peptide, is improved in comparison to the antibody fused to the signal peptide with the lowest secretion. Methods to create the polynucleotide and measure secretion will be explained in more details below and in the examples.

[0084] A threshold may be set to measure antibodies that may be excluded due to their low secretion and may be determined experimentally and empirically. Methods to measure secretion are given in the experimental section below. For example, Figure 3A is a histogram plot that represents the results provided by an enzyme-linked immunosorbent assay (ELISA) to measure the titer of a mAb produced by the method as described herein. More specifically, in Figure 3A, an antibody was created by fusing the heavy chain of Herceptin, an antibody of interest, to an Ig heavy chain human signal peptide selected by using the method as described herein. The heavy chain signal peptide was selected from one each of the eight clusters H1 to H8, described herein. Thus, in this example, eight different antibodies corresponding to a fusion of the heavy chain of the Herceptin antibody having the SEQ ID NO: 11 and a heavy

chain signal peptide having a sequence of SEQ ID NO: 1, 3, 5, 6, 7, 8, 9 or 10 were created. The antibody was created by fusing the polynucleotide sequence of the heavy chain of Herceptin of SEQ ID NO: 21 in frame with the polynucleotide sequence of the heavy chain signal peptide of any one of SEQ ID NO: 283 to 290 as described herein, inserted in a vector and transfected in CHO cells, together with a same or different vector, comprising the polynucleotide sequence of the light chain of Herceptin having the SEQ ID NO: 26 (encoding for the light chain of Herceptin having the amino acid sequence of SEQ ID NO: 16) fused in frame with the polynucleotide sequence of the light chain signal peptide of SEQ ID NO: 291 (encoding for the light chain signal peptide having the amino acid sequence of SEQ ID NO: 2) in the presence of a vector encoding for the green fluorescent protein (GFP) or any protein or molecule known in the art allowing measurement of transfection efficiency in individual cells or a population of cells . The secreted antibody comprising the heavy chain and the light chain is then titered by ELISA a method well known by a person skilled in the art. The transfection efficiency is measured in the host cells by measuring, for example, the mean fluorescent intensity of the GFP. Accordingly, a raw titer is obtained for all combinations of antibodies comprising the heavy chain of Herceptin fused to the heavy chain signal peptide, as described above, and the light chain of Herceptin fused to the light chain signal peptide, as described above.

[0085] In a next step, the values obtained above are first normalized using the mean fluorescence intensity given by the GFP to account for variation in transfection efficiency among the different vectors described above. The secretion efficiency of the heavy chain of Herceptin of SEQ ID NO: 11-heavy chain signal peptide combinations and Herceptin light chain of SEQ ID NO: 16-light chain signal peptide L1 of SEQ ID NO: 2 is then compared to, for example, the combination formed by the heavy chain of Herceptin and the H1 signal peptide of SEQ ID NO: 5 and the light chain of Herceptin SEQ ID NO: 16-light chain signal peptide L1 of SEQ ID NO: 2. Accordingly, as shown for example in Figure 3B, the Herceptin antibody combination of a heavy chain signal peptide H1 fused to the heavy chain of Herceptin and a light chain signal peptide fused to light chain of Herceptin as described above is given an arbitrary fold change of 1.0.

[0086] Thus, the antibody serves as a reference for normalization and comparison of the secretion of the different antibodies combination. For example, in the present invention forty combinations of heavy chain-heavy chain signal peptides were created corresponding to eight

different signal peptides fused to five different antibodies and ten combinations of light chain-light chain signal peptides were created corresponding to two different signal peptides fused to five different antibodies. As a result, when expressing the vectors encoding for the aforementioned heavy chain and light chain combinations, a total of eighty antibodies combinations was created.

[0087] Using H1 as a reference, any combination that has a fold-change higher than 1.0 has a higher secretion than H1 as described above, any combination that has a fold-change lower than 1.0 has a lower secretion than H1 as described above. Using Fig. 3B as an example, when combinations of Herceptin heavy chain-heavy signal peptide are co-transfected in the presence of Herceptin light chain-light chain signal peptide L1, H2/L1, H3/L1, H4/L1, H5/L1, H6/L1 and H8/L1 have a higher secretion than the H1/L1 combination. Conversely, the H7/L1 combination has a lower secretion than H1/L1. The H5/L1 has a secretion approximately 2.2-fold higher than H1/L1.

[0088] Based on the above analysis, the antibodies with the lowest secretion may be excluded. For example, when secretion is measured and compared between the library of antibodies that are created as described above, the antibodies whose secretion is no more than at least about 1.2-fold, or at least about 1.3-fold, or at least about 1.4-fold, or at least about 1.5-fold, or at least about 1.6-fold, or at least about 1.7-fold, or at least about 1.8-fold, or at least about 1.9-fold, or at least about 2.0-fold, or at least about 2.5-fold, or at least about 3.0-fold, or at least about 3.5-fold, or at least about 4.0-fold, or at least about 5.0-fold when compared to the secretion of the antibody with the lowest secretion are excluded.

[0089] As an example, in Fig. 4A and B as well as the experimental section below, the antibody with the lowest normalized secretion is the H7/L1 combination. The H1/L1 combination secretion is about 1.7-fold (i.e. fold change of H1/L1÷ fold change of H7/L1 or 1÷0.6) higher than the H7/L1 combination. Thus, by setting the exclusion threshold at 2.0-fold higher than the antibody with the lowest secretion (H7/L1 in this example), the H1/L1 combination would be excluded when Herceptin is the antibody of interest. If the 2.5-fold is selected, then the H6/L1 combination would be further excluded. The combinations with the highest fold-change (i.e. H5/L1 and H8/L1) produce approximately 3.75-fold more antibodies than the H7/L1 combination, when expressed in CHO cells.

[0090] By excluded, it is meant that the particular combination of the antibody of interest and signal peptide is not further considered for testing for improved secretion and expression

in the expression system into which the vector encoding the antibody of interest was transfected. However, it does not exclude using the antibody as a negative control or further testing the antibody of interest in other expression systems, or modifying the antibody, for example, by mutating the signal peptide sequence that was used.

[0091] In a further example, there is provided the antibody as described herein comprising a heavy chain and a light chain, as described herein, wherein the N-domain of the heavy chain comprises an amino acid sequence of formula (II): M-X_{aa1}-X_{aa2}-X_{aa3}-X_{aa4}-X_{aa5}, wherein X_{aa1} is any one of Asp, Asn or Lys; X_{aa2} is any one of Phe, Leu, His or Trp; X_{aa3} is any one of Gly, Leu or Thr; X_{aa4} is Leu or Trp; X_{aa5} is any one of Ser, Arg or Phe; and wherein the light chain comprises an amino acid sequence of the formula (V): M-X_{bb1}-X_{bb2}-X_{bb3}-X_{bb4}-X_{bb5}, wherein X_{bb1} is Asp or Lys; X_{bb2} is Met or Tyr; X_{bb3} is Arg or Leu; X_{bb4} is Val or Leu; X_{bb5} is Pro; wherein the amino acid sequence of formula (II) and/or formula (V) has independently, optionally one or two mutations. In yet another example, provided herein is the antibody comprising a heavy chain and a light chain, as described herein, wherein the C-domain of the heavy chain signal peptide comprises an amino acid sequence of the formula (III), X_{aa14}-X_{aa15}-X_{aa16}-X_{aa17}-X_{aa18}, wherein X_{aa14} is any one of Arg, Lys, Asp or Thr; X_{aa15} is Gly or Trp; X_{aa16} is Val or Ala; X_{aa17} is any one of Glu, Leu, or His; X_{aa18} is Cys or Ser; and wherein the light chain signal peptide comprises an amino acid sequence of the formula (VI): X_{bb15}-X_{bb16}-X_{bb17}-X_{bb18}-X_{bb19}- X_{bb20}-X_{bb21}, wherein X_{bb15} is Ala or Trp; X_{bb16} is Ala or Leu; X_{bb17} is Ser or Glu; X_{bb18} is Gly or Pro; X_{bb19} is Ala; X_{bb20} is Arg or Met; X_{bb21} is Ala or Cys; and wherein the amino acid sequence of formula (III) and/or formula (VI) has independently, optionally one or two mutations.

[0092] Accordingly, in one example, there is provided the antibody, as described herein, wherein the amino acid sequence of the N-domain of the heavy chain signal peptide may comprise but is not limited to, MELGLS (SEQ ID NO: 293), MELGLR (SEQ ID NO: 294), MKHLWF (SEQ ID NO: 295), MDWTWR (SEQ ID NO: 296) and MEFGLS (SEQ ID NO: 297) and wherein the amino acid sequence of the light chain signal peptide may comprise MDMRVP (SEQ ID NO: 311) or MKYLLP (SEQ ID NO: 312). In a further example, there is provided the antibody as described herein, wherein the amino acid sequence of the C-domain of the heavy chain signal peptide may comprise but is not limited to KGVQC (SEQ ID NO: 298), EGVQC (SEQ ID NO: 299), RWVLS (SEQ ID NO: 300), TGAHS (SEQ ID NO: 301), TGVQS (SEQ ID NO: 302) and RGVQC (SEQ ID NO: 303) and wherein the amino acid

sequence of the light chain signal peptide may further comprise SGARC (SEQ ID NO: 313) or QPAMA (SEQ ID NO: 314). In another example, there is provided the antibody, as described herein, wherein the amino acid sequence of the H-domain of the heavy chain signal peptide may comprise but is not limited to WIFLLAIL (SEQ ID NO: 304), WVFLVAIL (SEQ ID NO: 305), FLLLVAAP (SEQ ID NO: 306), ILFLVAAA (SEQ ID NO: 307), FLFVVAAA (SEQ ID NO: 308), WLFLVAIL (SEQ ID NO: 309) and WVFLVALF (SEQ ID NO: 310) and wherein the amino acid sequence of the light chain signal peptide is AQLLGLLLLWL (SEQ ID NO: 315) or TAAAGLLLLAA (SEQ ID NO: 316).

[0093] Thus, in one example there is provided the antibody of interest, as disclosed herein, comprising a heavy chain signal peptide, wherein the amino acid sequence of the heavy chain signal peptide comprises but is not limited to MELGLSWIFLLAILKGVQC (H1, SEQ ID NO: 5), MELGLRWVFLVAILEGVQC (H2, SEQ ID NO: 6), MKHLWFFLLLVAAPRWVLS (H3, SEQ ID NO: 7), MDWTWRILFLVAAATGAHS (H4, SEQ ID NO: 8), MDWTWRFLFVVAAATGVQS (H5, SEQ ID NO: 1), MEFGLSWLFLVAILKGVQC (H6, SEQ ID NO: 9), MEFGLSWVFLVALFRGVQC (H7, SEQ ID NO:3), MDLLHKNMKHLWFFLLLVAAPRWVLS (H8, SEQ ID NO: 10) and a light chain signal peptide, wherein the light chain of the antibody of interest is fused to a signal peptide comprising MDMRVPAQLLGLLLLWLSGARC (L1, SEQ ID NO: 2) or MKYLLPTAAAGLLLLAAQPAMA (L2, SEQ ID NO: 4).

[0094] Thus, in one example, the method, as disclosed herein, comprises the step of fusing polynucleotides encoding the heavy chain of the antibody to a polynucleotide encoding the heavy chain signal peptide selected, as described herein, and above thereby creating a recombinant heavy chain of the antibody of interest modified with a signal peptide. In another example, the method, as disclosed herein, comprises the step of fusing polynucleotides encoding the light chain of the antibody to a polynucleotide encoding the light chain signal peptide selected, as described herein, and above thereby creating a recombinant heavy chain of the antibody of interest modified with a signal peptide. Fusion of the polynucleotides may be performed using techniques known to the skilled artisan, such as enzymatic ligation, or polymerase chain reaction. The polynucleotides sequences may be obtained using publicly available information or by reverse translation of the amino acid sequences. It would be known to the person skilled in the art how to operably link the coding sequence of the signal peptide and the coding sequence of the antibody so that the resulting

protein product is a functional protein product with the desired amino acid sequence. Therefore, the polynucleotide sequence of the signal peptide and the antibody need to be in frame.

[0095] The fusion may be performed in such a way that the encoded signal peptide is at the N-terminus of the light chain or at the N-terminus of the heavy chain. In a further example, the signal peptide may be linked to the N-terminus of the variable region of light chain or to the N-terminus of the variable region of the heavy chain of the antibody of interest. In one example, the signal peptide that is fused to the light chain may comprise a signal peptide selected from a cluster created based on dataset generated from the amino acid sequence available of signal peptide from the kappa light chain of antibodies. In another example, the signal peptide that is fused to the light chain may comprise a signal peptide selected from a cluster created based on dataset generated from the amino acid sequence available of signal peptide from the heavy chain of antibodies.

[0096] The fused polynucleotide may include a linker to separate the variable region from the signal peptide. The sequence of the signal peptide may include a specific cleavable sequence that may be recognized and specifically cleaved by a signal peptidase. The polynucleotide sequence encoding for the light chain or the heavy chain of the antibody of interest may comprise the full sequence of the light chain or the heavy chain including the variable region and the constant region of the light chain or the heavy chain or a partial sequence of the light chain or the heavy chain that may encode for the region of the light chain or the heavy chain of the antibody that is of therapeutic interest.

[0097] The invention also provides for "derivatives" of the murine or chimeric antibodies, fragments, regions or derivatives thereof, which term includes those proteins encoded by truncated or modified genes to yield molecular species functionally resembling the immunoglobulin fragments. The modifications include, but are not limited to, the addition of genetic sequences coding for cytotoxic proteins such as plant and bacterial toxins. The fragments and derivatives can be produced from any of the hosts of this invention.

[0098] Fragments include, for example, Fab, Fab', F(ab')2 and Fv. These fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and can have less non-specific tissue binding than an intact antibody. These fragments are produced from intact antibodies using methods well known in the art, for example by proteolytic cleavage with enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')2 fragments).

[0099] In one example, the at least one polynucleotide encoding for the at least one signal peptide fused to a polynucleotide encoding for the light chain and/or the heavy chain is integrated in a vector. In view of the present disclosure, the nucleic acid molecule can be obtained using any molecular biology method known in the art. Those skilled in the art will know how to operably link the signal peptide coding sequence to a coding sequence for a heavy or light chain of an antibody of interest, to thereby obtain a nucleic acid molecule according to an embodiment of the present invention.

[00100] The cloning of the recombinant polynucleotide into a vector is well known to the person skilled in the art. A plurality of vectors is available to the skilled artisan to obtain expression of the recombinant light chain and/or heavy chain in the expression system of interest. In one example, at least one vector is created comprising the polynucleotide coding sequence for the signal peptide operably linked to the coding sequence of the heavy chain and/or the light chain of the antibody of interest. In one example, there is provided a vector encoding for at least one recombinant light chain comprising a signal peptide and a light chain which may further comprise a variable and a constant region and at least one recombinant heavy chain comprising a signal peptide and a heavy chain which may further comprise a variable and a constant region. In a further example, there is provided at least one vector encoding at least one recombinant light chain comprising a signal peptide and a light chain which may further comprise a variable and a constant region. In yet another example there is provided at least one vector encoding at least one recombinant heavy chain comprising a signal peptide and a light chain which may further comprise a variable and a constant region. It is meant herein by a "vector encoding", a nucleic acid molecule comprising the polynucleotide sequence of the recombinant antibody of interest that when introduced in a host cell is capable of expressing the recombinant antibody of interest (or the light chain and/or heavy chain thereof), thereby producing the amino acid sequence of the antibody of interest fused to a functional signal peptide. The nucleic acid molecule may be an expression vector.

[00101] Two coding DNA sequences are said to be "operably linked" if the linkage results in a continuously translatable sequence without alteration or interruption of the triplet reading frame. A DNA coding sequence is operably linked to a gene expression element if the linkage results in the proper function of that gene expression element resulting in expression of the coding sequence.

[00102] Expression vehicles include plasmids or other vectors. Preferred among these are vehicles carrying a functionally complete human C_H or C_L chain sequence having appropriate restriction sites engineered so that any V_H or V_L chain sequence with appropriate cohesive ends can be easily inserted therein. Human C_H or C_L chain sequence-containing vehicles thus serve as intermediates for the expression of any desired complete H or L chain in any appropriate host.

[00103] Oligonucleotides representing a portion of the variable region sequence are useful for screening for the presence of homologous genes and for the cloning of such genes encoding variable or constant regions of an antibody of interest. Such probes preferably bind to portions of sequences which encode light chain or heavy chain variable regions which bind an epitope of an antigen of interest. Such techniques for synthesizing such oligonucleotides are well known and disclosed in the art.

[00104] Because the genetic code is degenerate, more than one codon can be used to encode a particular amino acid. Using the genetic code, one or more different oligonucleotides can be identified, each of which would be capable of encoding the amino acid. The probability that a particular oligonucleotide will, in fact, constitute the actual XXX-encoding sequence can be estimated by considering abnormal base pairing relationships and the frequency with which a particular codon is actually used (to encode a particular amino acid) in eukaryotic or prokaryotic cells expressing an antibody of interest or fragment. Such "codon usage rules" are disclosed in the art. Using the "codon usage rules" of Lathe, a single oligonucleotide, or a set of oligonucleotides, that contains a theoretical "most probable" nucleotide sequence capable of encoding antibody of interest variable or constant region sequences is identified.

[00105] Although occasionally an amino acid sequence can be encoded by only a single oligonucleotide, frequently the amino acid sequence can be encoded by any of a set of similar oligonucleotides. Importantly, whereas all of the members of this set contain oligonucleotides which are capable of encoding the peptide fragment and, thus, potentially contain the same oligonucleotide sequence as the gene which encodes the peptide fragment, only one member of the set contains the nucleotide sequence that is identical to the nucleotide sequence of the gene. Because this member is present within the set, and is capable of hybridizing to DNA even in the presence of the other members of the set, it is possible to employ the

unfractionated set of oligonucleotides in the same manner in which one would employ a single oligonucleotide to clone the gene that encodes the protein.

[00106] The oligonucleotide or set of oligonucleotides, containing the theoretical "most probable" sequence capable of encoding an antibody or fragment including a variable or constant region is used to identify the sequence of a complementary oligonucleotide or set of oligonucleotides which is capable of hybridizing to the "most probable" sequence, or set of sequences. An oligonucleotide containing such a complementary sequence can be employed as a probe to identify and isolate the variable or constant region gene.

[00107] A suitable oligonucleotide, or set of oligonucleotides, which is capable of encoding a fragment of the variable or constant antibody region (or which is complementary to such an oligonucleotide, or set of oligonucleotides) is identified (using the above-described procedure), synthesized, and hybridized by means well known in the art, against a DNA or, more preferably, a cDNA preparation derived from cells which are capable of expressing antibodies or variable or constant regions thereof. Single stranded oligonucleotide molecules complementary to the "most probable" variable or constant region peptide coding sequences can be synthesized using procedures which are well known to those of ordinary skill in the art. Additionally, DNA synthesis can be achieved through the use of automated synthesizers. Techniques of nucleic acid hybridization are disclosed. Techniques such as, or similar to, those described above have successfully enabled the cloning of genes for human aldehyde dehydrogenases, fibronectin, the human estrogen receptor gene, tissue-type plasminogen activator and human term placental alkaline phosphatase complementary DNA.

[00108] In an alternative way of cloning a polynucleotide encoding a variable or constant region, a library of expression vectors is prepared by cloning DNA or, more preferably, cDNA (from a cell capable of expressing an antibody or variable or constant region) into an expression vector. The library is then screened for members capable of expressing a protein which competitively inhibits the binding of an antibody, and which has a nucleotide sequence that is capable of encoding polypeptides that have the same amino acid sequence as antibodies of interest or fragments thereof. In this embodiment, DNA, or more preferably cDNA, is extracted and purified from a cell which is capable of expressing an antibody of interest or fragment. The purified cDNA is fragmentized (by shearing, endonuclease digestion, etc.) to produce a pool of DNA or cDNA fragments. DNA or cDNA fragments from this pool are then cloned into an expression vector in order to produce a genomic library

of expression vectors whose members each contain a unique cloned DNA or cDNA fragment such as in a lambda phage library, expression in prokaryotic cell (e.g., bacteria) or eukaryotic cells, (e.g., mammalian, yeast, insect or, fungus). Once the nucleic acid sequence encoding such variable or constant regions is isolated, the nucleic acid sequence can be appropriately expressed in a host cell, along with other constant or variable heavy or light chain encoding nucleic acid, in order to provide recombinant MAbs. Such antibodies preferably include a murine or human variable region which contains a framework residue having complementarity determining residues which are responsible for antigen binding.

[00109] Human genes which encode the constant (C) regions of the murine and chimeric antibodies, fragments and regions of the present invention can be derived from a human fetal liver library, by known methods. Human C regions genes can be derived from any human cell including those which express and produce human immunoglobulins. The human C_H region can be derived from any of the known classes or isotypes of human H chains, including gamma, μ, α, δ or ε, and subtypes thereof, such as G1, G2, G3 and G4. Since the H chain isotype is responsible for the various effector functions of an antibody, the choice of CH region will be guided by the desired effector functions, such as complement fixation, or activity in antibody-dependent cellular cytotoxicity (ADCC). Preferably, the CH region is derived from gamma 1 (IgG1), gamma 3 (IgG3), gamma 4 (IgG4), or μ (IgM).

[00110] The human C_L region can be derived from human L chain isotype, kappa or lambda. Genes encoding human immunoglobulin C regions are obtained from human cells by standard cloning techniques. Human C region genes are readily available from known clones containing genes representing the two classes of L chains, the five classes of H chains and subclasses thereof. Chimeric antibody fragments, such as F(ab')2 and Fab, can be prepared by designing a chimeric H chain gene which is appropriately truncated. For example, a chimeric gene encoding an H chain portion of an F(ab')2 fragment would include DNA sequences encoding the CH1 domain and the hinge region of the H chain, followed by a translational stop codon to yield the truncated molecule.

[00111] Generally, the murine, human and chimeric antibodies, fragments and regions of the present invention are produced by cloning DNA segments encoding the H and L chain antigen-binding regions of an antibody of interest, and joining these DNA segments to DNA segments encoding C_H and C_L regions, respectively, to produce murine, human or chimeric immunoglobulin-encoding genes.

[00112] Thus, in a preferred embodiment, a fused chimeric gene is created which comprises a first DNA segment that encodes at least the antigen-binding region of non-human origin, such as a functionally rearranged V region with joining (J) segment, linked to a second DNA segment encoding at least a part of a human C region.

[00113] In a further example, the method can comprise quantifying and comparing secretion of the antibody after transfection of at least one vector into an expression system to determine the combination of heavy and light chain signal peptides, as to which combinations enhance expression and secretion of the antibody the most. Introduction of the vector into the expression system may be achieved by transfection, electroporation or any technique known in the art. As used herein the term "expression system" relates to a system that allows protein expression, that is the way in which proteins are synthesized, modified and regulated in living organism. The term can refer to the system required to manufacture proteins. Recombinant protein production in living organisms will depend on using cellular machinery.

[00114] The expression system may be a host cell or a cell-free expression system. For the purpose of the present invention where secretion of the recombinant antibody is to be assessed the expression system may be a host cell. The vector can be introduced in a host cell to obtain a recombinant cell. The recombinant cell can be grown under conditions suitable for the expression and secretion of the antibody of interest.

[00115] Transfection may be a single transfection if the vector encodes for the recombinant heavy chain and the recombinant light chain or may be a co-transfection, to introduce at least two vectors independently encoding for a recombinant heavy chain comprising the heavy chain of the antibody of interest and the signal peptide to be tested and for a recombinant light chain comprising the light chain of the antibody of interest and the signal peptide to be tested. According to an example, transfection of the at least one vector in the expression system may be transient or stable. In case the transfection is a stable transfection, at least one vector may encode for a selectable marker, thereby allowing identification of expression system having the vector of interest.

[00116] Quantification of the secreted recombinant antibody may comprise but is not limited to IgG ELISA assay, Western Blotting, immunoprecipitation, immunofluorescence, mass spectrometry or Fluorescence activated cell sorting (FACS). The quantification of the secretion of the recombinant antibody may provide numerical values thereby allowing comparison between different recombinant antibodies comprising different signal peptides.

Thus, the method allows determining which signal peptide enhances expression and secretion of the antibody of interest the most.

[00117] In one embodiment, the method as described above may further comprise mutating at least one amino acid of the signal peptide to be fused to the antibody of interest. For example, the signal peptide amino acid sequence may comprise at least one, at least two, at least three, at least four or at least five mutations. In one example, disclosed herein, is the antibody of interest, as disclosed herein, wherein the antibody of interest comprises the heavy chain signal peptide that may comprise one or two or three or four mutations in the heavy chain signal peptide. In another example, there is proved the antibody, as disclosed herein, wherein the antibody of interest comprises the light chain signal peptide comprising one or two or three or four mutations in the light chain signal peptide.

[00118] Mutating may comprise changing the nucleotide sequence of the signal peptide thereby changing the encoded amino acid (missense mutation), deleting an amino acid from the signal peptide sequence or inserting a new amino acid into the signal peptide sequence.

[00119] In a further example, the at least one amino acid is mutated after determining a consensus sequence as indicated above. The consensus sequence is created based on comparison of the amino acid signal peptide sequences within one of the groups clusters created as indicated above. In another example, the amino acid to be mutated within the signal peptide sequence is not an amino acid of the consensus sequence.

[00120] A chimeric antibody, such as a mouse-human, will typically be synthesized from genes driven by the chromosomal gene promoters native to the mouse H and L chain V regions used in the constructs; splicing usually occurs between the splice donor site in the mouse J region and the splice acceptor site preceding the human C region and also at the splice regions that occur within the human C region; polyadenylation and transcription termination occur at native chromosomal sites downstream of the human coding regions.

[00121] In some embodiments, the host cell may comprise a eukaryotic cell or a prokaryotic cell. The eukaryotic cell may comprise but is not limited to a mammalian cell, an avian cell, an insect cell, a fungal cell, a yeast cell and a plant cell. The prokaryotic cell can be a bacterial cell. In a further example, the expression system, as disclosed herein, can comprise, but is not limited, to a Chinese Hamster Ovary Cell (CHO cell), a baculovirus expression vector system, a *Leishmania tarentolae* expression system, a galline (chicken) cell, a yeast expression system, and a plant expression system. The person skilled in the art

would know how to select a suitable expression system for the purpose of producing recombinant antibodies of interest.

[00122] In one example, the antibody of interest whose expression and secretion is monitored to determine the signal peptide which enhances the expression the most, comprises, but is not limited to, a monoclonal antibody, a chimeric antibody, an humanized antibody, derivatives, regions or fragments thereof. In one example, the method provides a monoclonal antibody.

[00123] The antibody (also referred to as immunoglobulins) as defined herein may have an isotype comprising but not limited to IgG, IgM, IgD, IgA and IgE. The isotype of the antibody is determined by the nature of its heavy chain. For example, in mammals there are five types of Ig heavy chains, denoted by the Greek letters gamma (γ), mu (μ), delta (δ), alpha (α) and epsilon (ε). As indicated herein, in mammals there are two types of immunoglobulin light chain, which are called lambda (λ) and kappa (κ).

[00124] Any antibody of interest can be expressed and secreted by the method, as disclosed herein. As will be seen herein, examples of antibody of interest may comprise recombinant therapeutic antibodies such as trastuzumab (Herceptin, CAS number 180288-69-1; Heavy Chain (H) SEQ ID NO: 11; Light Chain (L) SEQ ID NO: 16), bevacizumab (Avastin, CAS number 216974-75-3; (H) SEQ ID NO:12; (L) SEQ ID NO: 17), infliximab (Remicade, CAS number 170277-31-3; (H) SEQ ID NO:13; (L) SEQ ID NO: 18), rituximab (Rituxan, CAS number 174722-31-7; (H) SEQ ID NO: 14; (L) SEQ ID NO: 19) and adalimumab (Humira, CAS number 331731-18-1; (H) SEQ ID NO: 15; (L) SEQ ID NO: 20).

[00125] As indicated above, the at least one vector encoding the recombinant antibody of interest may comprise, for example, the recombinant heavy chain comprising the heavy chain of the antibody of interest and the signal peptide and the recombinant light chain comprising the light chain of the antibody of interest and the signal peptide. For example, a first vector encoding the recombinant light chain of interest and a second vector encoding the recombinant heavy chain of interest may be co-transfected into the host cell.

[00126] Therefore, in one example, there is provided an antibody of interest that may comprise combinations of one heavy and one light chain signal peptide, as defined herein, and above for the heavy chain and for the light chain. In a further example, there is provided the antibody of interest may comprise combinations of one heavy and one light chain signal peptide wherein a signal peptide comprising, but not limited to,

MELGLSWIFLLAILKGVQC (H1, SEQ ID NO: 5), MELGLRWVFLVAILEGVQC (H2, SEQ ID NO: 6), MKHLWFFLLLVAAPRWVLS (H3, SEQ ID NO: 7), MDWTWRILFLVAAATGAHS (H4, SEQ ID NO: 8), MDWTWRFLFVVAAATGVQS (H5, SEQ ID NO: 1), MEFGLSWLFLVAILKGVQC (H6, SEQ ID NO: 9), MEFGLSWVFLVALFRGVQC (H7, SEQ ID NO:3), MDLLHKNMKHLWFFLLLVAAPRWVLS (H8, SEQ ID NO: 10) for the heavy chain and a signal peptide, comprising but not limited to, MDMRVPAQLLGLLLLWLSGARC (L1, SEQ ID NO: 2) and MKYLLPTAAAGLLLLAAQPAMA (L2, SEQ ID NO: 4) for the light chain.

[00127] As indicated above, an antibody of interest may comprise a recombinant therapeutic antibody. Thus in one example, there is provided an antibody, as defined herein, wherein the antibody of interest comprises, but is not limited to, an antibody used for the treatment of breast cancer, colorectal cancer, lung cancer, glioblastoma, kidney cancer and ovarian cancer; an autoimmune disease, rheumatoid arthritis (RA), lymphoma, leukemia, and transplant rejection.

[00128] In one embodiment, there is provided the antibody of interest, as described herein, wherein the antibody of interest used for the treatment of breast cancer is trastuzumab ((H) SEQ ID NO: 11; (L) SEQ ID NO: 16) or bevacizumab ((H) SEQ ID NO:12; (L) SEQ ID NO: 17). In a further embodiment, there is provided the antibody, as described herein, wherein the antibody of interest used for the treatment of colorectal cancer, lung cancer, glioblastoma, kidney cancer and ovarian cancer is bevacizumab ((H) SEQ ID NO: 12; (L) SEQ ID NO: 17).

[00129] In yet another embodiment, there is provided the antibody, as described herein, and above, wherein the antibody of interest used for the treatment of an autoimmune disease and rheumatoid arthritis is infliximab ((H) SEQ ID No. 13; (L) SEQ ID NO: 18) or rituximab ((H) SEQ ID No. 14; (L) SEQ ID NO: 19) or adalimumab ((H) SEQ ID No. 15; (L) SEQ ID NO: 20). In an embodiment, there is provided the antibody, as described herein, wherein the autoimmune disease may comprise, but is not limited to, psoriasis, Crohn's disease, ankylosing spondylitis, psoriatic arthritis, rheumatoid arthritis, and ulcerative colitis. Thus, in one example, there is provided the antibody as described above, wherein the antibody of interest used for the treatment of lymphoma, leukemia and transplant rejection is rituximab ((H) SEQ ID No. 14; (L) SEQ ID NO: 19).

[00130] The present disclosure also provides a nucleotide sequence encoding the antibody of interest, as described herein. For example, the nucleotide sequence comprises, but is not

limited to, the nucleotide sequence encoding for trastuzumab (Herceptin; (H) SEQ ID NO: 21; (L) SEQ ID NO: 26), bevacizumab (Avastin; (H) SEQ ID NO: 22; (L) SEQ ID NO: 27), infliximab (Remicade; (H) SEQ ID NO: 23; (L) SEQ ID NO: 28), rituximab (Rituxan; (H) SEQ ID NO: 24; (L) SEQ ID NO: 29) and adalimumab (Humira; (H) SEQ ID NO: 25; (L) SEQ ID NO: 30).

[00131] The present disclosure also provides a vector comprising the nucleotide sequence, as described herein. As indicated above, there is provided a host cell comprising the nucleotide sequence and/or the vector, as disclosed herein. The vector or nucleotide sequence may be introduced in the host cell, as disclosed above.

[00132] The techniques to raise antibodies of the present invention to small peptide sequences that recognize and bind to those sequences in the free or conjugated form or when presented as a native sequence in the context of a large protein are well known in the art. Such antibodies include murine, murine-human and human-human antibodies produced by hybridoma or recombinant techniques known in the art.

[00133] Antibodies, fragments or derivatives having chimeric H chains and L chains of the same or different variable region binding specificity, can also be prepared by appropriate association of the individual polypeptide chains, according to known method steps.

[00134] With this approach, hosts expressing chimeric H chains (or their derivatives) are separately cultured from hosts expressing chimeric L chains (or their derivatives), and the immunoglobulin chains are separately recovered and then associated. Alternatively, the hosts can be co-cultured and the chains can be allowed to associate spontaneously in the culture medium, followed by recovery of the assembled immunoglobulin, fragment or derivative.

[00135] The hybrid cells are formed by the fusion of a non-human antibody of interest-producing cell, typically a spleen cell of an animal immunized against either natural or recombinant human ligand comprising the antigen, or a peptide fragment of the human epitope protein sequence. Alternatively, the non-human antibody of interest-producing cell can be a B lymphocyte obtained from the blood, spleen, lymph nodes or other tissue of an animal immunized with a ligand comprising the antigen.

[00136] The second fusion partner, which provides the immortalizing function, can be a lymphoblastoid cell or a plasmacytoma or myeloma cell, which is not itself an antibody producing cell, but is malignant. Preferred fusion partner cells include the hybridoma SP2/0-

Ag14, abbreviated as SP2/0 (ATCC CRL1581) and the myeloma P3X63Ag8 (ATCC TIB9), or its derivatives.

[00137] Murine hybridomas, which produce mAbs specific for ligands, are formed by the fusion of a mouse fusion partner cell, such as SP2/0, and spleen cells from mice immunized against purified antigen containing protein, recombinant antigen containing protein, natural or synthetic antigen containing peptides, including peptides including 5 or more amino acids or other biological preparations comprising antigen containing protein. To immunize the mice, a variety of different conventional protocols can be followed. For example, mice can receive primary and boosting immunizations of antigen containing protein.

[00138] The antibody-producing cell contributing the nucleotide sequences encoding the antigen-binding region of the chimeric antibody of the present invention can also be produced by transformation of a non-human, such as a primate, or a human cell. For example, B lymphocyte can be transformed by providing a transforming gene or transforming gene product, as is well-known in the art.

[00139] The cell fusions are accomplished by standard procedures well known to those skilled in the field of immunology. Fusion partner cell lines and methods for fusing and selecting hybridomas, and screening for mAbs are well known in the art.

[00140] The antibody of interest-specific murine or chimeric mAbs of the present invention can be produced in large quantities by injecting hybridoma or transfectoma cells secreting the antibody into the peritoneal cavity of mice and, after appropriate time, harvesting the ascites fluid which contains a high titer of the mAb, and isolating the mAb therefrom. For such *in vivo* production of the mAb with a non-murine hybridoma (e.g., rat or human), hybridoma cells are preferably grown in irradiated or athymic nude mice. Alternatively, the antibodies can be produced by culturing hybridoma or transfectoma cells *in vitro* and isolating secreted mAbs from the cell culture medium or recombinantly, in eukaryotic or prokaryotic cells.

[00141] Thus in an example, there is provided a hybridoma cell line capable of producing the antibody of interest. In some examples, there is provided the amino acid sequence comprising the amino acid sequence for a recombinant therapeutic antibody of interest, a heavy chain signal peptide amino acid sequence fused to the heavy chain amino acid sequence of the antibody of interest, and a light chain amino acid sequence. In case the antibody heavy chain amino acid sequence already comprises an amino acid signal peptide sequence, then said amino acid signal peptide sequence is replaced by the optimized amino

acid signal peptide sequence of the invention. Accordingly, in the specific non-limiting examples below it is understood that the amino acid sequence of the heavy or light chain of the exemplary antibodies of interest should not contain the amino acid sequence of a signal peptide. In case such an amino acid signal peptide sequence comprises the heavy or light amino acid sequence at the C-terminus of the molecule, then said amino acid signal peptide sequence should be replaced by an optimised amino acid signal peptide sequence.

[00142] In a further example, disclosed herein is an amino acid sequence comprising the amino acid sequence for trastuzumab (SEQ ID NO: 11); a heavy chain signal peptide of SEQ ID NO: 1 fused to the heavy chain amino acid sequence of trastuzumab; and a light chain signal peptide of SEQ ID NO: 2. In another example, there is provided an amino acid sequence comprising the amino acid sequence for bevacizumab, a heavy chain signal peptide of SEQ ID NO: 3 fused to the heavy chain amino acid sequence of bevacizumab (SEQ ID NO: 12), and a light chain signal peptide of SEQ ID NO: 5 fused to the light chain amino acid sequence of bevacizumab (SEQ ID NO: 16).

[00143] In a further example, there is provided an amino acid sequence comprising the amino acid sequence for infliximab, a heavy chain signal peptide of SEQ ID NO: 3 fused to the heavy chain amino acid sequence of infliximab (SEQ ID NO: 13), and a light chain signal peptide of SEQ ID NO: 4 fused to the light chain amino acid sequence of infliximab (SEQ ID NO: 18).

[00144] Disclosed herein is an amino acid sequence comprising the amino acid sequence for rituximab, a heavy chain signal peptide of SEQ ID NO: 3 fused to the heavy chain amino acid sequence of rituximab (SEQ ID NO: 14), and a light chain signal peptide of SEQ ID NO: 4 fused to the light chain amino acid sequence of rituximab (SEQ ID NO: 19).

[00145] Also disclosed herein, is an amino acid sequence comprising the amino acid sequence for adalimumab, a heavy chain signal peptide of SEQ ID NO: 3 fused to the heavy chain amino acid sequence of adalimumab (SEQ ID NO: 15), and a light chain signal peptide of SEQ ID NO: 5 fused to the light chain amino acid sequence of adalimumab (SEQ ID NO: 20).

[00146] Accordingly, there is provided a nucleotide sequence encoding an antibody as disclosed herein and above. There is also provided a vector comprising a nucleotide sequence as defined above.

[00147] Whilst the issue of secretion efficiency of heavy and light chains of recombinant antibodies of interest is primarily considered in relation to the signal peptides, the inventors also determined whether the secreted recombinant antibody conserved its activity. Cleavage heterogeneity may arise from nonspecific cleavage of the signal peptide by the signal peptidase. As this occurs within the variable region in the N-terminus of both the heavy and light chains, it may affect the specifity of the secreted antibody in view of antigen recognition. Another heterogeneity issue that may affect recombinant antibody production is the glycan heterogeneity present at the N-glycosylation site of the C_{H2} constant domain of the heavy chains.

[00148] Thus, there is provided a method wherein the secreted recombinant antibody is analyzed by mass spectrometry to detect whether the cleaved signal peptide corresponds to the signal peptide that was fused to the heavy and/or light chain of the antibody of interest. The presence of the signal peptides that were originally fused indicates a homogenous cleavage, which further indicates that the cleavage by the signal peptidase is specific. Thus the secreted antibody maintains its integrity and specificity.

[00149] The invention illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising", "including", "containing", etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention. It must be noted that as used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural reference unless the context clearly dictates otherwise.

[00150] The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or

negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein. Other embodiments are within the following claims and non- limiting examples.

EXPERIMENTAL SECTION

[00151] *Generation of antibody heavy chain and light chain constructs*

[00152] Signal peptides from 173 human antibody heavy chains (SEQ ID NO: 1, 3, 5-10 and 43 to 214) were analyzed by a sequencing alignment. These antibodies include IgG, IgM, IgD, IgA and IgE. Signal peptides from 57 human kappa chains (SEQ ID NO: 2, 4 and 215 to 276) were also analyzed by sequencing alignment. The sequence alignment results are shown in Fig. 1. Based on these results, eight heavy chain signal peptides and two kappa light chain signal peptides (Fig. 2) were chosen and compared for their impact on antibody secretion. Five best-selling therapeutic recombinant antibodies, namely Avastin, Herceptin, Humira, Remicade and Rituxan, were used as model molecules. Each antibody heavy chain and light chain was fused to eight signal peptides and two signal peptides, respectively, by overlapping PCR and ligated into the vector pcDNA3.1 (Invitrogen). Each of the eight heavy chains was co-transfected with two different light chain constructs separately (Fig. 3). Therefore, for each antibody, 16 heavy chain and light chain combinations were transfected into CHO-K1 cells as described below. Combinations of Rituxan with its original heavy chain and light chain signal peptides were also constructed in pcDNA3.1.

[00153] *Transient Transfection of antibody constructs*

[00154] Transient antibody expression using eight different heavy chain signal peptides and two light chain signal peptides were performed. Transfections were carried out in 6-well tissue culture plate using Fugene 6 (Roche, Indianapolis, IN). 2 mL of adherent CHO K1 cells at exponential phase were seeded at density of 3×10^5 cells/mL in 6-well plates 24 h prior to transfection. Duplicate transient transfections for each set of mAb vectors were performed using a Fugene 6 to plasmid ratio of 6 μL:2 μg. 1ug of each light chain and heavy chain bearing plasmids was used in each transfection. To normalize transfection efficiency, a third transfection was carried out in parallel, with an added 0.2 μg of plasmid bearing a gene coding for green fluorescence protein (GFP), pMax-GFP (Amaxa, Gaithersburg, MD). At 48 h post-transfection, supernatant from cultures transfected with only mAb vectors was collected for analysis of mAb concentration using ELISA, and cells from cultures co-

transfected with GFP were collected to measure the fluorescence intensity using a FACS Calibur (Becton Dickinson, Bedford, MA). Results were normalized to the GFP expression and expression level of S1, the first heavy chain signal peptide.

[00155] *Quantification of Secreted Antibody by IgG ELISA Assay*

[00156] The enzyme-linked immunosorbent assay (ELISA) was carried out in 96-well flat bottom MaxiSorp immunoplates (NUNC). The plates were first coated with 50 µL of goat anti-human IgG+IgA+IgM (heavy chain and light chain) antibody (KPL) in PBS at 10 µg/mL for 1 h at 37°C, followed by washing three times with washing buffer (0.1% Tween-20 in PBS). Afterwards, the plates were incubated overnight in blocking solution (3% BSA in PBS) at 4°C. The following day, the plates were washed as above, and 50 µL of standard in-house purified human anti-Rhesus (D) antibody and 1:10 diluted supernatants from the transient transfection, all diluted in dilution buffer (1% BSA in PBS), were added in duplicates and incubated for 1 h at 37°C. Following another round of washes, 50 µL of alkaline phosphatase-conjugated anti-human IgG (Fc-specific) antibody (Sigma) were added to each well and incubated for 1 h at 37°C. After another three washes, 50 µL of SIGMA*FAST*™ p-Nitrophenyl phosphate substrate were added to each well of the plates and incubated for 15 min at room temperature before the absorbance was read at a wavelength of 405 nm using VersaMax ELISA microplate reader (Molecular Devices). A reference wavelength of 620 nm was used.

[00157] *Generation of Avastin heavy chain constructs with hybrid signal peptides*

[00158] Based on the sequence alignment of the 8 IgG heavy chain signal peptides S1 - S8, several conserved amino acid residues were observed. The inventors prepared six Avastin heavy chain constructs with hybrid leader peptides, that had one or more amino acid residues mutated from the parental H7, designated H7a – H7f. These were engineered using QuikChange II XL site-directed mutagenesis kit (Agilent), following the manufacturer's protocol.

[00159] The primer pairs used to generate each construct from the parental H7 (unless specified) were the following:

[00160] H7a, 5′-ccaccatggagtttgggtggagctgggttttcctcg-3′ (SEQ ID NO: 31) and 5′-cgaggaaaacccagctccacccaaactccatggtgg-3′ (SEQ ID NO: 32);

[00161] H7b, 5′-gagtttgggctgagctggctcttcctcgttgctcttttt-3′ (SEQ ID NO: 33) and 5′-aaaaagagcaacgaggaagagccagctcagcccaaactc-3′ (SEQ ID NO: 34);

[00162] H7c 5′-ctgggttttcctcgttgctgcttttagaggtgtccagtgt-3′ (SEQ ID NO: 35) and 5′-acactggacacctctaaaagcagcaacgaggaaaacccag-3′ (SEQ ID NO: 36);

[00163] H7d, 5′-tttttagaggtgtccagtccgaggttcagctggtggag-3′ (SEQ ID NO: 37) and 5′-ctccaccagctgaacctcggactggacacctctaaaaa-3′ (SEQ ID NO: 38);

[00164] H7e, 5′-gccaccatggagtttgggtggagctggctcttcctcgttgctgctttt-3′ (SEQ ID NO: 39) and 5′-aaaagcagcaacgaggaagagccagctccacccaaactccatggtggc-3′ (SEQ ID NO: 40) using the construct H7c as a template;

[00165] H7f, 5′-cttttagaggtgtccagtccgaggttcagctggtggag-3′ (SEQ ID NO: 41) and 5′-ctccaccagctgaacctcggactggacacctctaaaag-3′ (SEQ ID NO: 42) by using construct H7e as a template, where the underlined bases are the mutated targets. Subsequently, two batches of the hybrid constructs and one batch of the Avastin light chain construct were prepared using QIAGEN Plasmid Maxi Kit (Qiagen, Basel, Switzerland).

[00166] *Transient transfection of the hybrid constructs*

[00167] For each pair of Avastin light and heavy chain constructs, 2 mL of CHO-K1 cells were seeded onto each well of a 6-well plate at a concentration of 3 x 10^5 cells per mL. On the following day, transfection was carried out using FuGENE 6 transfection reagent (Roche Applied Science, Rotkreuz, Switzerland) at a reagent: light chain: heavy chain ratio of 6:1:1, following the manufacturer's protocol. Each transfection was performed in duplicate and a third well was transfected with 0.2 μg of GFP construct pmaxGFP® in addition to the Avastin constructs. This served as transfection efficiency control. After 48 h of incubation, the media were collected and centrifuged at 6,000 x *g* for 10 min to remove cell debris, and the supernatant subsequently subjected to ELISA. For the wells transfected with GFP, the cells were trypsinised and subjected to flow cytometry analysis for green fluorescence.

[00168] *Large scale transfection of the antibody constructs for antibody production and purification*

[00169] For each of the antibodies, CHO-K1 cells were seeded into ten T-175 flasks and transfected on the following day with the constructs encoding the best signal peptides as determined by the IgG ELISA assay. Six hours after transfection, the cells were washed with DPBS and the medium replaced with chemically defined serum-free medium. Conditioned medium containing the secreted antibody was collected every 2-3 days over the course of 7 days. The antibodies were then purified with fast protein liquid chromatography (FPLC-

AKTA purifier) system on a HiTrap Protein A HP column (GE Healthcare), equilibrated with 20 mM, pH 7.0 sodium phosphate buffer.

[00170] *NanoLC-MS/MS analysis*

[00171] Diafiltration cartridges (30 kDa; Millipore, Billerica, MA) were used to concentrate 20 µg of each antibody into PBS. Antibodies were then supplemented with 20 mM triethylammonium bicarbonate, pH 8.5, reduced with 30 mM tris(2-carboxyethyl)phosphine (TCEP) at 60°C for 1 h, and cysteine alkylated with 60 mM iodoacetamide at room temperature in the dark for 40 min. Digestion was carried out using sequencing-grade modified trypsin (1:25) (Promega, Madison, WI) overnight at 37°C. Peptide samples were dried down in Savant SpeedVac (Thermo Scientific, Asheville, NC), and resuspended with 25 µl buffer A (0.1% formic acid).

[00172] Nanoscale liquid chromatography (NanoLC) was performed on nanoACQUITY UPLC System (Waters, Milford, MA). Peptide sample (2 µl) was loaded onto Symmetry C18 trapping column, 5 µm, 180 µm x 20mm (Waters) and desalted for 8 min with 2% buffer B (0.1% formic acid in acetonitrile) at 8 µl/min. The trapping column was subsequently switched online to a nanoACQUITY UPLC BEH130 C18 column, 1.7 µm, 75 µm x 150 mm (Waters), and the peptides were separated at a flow of 300 nl/min with a gradient consisting of 60 min 2 – 28% buffer B, 8 min 28 – 40% buffer B and 5 min 97% buffer B.

[00173] Mass spectrometer (MS) detection was performed on a LTQ Orbitrap Velos MS (Thermo Scientific) operating in CID top 10 mode, with nanoelectrospray potential at 1.7 kV. Full scan MS spectra (from m/z 300 – 1,800) were obtained by data dependent acquisition with the resolution set at 60,000. The 10 most intense peptide ions with charge state ≥2 were sequentially fragmented with normalised collision energy of 35 V. Minimum signal threshold for MS/MS was set at 500 counts, activation q value at 0.25 and activation time at 10 ms. Ion trap and orbitrap maximal injection times were set to 100 ms and 10 ms respectively.

[00174] Raw data files were processed by version 1.3.0.339 of Proteome Discoverer (Thermo Scientific) using SEQUEST algorithm, and searched against respective compiled databases consisting of sequentially shortened antibody sequences from the N-terminal.

[00175] *Evaluation of Human Immunoglobulin Signal Peptides for Antibody Secretion in CHO cells*

[00176] Sequences of human Ig heavy chains and kappa chains with complete coding regions were collected from the PubMed database. In total, 173 human Ig heavy chains and

62 human kappa chains were gathered. Majority of the heavy chain signal peptides contain 19 amino acids and all of the kappa light chain signal peptides contain 22 amino acids. A database of signal peptide sequences was built using these heavy chains and light chains. The signal peptides were then clustered according to sequence similarity and the phylogenetic tree is shown in Fig. 1. Based on the phylogenetic tree, eight heavy chain signal peptides (H1-H8) and two kappa light chain signal peptides (L1 and L2) were selected. The amino acid sequences and the DNA sequences of these selected signal peptides are shown in the table in Figure 2. These signal peptides were then assessed for their impact on antibody secretion in CHO-K1 cells.

[00177] The variable regions and the constant regions of the heavy chain and light chain of Herceptin, Avastin, Remicade, Rituxan and Humira were generated based on publicly available information. Each antibody heavy chain was then fused to any one of the eight signal peptides (H1-H8) to generate eight different heavy chain constructs. Each antibody light chain was fused to two signal peptides (L1 & L2) to generate two different light chain constructs (Table in Figure 2).

[00178] For analyzing the impact of signal peptides on the secretion of each antibody, 16 heavy and light chain combinations were transfected into CHO-K1 cells. Duplicate transfections for each pair of heavy and light chain were performed. To normalize transfection efficiency, a third transfection was also performed. In this transfection, in addition to the same pair of heavy and light chain constructs, a construct expressing GFP was also included in the transfection as the control for transfection efficiency as known in the art. The antibody concentrations were determined 2 days after transfection by ELISA. The Figures 3 to 8 (A and C) show the raw data of the ELISA and the expression level of GFP. The Figures labeled A and C assessed the secretion efficiency of L1 and L2 light chain respectively. Within each box, the secretion efficiency of the H1 to H8 was compared. GFP fluorescence was also assessed in each of the samples as a control for transfection efficiency.

[00179] Taking the heavy chain signal peptide 1 (H1) as a reference as explained above, the relative productivity of each antibody was plotted and shown in Fig. 3 to 8 (B and D). As clearly shown, the amount of the antibody in the media is highly dependent on the signal peptide used. Interestingly, the heavy chain signal peptide 7 (H7) resulted in a significantly increased secretion for Avastin, Remicade, Rituxan and Humira. This observation was seen with the use of both light chain signal peptides L1 and L2. For Herceptin, the light chain

signal peptide factored in more significantly with respect to increasing the titre of the antibody. The inventors discovered that several combinations of heavy chain/light chain signal peptides lead to increased secretion of the Herceptin antibody, namely H4/L1, H5/L1, H8/L1, H1/L2 and H7/L2.

[00180] However, the impact of the light chain signal peptides (L1 & L2) could not be conclusively determined in Fig. 6 A to D. As shown in Fig. 6 A and C, the amount of Rituxan antibody produced by H7/L1 was significantly different from that produced by H7/L2. The GFP levels between these two sets of transfection were also very different, suggesting differences in transfection efficiencies. To further confirm these results, combinations of signal peptide H7 with either L1 or L2 were tested for their impacts on the production of Avastin, Remicade and Rituxan, in one single experiment. The signal peptides H1, H5 and H7 in combination with L1 or L2, were assessed for the ability to increase production of Herceptin.

[00181] The results in Fig. 8 correlate well with that of Fig. 6 which allowed the inventors to identify the ideal signal peptide combinations for increased production of specific antibodies. The signal peptide combination for Herceptin, Avastin, Remicade, Rituxan and Humira were H5/L1, H7/L1, H7/L2, H7/L2 and H7/L1 respectively. Based on the results of the transient transfection of these antibodies in the CHO cell expression system, it was determined that Herceptin is produced most efficiently, followed by Rituxan, Avastin, Humira, and lastly Remicade.

[00182] *Assessing the Impact of the Optimized Rituxan Signal Peptides against the Original Peptides*

[00183] Among the five antibodies, 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 peptides are optimal for the secretion of the antibody in CHO cells. Rituxan's original light chain signal peptide is MDFQVQIISFLLISASVIMSRG, while the original heavy chain signal peptide is MGWSLILLFLVAVATRVLS. As indicated above, the H7/L2 is the optimal signal peptides combination for Rituxan (Fig. 8D). In order to compare the effects of the original signal peptide with the optimized peptides, the original heavy and light chain signal peptides were fused to Rituxan and transfected into CHO-K1 cells. As a control, Rituxan H1/L2 constructs were also transfected into CHO-K1 cells. Two days after transfection, conditioned media was

harvested and the titre of Rituxan heavy chain was determined with ELISA. The results in Fig. 4 showed that the optimized signal peptides (H7/L2) resulted in more than a two-fold increase in antibody titre compared to the original signal peptides.

[00184] *Sequence Comparisons of the Eight Ig Heavy Chain Signal Peptides*

[00185] Three domains were defined in the signal peptides, the positively charged N-terminal domain (N-domain), the hydrophobic domain (H-domain) and the polar C-terminal domain (C-domain). The eight heavy chain signal peptides identified in this study show certain common features and yet each of them is quite different from others. As the sequence alignment shown in Fig. 10A, all the signal peptides contain 19 amino acids except for H8, which contains 26. In all the signal peptides except for H3, the amino acid that follows the translation starting methionine (M) is a negatively charged glutamic acid or aspartic acid (E or D), however, it is a lysine (K) at the same position in H3. In fact, in almost all of Ig heavy chains in our database, the second amino acid is either E or D. The second amino acid is a K in almost all signal peptides in cluster 3 represented by H3 (Fig. 1C, bottom panel). Amino acids 7 to 14 form the hydrophobic H-domain and amino acids 15 to 19 form the C-domain. Four signal peptides (H1, H2, H6, H7) terminate with a cysteine (C), whereas others terminate with a serine (S). There is only one negatively charged amino acid (E), and no positively charged amino acid; in the N-domains of the signal peptide clusters 1, 6 and 7 (represented by H1, H6 and H7). Therefore, not all N-domains of the signal peptides are positively charged.

[00186] *Characterization of the Optimal Signal Peptide for Avastin (H7)*

[00187] As discussed above, H7 is shown to be the most optimal signal peptide for the heavy chains in 4 out of 5 antibodies studied. Avastin was used as the model molecule to further optimize the amino acid sequence of H7. A sequence comparison of all signal peptides used in the present invention (H1-H8) showed several highly conserved amino acids: M***W/L**LFLVAA**GVQS/C. (Fig. 10A). The alignment of this highly conserved sequence with H7 (MEFGLSWVFLVALFRGVQC), revealed that three amino acid residues in H7 (underlined) were divergent from the highly conserved sequence. To investigate the functional significance of these amino acids, three H7 mutants were generated by individually mutating each of the amino acids. As shown in Fig. 10B, H7a carries a L5W mutation, H7b carries a V8L mutation and H7c carries a L13A mutation. To compare the difference between C and S at the cleavage site, we generated H7d which carries a C19S

mutation. In H7e, all three amino acids in H7 were mutated to investigate the combinatory effect of these three amino acids. In H7f, four amino acids were mutated, including a C to S mutation (Fig. 10B).

[00188] These mutated signal peptides fused to the Avastin heavy chains were co-transfected with L1 Avastin light chain into CHO-K1 cells. Two days after transfection, the amount of antibody in each conditioned medium was determined by ELISA. Interestingly, the results show that the substitution of any of the three amino acids (H7a, H7b and H7c) dramatically reduced the amount of Avastin heavy chain secreted into the medium (Fig. 10C), suggesting that each of these three amino acids on H7 is important for its function as a signal peptide H7 for Avastin. This is despite the fact that these 3 amino acids were highly conserved in many heavy chain signal peptides. Substituting all three amino acids together (H7e and H7f) further reduced the secretion of Avastin. In contrast, substitution of cysteine to serine (H7d) at the cleavage site does not affect the secretion of the antibody, suggesting that both amino acids are equally effective at the cleavage site.

[00189] *Analysis of Antibody Heterogeneity due to N-terminal Processing and Cleavage of the Signal Peptides*

[00190] In addition to improving secretion efficiency, the inventors also provided some surprising results in view of cleavage heterogeneity problem which occurs as a result of non-specific cleavage of the signal peptide by SPP. This phenomenon leads to the elongation or truncation of the N-terminus of the heavy and light chains which can have a direct effect on the antigen recognition site of the antibody. Such antibodies with variable heterogeneities may not be suitable for biopharmaceutical therapeutics. To determine whether the best signal peptides presently identified can be cleaved efficiently at their expected sites, large scale transfections of each recombinant antibody were performed. Antibodies were harvested from the conditioned media and purified by protein A affinity chromatography. Purified antibodies were subsequently digested by trypsin, and the resultant peptides analysed by mass spectrometry.

[00191] Detection of alternative cleavage sites of the N-terminal peptides of both heavy and light chains of each antibody were carried out by tryptic peptide mapping using LC-MS/MS. N-terminal peptides were identified by high-resolution tandem mass spectrometry (MS/MS), and corresponding peptide precursor peak areas from extracted ion chromatograms (XICs) were used for relative quantification.

[00192] The results obtained for Avastin showed that the correct heavy chain N-terminal peptide EVQLVESGGGLVQPGGSLR (m/z 941.51) accounted for 99.4% of total heavy chain N-terminal peptides detected, while an erroneously processed peptide ESGGGLVQPGGSLR (m/z 657.35), cleaved 5 residues downstream of the expected cleavage site, accounted for 0.6% (Fig. 11A & B). For the light chain, only the correctly processed N-terminal peptide DIQMTQSPSSLSASVGDR (m/z of 939.95) was detected (Fig. 11C), thus suggesting the absence of alternative cleavage site in signal peptide processing. N-terminal peptides from all antibodies were similarly identified and quantified in triplicate mass spectrometry analyses. The results are summarized in Table 2 below and the detailed breakdown is shown in Table 3. As depicted, the optimized signal sequences used in the expression of the antibodies did not give rise to significant cleavage heterogeneity of the signal peptides. Efficacy of N-terminal processing at the expected cleavage site ranged from ~99.2% to 100%, while N-terminal peptides resulting from erroneous cleavage, when present and summed, accounted for less than 1% of total N-terminal peptide population.

Table 2. Proportion of N-terminal peptide(s) quantified by triplicate analyses using mass spectrometry.

Antibody	Sequence (N-terminal)	Percentage (%)
Avastin – LC	...RC▾DIQMTQSPSSLSASVGDR	100.00
Avastin – HC	...QC▾EVQLVESGGGLVQPGGSLR ...QCEVQLV▿ESGGGLVQPGGSLR	99.38 0.62
Herceptin – LC	...RC▾DIQMTQSPSSLSASVGDR ...RCDIQMTQ▿SPSSLSASVGDR ...RCDIQMTQSP▿SSLSASVGDR	99.18 0.71 0.11
Herceptin – HC	...QS▾EVQLVESGGGLVQPGGSLR ...QSEVQ▿LV▿ES▿GG▿GLVQPGGSLR	99.91 0.09
Humira – LC	...RC▾DIQMTQSPSSLSASVGDR	100.00
Humira – HC	...QC▾EVQLVESGGGLVQPGR	100.00
Rituxan – LC	...RC▾QIVLSQSPAILSASPGEK ...RCQIVL▿SQSPAILSASPGEK ...RCQIVLSQ▿SP▿AILSASPGEK	98.95 0.90 0.15
Rituxan – HC	...QC▾QVQLQQPGAELVKPGASVK ...QCQVQ▿L▿QQP▿G▿A▿ELVKPGASVK	99.82 0.18

Rituxan_k01 – LC	...MA▼QIVLSQSPAILSASPGEK ...MAQ▽I▽VLSQSPAI▽LSASPGEK	99.81 0.19
Rituxan_k01 – HC	...LS▼QVQLQQPGAELVKPGASVK ...LSQVQL▽QQ▽PGAELVKPGASVK	99.62 0.38
Remicade – LC	...RC▼DILLTQSPAILSVSPGER	100.00
Remicade – HC	...QC▼EVK↓LEESGGGLVQPGGSMK	*100.00

Light chain (LC); heavy chain (HC); signal peptide processing site: actual N-terminal (▼), erroneous N-terminal (▽); tryptic site (↓). Signal peptide sequences are depicted in smaller font size.
Erroneously processed N-terminal peptides >0.5% are presented individually, while those <0.5% are grouped together where applicable.
*Estimation of proportion of Remicade-HC N-terminal peptides may not be accurate as it relied on the quantification of peptides with one missing trypsin cleavage.

[00193] The translocation of secretory proteins into the lumen of the ER represents a rate limiting step within the classical secretory pathway. Several studies have shown that protein production can be enhanced through the use of alternative signal peptides. This work is the first systematic analysis for identifying the best signal peptides for recombinant antibody production. Advantageously, the strategy for selecting optimal signal peptides begins by generating a database of known antibody signal peptides of human Ig heavy chains and kappa light chains from complete cDNA sequences in the public database. The present exemplary signal peptides collected in the database described herein are only signal peptides of human origin, partly due to the fact that majority of antibody drugs are now either humanized or fully human antibodies.

Table 3. Proportion of N-terminal peptide(s) quantified by triplicate analyses using mass spectrometry (detailed breakdown).

Antibody	Sequence (N-terminal)	Proportion (%)			Average (%)
		1	2	3	
Avastin – LC	...RC▼DIQMTQSPSSLSASVGDR	100.00	100.00	100.00	100.00
Avastin – HC	...QC▼EVQLVESGGGLVQPGGSLR	99.41	99.29	99.43	99.38
	...QCEVQLV▽ESGGGLVQPGGSLR	0.59	0.71	0.57	0.62
Herceptin – LC	...RC▼DIQMTQSPSSLSASVGDR	99.16	99.34	99.06	99.18
	...RC DIQMTQ▽SPSSLSASVGDR	0.68	0.59	0.84	0.71
	...RC DIQMTQSP▽SSLSASVGDR	0.16	0.07	0.10	0.11
Herceptin – HC	...QS▼EVQLVESGGGLVQPGGSLR	99.91	99.92	99.90	99.91
	...QS EVQ▽LVESGGGLVQPGGSLR	0.02	0.02	0.02	0.02
	...QS EVQLV▽ESGGGLVQPGGSLR	0.03	0.02	0.03	0.03
	...QS EVQLVES▽GGGLVQPGGSLR	0.03	0.02	0.04	0.03

	...QS EVQLVESGG▽GLVQPGGSLR	0.01	0.02	0.01	0.01
Humira – LC	...RC▼DIQMTQSPSSLSASVGDR	100.00	100.00	100.00	100.00
Humira – HC	...QC▼EVQLVESGGGLVQPGR	100.00	100.00	100.00	100.00
Rituxan – LC	...RC▼QIVLSQSPAILSASPGEK	98.86	99.09	98.89	98.95
	...RC QIVL▽SQSPAILSASPGEK	0.96	0.77	0.96	0.90
	...RC QIVLSQ▽SPAILSASPGEK	0.10	0.08	0.09	0.09
	...RC QIVLSQSP▽AILSASPGEK	0.08	0.05	0.05	0.06
Rituxan – HC	...QC▼QVQLQQPGAELVKPGASVK	99.80	99.84	99.81	99.82
	...QC QVQ▽LQQPGAELVKPGASVK	0.00	0.00	0.01	0.00
	...QC QVQL▽QQPGAELVKPGASVK	0.02	0.02	0.02	0.02
	...QC QVQLQQP▽GAELVKPGASVK	0.01	0.00	0.00	0.01
	...QC QVQLQQPG▽AELVKPGASVK	0.01	0.00	0.00	0.00
	...QC QVQLQQPGA▽ELVKPGASVK	0.16	0.14	0.15	0.15
Rituxan_k01 – LC	...MA▼QIVLSQSPAILSASPGEK	99.94	99.74	99.74	99.81
	...MA Q▽IVLSQSPAILSASPGEK	0.03	0.02	0.11	0.05
	...MA QI▽VLSQSPAILSASPGEK	0.01	0.05	0.14	0.07
	...MA QIVLSQSPAI▽LSASPGEK	0.01	0.19	0.01	0.07
Rituxan_k01 – HC	...LS▼QVQLQQPGAELVKPGASVK	99.53	99.71	9.61	99.62
	...LS QVQL▽QQPGAELVKPGASVK	0.38	0.25	0.27	0.30
	...LS QVQLQQ▽PGAELVKPGASVK	0.09	0.05	0.13	0.09
Remicade – LC	...RC▼DILLTQSPAILSVSPGER	100.00	92.87	95.91	]100.00
	...RC▼DILLTQSPAILSVSPGERVSFSCR	absent	7.13	4.09	]
Remicade – HC	...QC▼EVKLEESGGGLVQPGGSMK	2.13	100.00	100.00	]*100.00
	...QC EVK↓LEESGGGLVQPGGSMK	97.87	absent	absent	]

Light chain (LC); heavy chain (HC); signal peptide processing site: actual N-terminal (▼); erroneous N-terminal (▽); tryptic site (↓). Signal peptide sequences are depicted in smaller font size.

*The presence of a tryptic site near the N-terminal of Remicade-HC required reliance on an N-terminal peptide with one missed cleavage for identification and quantification. As such, estimation of the proportion of N-terminal peptides is likely to be unreliable, as the majority of the N-terminal peptides would have been processed at the tryptic site, leading to difficulties in identifying any erroneous N-terminal peptides that, if present, would have been at low concentrations to begin with. Note that Remicade sample-1 was digested under normal conditions as described while sample-2 and -3 were digested under conditions that favoured partial digestion (at pH 7.4, room temperature for 4hr).

[00194] Based on sequence similarities, eight heavy chain and two light chain signal peptides were fused to each of the five antibodies for secretion efficiency analysis. Our results showed that some antibodies can tolerate different signal peptides, whereas others are more restricted. The best signal peptide combination identified for Herceptin, Avastin, Remicade, Rituxan, and Humira were H5/L1, H7/L1, H7/L2, H7/L2, and H7/L1, respectively. In the case of Avastin, Remicade and Rituxan, the impacts of heavy chain signal peptides on the production of these antibodies showed similar patterns, with H7 being the most optimal signal peptide (Fig. 2). For some antibodies such as Herceptin, Avastin and

Humira, the antibody productivity was higher when the L1 light chain was produced, whereas for others (Rituxan and Remicade), the L2 light chain resulted in higher productivity of the antibody (Fig. 3 to 10), suggesting that, when used in combination the light chain signal peptide also affects the overall productivity of the antibody.

[00195] As Rituxan is the only antibody with its heavy and light chain signal peptide information available in public database, the secretion efficiency of its native mouse signal peptides was compared with that of human signal peptides and the result showed that the latter improved the production of a chimeric Rituxan by a factor of 2. Advantageously, the method for signal peptide optimization, as disclosed herein, may improve the production of recombinant antibodies.

[00196] In summary, the inventors identified the best signal peptide pairs for each of five therapeutic antibody drugs from a collective pool of 230 human IgG signal peptides. Additionally, the method of the invention may be used to identify the best signal peptides for producing new antibody drugs.

Claims

1. An amino acid sequence of interest comprising a polypeptide that corresponds to a heavy chain of an antibody of interest, a heavy chain signal peptide, a polypeptide that corresponds to a light chain of the antibody of interest, and a light 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), wherein the light chain signal peptide is bound to the polypeptide that corresponds to a light chain of the antibody of interest, and wherein the light chain signal peptide comprises an amino acid sequence selected from the group consisting of MDMRVPAQLLGLLLLWLSGARC (L1, SEQ ID NO: 2) and MKYLLPTAAAGLLLLAAQPAMA (L2, SEQ ID NO: 4), 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 host cell of claim 5, wherein the host cell is a mammalian or bacterial cell.

14. 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), 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.

15. The method of claim 14, wherein the expression system is a host cell.

16. The method of claim 15, wherein the host cell is a eukaryotic cell or a prokaryotic cell.

17. The method of claim 16, 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.

18. The method 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.

19. The method of any one of claims 14 to 18, wherein the antibody is a monoclonal antibody.

20. The method of any one of claims 14 to 19, wherein the antibody has an isotype selected from the group consisting of mammalian IgG, IgM, IgD, IgA and IgE.