

(51) International Patent Classification:
G01N 33/68 (2006.01)(21) International Application Number:
PCT/US2017/016549(22) International Filing Date:
3 February 2017 (03.02.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/291,216 4 February 2016 (04.02.2016) US

(71) Applicant: ONCOBIOLOGICS, INC. [US/US]; 7 Clarke Drive, Cranbury, New Jersey 08512 (US).

(72) Inventors: XIAO, Xiaoyao; 7 Clarke Drive, Cranbury, New Jersey 08512 (US). LI, Linghui; 7 Clarke Drive, Cranbury, New Jersey 08512 (US).

(74) Agents: ELRIFI, Ivor R. et al.; Cooley LLP, 1299 Pennsylvania Avenue, NW, Suite 700, Washington, District of Columbia 20004 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM,

DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))

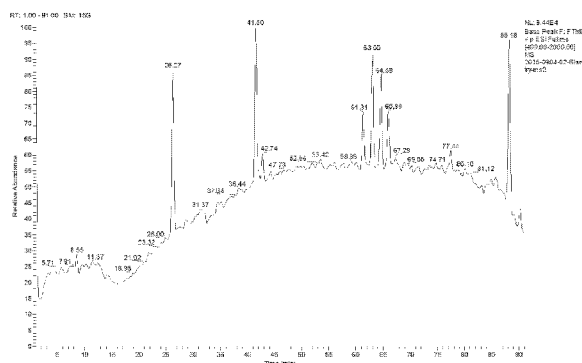
Published:

— with international search report (Art. 21(3))
— with sequence listing part of description (Rule 5.2(a))

(54) Title: METHODS FOR IDENTIFYING AND ANALYZING AMINO ACID SEQUENCES OF PROTEINS

Trypsin Digested

Fig. 1a



(57) Abstract: The disclosure provides methods for determining the biosimilarity of a test protein in relation to a target biologic in which the test protein is digested by two distinct proteases, the resultant digested peptide fragments are analyzed by column chromatography-tandem mass spectrometry to achieve 100% of the amino acid sequence coverage and 100% amino acid sequence accuracy of the test protein.

METHODS FOR IDENTIFYING AND ANALYZING AMINO ACID SEQUENCES OF PROTEINS

RELATED APPLICATIONS

[0001] This application claims priority to U.S. Patent Application No. 62/291,216, filed February 4, 2016, the contents of which are herein incorporated by reference in their entirety.

INCORPORATION OF SEQUENCE LISTING

[0002] The contents of the text file named “ONBI-008001WO_SeqList.txt”, which was created on February 2, 2017 and is 81.7 KB in size, are hereby incorporated by reference in their entirety.

FIELD OF THE DISCLOSURE

[0003] The disclosure relates generally to improved protein sequencing methods that use a reduced incubation time for protease digestion of denatured protein and includes an increased aqueous mobile phase during column chromatography and tandem mass spectrometry (LC-MS/MS) analysis to increase sequence coverage and accuracy up to 100%, as well as improved protein sequencing methods for use in development of therapeutic recombinant proteins and quality control analysis for the manufacture of approved biologics.

BACKGROUND

[0004] Recombinant proteins including recombinant monoclonal antibodies (mAbs) and recombinant versions of natural proteins have been used as reagents for biomedical research, as well as diagnostic and therapeutic agents for humans. One example of recombinant proteins includes biosimilar molecules (also referred to as “biologics”). In order to be approved for use as therapeutic agents for humans, biosimilar molecules must be shown to have an identical amino acid sequence and be very nearly similar in posttranslational modifications, e.g., have “sameness”, to the parent innovator biologic product. Assessing the sameness of a biosimilar molecule is critical because recombinant proteins are complex in nature. Recombinant proteins are engineered using genetically-modified, living organisms (e.g., bacteria, yeast, animal or

human cell lines). The living organisms produce recombinant proteins that are long chain amino acids and/or modified amino acids folded by complex mechanisms. Consequently, recombinant proteins exhibit high molecular complexity and are highly sensitive to changes in the manufacturing process.

[0005] The specificity and effector function of a recombinant protein is highly dependent on the amino acid sequence and the presence or absence of specific modifications. Accordingly, DNA sequencing is routinely used to initially characterize biologics, such as monoclonal antibodies. However, protein-level rearrangements such as subsequent mutations and posttranslational modifications (PTMs) of recombinant proteins, e.g., a monoclonal antibody, are recognized by analysis at the protein level because such rearrangements can only be revealed by protein level analysis. Therefore, amino acid sequencing of monoclonal antibodies is required when the cDNA or the original cell line for the antibody is not available, or when characterization of an amino acid sequence is necessary to verify similarity of the recombinant antibody for approved use as a therapeutic agent, as well as for quality control during manufacture.

[0006] Despite the importance of sequence identification of amino acids in proteins, no methods have been developed for sequencing unknown proteins that provide a high level (100%) of sequence accuracy and coverage. Sequencing recombinant proteins in particular remains a challenge. Two general approaches are used for sequencing proteins using mass spectrometry. In the first, intact proteins are ionized and then introduced to a mass analyzer for mass measurement and tandem mass spectrometry (MS/MS) analysis. This approach is referred to as “top-down” proteomics. In the second, proteins are enzymatically digested into smaller peptides using a protease such as trypsin. Subsequently, the peptides are introduced into a mass spectrometer and identified by peptide mass fingerprinting or tandem mass spectrometry (MS/MS). This latter approach is called “bottom-up” proteomics and uses identification at the peptide level to infer the existence of proteins. Bottom-up proteomics is a preferred process for identifying proteins and characterizing their amino acid sequences, as well as PTMs.

[0007] One well-known method of bottom-up proteomics is Edman degradation. In this method, the amino-terminal residue is labeled and cleaved from a peptide without disrupting the peptide bonds between other amino acid residues. Because Edman degradation proceeds from the N-terminus of the protein, it is unreliable if the N-terminal amino acid has been chemically modified or if it is concealed within the body of the native protein. It also requires guesswork or

a separate procedure to determine the positions of disulfide bridges, as well as peptide concentrations of 1 picomolar or above, for discernible results. Consequently, the Edman process is unsuitable for sequencing proteins longer than 50 amino acids or proteins with PTMs.

[0008] Mass spectrometry-based methods characterize a protein by assembling tandem mass (MS/MS) spectra of overlapping peptides generated from multiple proteolytic digestions of the protein. Each tandem mass (MS/MS) spectrum covers only a short peptide of the target protein. Thus, the key to high coverage protein sequencing is to find spectral pairs from overlapping peptides in order to assemble tandem mass spectrometry (MS/MS) spectra to long ones. However, overlapping regions of peptides may be too short to be confidently identified. Further, automated de novo sequencing methods that rely on interpreting individual tandem mass spectrometry (MS/MS) spectra are limited because these methods typically cannot reconstruct long (8+ amino acid) sequences without misidentifying 1 in 5 amino acids on average. Advances in de novo peptide sequencing have improved sequencing accuracy to over 95%, but at limited sequence coverage, e.g., only 55% sequence coverage. All current per-spectrum de novo sequencing strategies face a tradeoff between sequencing accuracy versus coverage as spectra exhibiting complete peptide fragmentation rarely cover entire target proteins, yet are required to accurately reconstruct full-length peptide sequences.

[0009] An alternative approach to separately sequencing individual spectra is to simultaneously interpret multiple MS/MS spectra from overlapping peptides using another process called Shotgun Protein Sequencing (SPS). SPS has been found to generate sequences that frequently cover 90–95% of the target protein sequence(s) while only misidentifying 1 out of every 20 amino acids on a high resolution MS/MS spectra. SPS has limitations. It generates fragmented sequences that do not singularly cover large regions of the target protein sequences, much less complete proteins. SPS sequences have an average length of 10–15 amino acids and the longest recovered SPS de novo sequence is less than 45 amino acids long.

[0010] In order to be approved for therapeutic use in humans or animals, biosimilars must be shown to be as close to identical, e.g., have “sameness,” to the parent innovator biologic product based on data compiled through clinical, animal, and analytical studies, as well as conformational status. None of the top-down or bottom-up reversed-phase chromatographic methods provides a reliable and simple basis (e.g., 100% sequence accuracy and coverage) for determining biosimilarity of a recombinant protein.

[0011] Therefore, there is a present need for a method for determining analytical similarity or “sameness” of recombinant proteins, e.g., monoclonal antibodies, in comparison to a parent innovator biologic product, wherein the method accurately analyzes amino acid sequence coverage up to 100% with high confidence using a significantly reduced time frame (when compared to well-used protease digestion protocols) for protease digestion of the recombinant protein and enhanced conditions for peptide exposure and consequently increased adherence of peptides to a chromatography column. The method is useful for developing approved biosimilars, as well as quality control analyses during the manufacture of approved biosimilars.

SUMMARY

[0012] The disclosure provides methods for use in evaluating, selecting, and/or manufacturing biologics, including, for example, biosimilars, including interchangeable compositions related thereto (e.g., pharmaceutical preparations). For example, the disclosure provides methods whereby a target protein (e.g., parent innovator biologic product approved under a biologics license application (BLA)) is defined by characteristic signatures, e.g., amino acid sequence, and such signatures are used in the evaluation, identification, and/or manufacture of biologics having the required “sameness” to the target protein for use in diagnostics or approval for use as a therapeutic. The disclosed methods are also useful, for example, for monitoring product changes and controlling product drift that may occur as a result of the use of recombinant technologies with living cells during manufacture of the biologics. The methods include steps for evaluating the similarity of the test protein with a target protein with high reliability on the coverage and accuracy up to 100% of the amino acid sequence of the biologic. For example, the test protein can be evaluated to determine if it has a predetermined level of similarity, or “sameness” with a target protein that is commercially available and/or approved for therapeutic use in humans or animals. This is of particular benefit wherein one or more, or all, of the following conditions is present: (1) the test protein is made by a different method than the target protein or the method used to make the target protein is not known to the maker of the test protein; (2) the test protein is made by an entity having a different marketing approval (or no approval at all) than the entity that makes the target protein; or (3) the test protein was approved in a process that relied on or referred to clinical information regarding the target protein for its approval.

[0013] The disclosure provides a method for determining the biosimilarity of a test protein in relation to a target biologic, the method comprising the steps of: (a) digesting a first sample of a test protein for a first incubation time using a first protease and digesting a second sample of the test protein for a second incubation time using a second protease, wherein the first sample and the second sample are physically separated; (b) applying column chromatography and tandem mass spectroscopy to the first sample under conditions sufficient to enhance binding of small peptides to the column, and generating a sequence of the test protein in the first sample; (c) applying column chromatography and tandem mass spectroscopy to the second sample under conditions sufficient to enhance binding of small peptides to the column, and generating the sequence of the test protein in the second sample, wherein the first sample and second sample are physically separated; (d) identifying the test protein as biosimilar to the target biologic when the test protein comprises 100% sequence identity to the target biologic; and (e) identifying the test protein as not biosimilar to the target biologic when the test protein does not comprise 100% sequence identity to the target biologic.

[0014] In certain embodiments of method for determining the biosimilarity of a test protein in relation to a target biologic of the disclosure, the monoclonal antibody comprises Adalimumab.

[0015] In certain embodiments of method for determining the biosimilarity of a test protein in relation to a target biologic of the disclosure, the first protease is Trypsin. Alternatively, or in addition, in certain embodiments of method for determining the biosimilarity of a test protein in relation to a target biologic of the disclosure, the second protease is Chymotrypsin.

[0016] In certain embodiments of method for determining the biosimilarity of a test protein in relation to a target biologic of the disclosure, the first digestion period is about 0.1 to about 1.0 hour. In certain embodiments, the first digestion period is about 0.1 to about 0.5 hour. In certain embodiments, the first digestion period is about 0.6 to about 1.0 hour. In certain embodiments, the first digestion period is about 0.5 hours.

[0017] In certain embodiments of method for determining the biosimilarity of a test protein in relation to a target biologic of the disclosure, the second digestion period is about 0.1 to about 2.0 hours. In certain embodiments, the second digestion period is about 0.1 to about 1.5 hours. In certain embodiments, the second digestion period is about 1.5 to about 2.0 hours. In certain embodiments, the second digestion period is about 1.5 hours. The disclosure provides a method for determining the biosimilarity of a test protein in relation to a target biologic, the method

comprising the steps of: digesting a first sample of a test protein for a first incubation time using a first protease and a second sample of the test protein for a second incubation time using a second protease, wherein the test protein is digested separately in the first sample and the second sample into peptide sequences; and analyzing the peptide sequences of the first sample separately from the peptide sequences of the second sample using column chromatography to determine 100% of the amino acid sequence coverage and 100% of the amino acid sequence accuracy of the test protein, wherein the column chromatography includes conditions that enhance binding of small peptides to the column.

[0018] In certain embodiments of the method for determining the biosimilarity of a test protein in relation to a target biologic, the test protein is one of a protein, a glycoprotein, a fusion protein, a growth factor, a vaccine, a blood factor, a thrombolytic agent, a hematopoietic protein, a hormone, an interferon, an interleukin-based product, an antibody, a monospecific (e.g., monoclonal) antibody, a pegylated antibody, an antibody drug conjugate, a therapeutic enzyme, a cytokine, or a soluble receptor fragment.

[0019] In certain embodiments of the method for determining the biosimilarity of a test protein in relation to a target biologic, the first protease is Trypsin. Alternatively, or in addition, in certain embodiments of the method for determining the biosimilarity of a test protein in relation to a target biologic, the second protease is Chymotrypsin.

[0020] In certain embodiments of the method for determining the biosimilarity of a test protein in relation to a target biologic, the first protease is Trypsin. In certain embodiments, including those in which the first protease is Trypsin, the first digestion period is about 0.1 to about 1.0 hour. In certain embodiments, including those in which the first protease is Trypsin, the first digestion period is about 0.1 to about 0.5 hour. In certain embodiments, including those in which the first protease is Trypsin, the first digestion period is about 0.6 to about 1.0 hour. In certain embodiments, including those in which the first protease is Trypsin, the first digestion period is about 0.5 hours.

[0021] In certain embodiments of the method for determining the biosimilarity of a test protein in relation to a target biologic, the second protease is Chymotrypsin. In certain embodiments, including those in which the second protease is Chymotrypsin, the second digestion period is about 0.1 to about 2.0 hours. In certain embodiments, including those in which the second protease is Chymotrypsin, the second digestion period is about 0.1 to about 1.5 hours. In certain

embodiments, including those in which the second protease is Chymotrypsin, the second digestion period is about 1.5 to about 2.0 hours. In certain embodiments, including those in which the second protease is Chymotrypsin, the second digestion period is about 1.5 hours.

[0022] In certain embodiments of the method for determining the biosimilarity of a test protein in relation to a target biologic, the target biologic is a commercially available or approved biologic for therapeutic use in humans or animals, a reference listed drug for a secondary approval process, a protein, a glycoprotein, a fusion protein, a growth factor, a vaccine, a blood factor, a thrombolytic agent, a hematopoietic protein, a hormone, an interferon, an interleukin-based product, an antibody, a monospecific (e.g., monoclonal) antibody, a pegylated antibody, an antibody drug conjugate, a therapeutic enzyme, a cytokine, or a soluble receptor fragment. In certain embodiments, the target biologic is one of Adalimumab (Humira®), Bevacizumab (Avastin®), Denosumab (Xgeva®), Cetuximab (Erbix®); Rituximab (Rituxan®); Mabthera®; Campath®; Herceptin®; Xolair®; Prolia®; Vectibix®; ReoPro®; Zenapax®; Simulect®; Synagis®; Remicade®; Mylotarg®; Campath®; Raptiva®; Zevalin®; Erbitux®; Tysabri®; Lucentis®; Soliris®; Cimzia®; Ilaris®; Arzerra®; Bexxar®; Simponi®; Actemra®; Benlysta®; Adcetris®; or Yervoy®. In certain embodiments of the method for determining the biosimilarity of a test protein in relation to a target biologic, the target biologic is Adalimumab (Humira®).

[0023] The disclosure provides a method for analyzing the biosimilarity of a recombinant monoclonal antibody in relation to Adalimumab or its bioequivalent, the method comprising the steps of: determining up to 100% of an amino acid sequence of the recombinant monoclonal antibody by digesting a first sample of the recombinant monoclonal antibody with a first protease and separately digesting a second sample of the recombinant monoclonal antibody with a second protease, wherein the protease digestion steps include incubation times that are no longer than 2 hours collectively; and comparing the amino acid sequence of the recombinant monoclonal antibody to an amino acid sequence of the Adalimumab or its bioequivalent to determine sameness.

[0024] In certain embodiments of the method for analyzing the biosimilarity of a recombinant monoclonal antibody in relation to Adalimumab or its bioequivalent, the sameness comprises 100% similarity between the amino acid sequence of the recombinant monoclonal antibody and the amino acid sequence of Adalimumab or its bioequivalent.

[0025] The disclosure provides a method for manufacturing a pharmaceutical product comprising a recombinant monoclonal antibody, the method comprising the steps of: providing a recombinant monoclonal antibody, wherein the recombinant monoclonal antibody is not approved under a BLA or a supplemental BLA; acquiring input values for the recombinant monoclonal antibody, wherein one or more of the input values are amino acid sequence(s) of a target biologic; acquiring a plurality of assessments made by comparing the input values with a plurality of amino acid sequence(s) for the target biologic, wherein the target biologic is approved under a biologics license application (BLA) or a supplemental BLA; and processing the recombinant monoclonal antibody into a pharmaceutical product if the input values are indistinguishable from target values for said amino acid sequence(s) for the target biologic.

[0026] In certain embodiments of the method for manufacturing a pharmaceutical product comprising a recombinant monoclonal antibody, the recombinant monoclonal antibody is engineered to be a biosimilar to one of Adalimumab (Humira®), Bevacizumab (Avastin®), Denosumab (Xgeva®), Cetuximab (Erbix®); Rituxan®; Mabthera®; Campath®; Herceptin®; Xolair®; Prolia®; Vectibix®; ReoPro®; Zenapax®; Simulect®; Synagis®; Remicade®; Mylotarg®; Campath®; Raptiva®; Zevalin®; Erbitux®; Tysabri®; Lucentis®; Soliris®; Cimzia®; Ilaris®; Arzerra®; Bexxar®; Simponi®; Actemra®; Benlysta®; Adcetris®; or Yervoy®.

[0027] In certain embodiments of the method for manufacturing a pharmaceutical product comprising a recombinant monoclonal antibody, the recombinant monoclonal antibody is engineered to be a biosimilar to Adalimumab (Humira®).

[0028] In certain embodiments of the method for manufacturing a pharmaceutical product comprising a recombinant monoclonal antibody, the input values comprise 100% coverage of the amino acid sequence of the recombinant monoclonal antibody.

[0029] The disclosure provides a method for analyzing up to 100% of the sequence of amino acids of a recombinant monoclonal antibody to determine sameness to a pharmaceutical product, the method comprising the steps of: fragmenting a denatured recombinant monoclonal antibody into discrete peptides by digesting a first sample of the denatured recombinant monoclonal antibody for a first incubation time using a first protease and a second sample of the denatured recombinant monoclonal antibody for a second incubation time using a second protease, wherein the first incubation time is about 0.1 to about 1.0 hours whereafter the first protease is quenched,

and wherein the second incubation time is about 1.0 to about 2.0 hours whereafter the second protease is quenched; analyzing the discrete peptides of the recombinant monoclonal antibody to determine the sequence of amino acids that form the recombinant monoclonal antibody; and comparing the sequence of amino acids of the recombinant monoclonal antibody against a sequence of amino acids of the pharmaceutical product, wherein the pharmaceutical product is approved under a biologics license application (BLA) or a supplemental BLA.

[0030] Methods are also provided for the generation of, or evaluation of, a predetermined plurality of target values for the generation of, or evaluation of, a signature, e.g., amino acid sequence, for a test protein, and/or use or application of such information to acquire a sameness/identity value describing the relationship (e.g., structural relationship) between the test protein and the target protein. In some instances, a sameness/identity value can be used to evaluate, identify, and/or produce (e.g., manufacture) a test protein. In some instances, a sameness/identity value is a specification for release of a test protein. Accordingly, disclosed herein are methods useful for evaluating, identifying, and manufacturing an approved biologic.

[0031] The method optionally includes a preparation step of separating a test biologic preparation from other isoforms or variants of the test biologic, as well as by-products from manufacturing the same, in a highly purified preparation, e.g., a test protein preparation, wherein the test biologic is not approved under a biologics license application (BLA), a supplemental BLA, or equivalents thereof; and then processing the highly purified test biologic preparation using input values for one or more amino acid sequences for a target biologic.

[0032] In one embodiment, the test protein is determined to have an amino acid sequence (e.g., a primary amino acid sequence) that is identical or nearly identical to the target protein amino acid sequence (e.g., 100% match with 0.5% tolerance for sequence variance due to translational errors), and the target protein is approved under a BLA, a supplemental BLA, or equivalents thereof.

[0033] In one embodiment, the method comprises the steps of: (1) producing an enriched test protein preparation, wherein the test protein may or may not be approved under a biologics license application (BLA), a supplemental BLA, or equivalents thereof; and (2) processing the test protein preparation to determine that the amino acid sequence is indistinguishable from of the amino acid sequence a target protein, wherein the test protein has an amino acid sequence (e.g., a primary amino acid sequence) that is up to 100% identical to the target protein amino

acid sequence, and wherein the target protein is approved under a BLA, a supplemental BLA, or equivalents thereof, thereby manufacturing a pharmaceutical product comprising a protein, e.g., a monoclonal antibody (mAb).

[0034] In an embodiment, the target protein is an antibody, e.g., a monoclonal antibody, a humanized antibody, or a human antibody. In alternative embodiments, the target protein can be an antibody conjugated with polyethylene glycol (PEG) polymer chains, e.g., a pegylated antibody. For a pegylated monoclonal antibody, depending on the degree of pegylation and the range of mass size of pegylation, the methods of the disclosure can be adopted to sequence the amino acids of the monoclonal antibody after a step of releasing PEG prior to sample preparation for peptide mapping. In further embodiments, the target protein can be an antibody-drug conjugate (ADC) complex molecule composed of an antibody, e.g., whole mAb or an antibody fragment such as a single-chain variable fragment (scFv)) that is linked, via a stable, chemical linker with labile bonds, to a biological active cytotoxic (anticancer) payload or drug. In an ADC complex molecule, wherein the reactive residue with the drug is modified, the methods of the disclosure can be used to map the drug conjugation sites by including the molecular weight of the drug as a modification in the sequence database.

[0035] For example, the target protein can be selected from the products marketed as Adalimumab (Humira®), Bevacizumab (Avastin®), Denosumab (Xgeva®), Cetuximab (Erbix®); Rituxan®; Mabthera®; Campath®; Herceptin®; Xolair®; Prolia®; Vectibix®; ReoPro®; Zenapax®; Simulect®; Synagis®; Remicade®; Mylotarg®; Campath®; Raptiva®; Zevalin®; Erbix®; Tysabri®; Lucentis®; Soliris®; Cimzia®; Ilaris®; Arzerra®; Bexxar®; Simponi®; Actemra®; Actemra®; Benlysta®; Adcetris®; or Yervoy®, as well as other biologics.

BRIEF DESCRIPTION OF THE DRAWINGS

[0036] Additional aspects, features, and advantages of the disclosure, both as to its methods and use, will be understood and become more readily apparent when the disclosure is considered in light of the following description of illustrative embodiments made in conjunction with the accompanying drawings, wherein:

[0037] **Figures 1A-B** are a pair of graphs illustrating chromatographic profiles of trypsin-digested and chymotrypsin-digested chromatography matrix, respectively, that were run for

specificity. The matrix showed no hit on any target amino acid sequence on Sequence Discoverer. Small peaks on the matrix chromatograms show system peaks and enzyme peaks.

[0038] Figures 2A-D are a series of alignments illustrating sequence coverage for trypsin-digested heavy chain (FIG. 2A (SEQ ID NO: 1, with potential modifications (SEQ ID NO: 2)), chymotrypsin-digested heavy chain (FIG. 2B (SEQ ID NO: 3, with potential modifications (SEQ ID NO: 4)), trypsin-digested light chain (FIG. 2C (SEQ ID NO: 5, with potential modifications (SEQ ID NO: 6)), and chymotrypsin-digested light chain (FIG. 2D (SEQ ID NO: 7, with potential modifications (SEQ ID NO: 8)) of the ONS-3010 reference standard (*Adalimumab*), respectively. These figures demonstrate that the method of the disclosure is capable of 100% sequence coverage.

[0039] Figures 3A-B are a pair of graphs illustrating chromatographic profiles for trypsin-digested and chymotrypsin-digested ONS-3010 reference standards, respectively, that were run for specificity.

[0040] Figures 4A-D are a series of alignments illustrating 100% sequence coverage for trypsin-digested heavy chain (FIG. 4A (SEQ ID NO: 9, with potential modifications (SEQ ID NO: 10)), chymotrypsin-digested heavy chain (FIG. 4B (SEQ ID NO: 11, with potential modifications (SEQ ID NO: 12)), trypsin-digested light chain (FIG. 4C (SEQ ID NO: 13, with potential modifications (SEQ ID NO: 14)), and chymotrypsin-digested light chain (FIG. 4D (SEQ ID NO: 15, with potential modifications (SEQ ID NO: 16)) of the positive control Adalimumab (Humira®) (sample test ID H35), respectively. These figures confirm that the target sequence is an accurate amino acid sequence for Adalimumab (Humira®).

[0041] Figures 5A-B are a pair of graphs illustrating chromatographic profiles for trypsin-digested and chymotrypsin-digested positive control Adalimumab (Humira®) (sample test ID H35), respectively.

[0042] Figures 6A-D are a series of alignments illustrating sequence coverage for trypsin-digested heavy chain (FIG. 6A (SEQ ID NO: 17, with potential modifications (SEQ ID NO: 18)), chymotrypsin-digested heavy chain (FIG. 6B (SEQ ID NO: 19, with potential modifications (SEQ ID NO: 20)), trypsin-digested light chain (FIG. 6C (SEQ ID NO: 21, with potential modifications (SEQ ID NO: 22)), and chymotrypsin-digested light chain (FIG. 6D (SEQ ID NO: 23, with potential modifications (SEQ ID NO: 24)) of the negative control Rituximab

(Rituxan®) (sample test ID M6), respectively. This demonstrates that the method of the disclosure is capable of identifying sequences accurately.

[0043] Figures 7A-B are a pair of graphs illustrating chromatographic profiles of trypsin-digested and chymotrypsin-digested negative control Rituximab (Rituxan®) (sample test ID M6), respectively.

[0044] Figure 8 is a schematic diagram illustrating the theoretical amino acid sequences of the light chain ((SEQ ID NO: 25) and the heavy chain of Adalimumab (Humira®) ((SEQ ID NO: 26).

DETAILED DESCRIPTION

[0045] As used herein, the term “biologic” (singular or plural) refers to peptide and protein products. For example, biologics include naturally-derived or recombinant products expressed in cells, such as, e.g., proteins, glycoproteins, fusion proteins, growth factors, vaccines, blood factors, thrombolytic agents, hormones, interferons, interleukin-based products, monospecific (e.g., monoclonal) antibodies, therapeutic enzymes. Biologics may be approved under a biologics license application (BLA), under Section 351(a) of the Public Health Service (PHS) Act, whereas biosimilar and interchangeable biologics referencing a BLA as a reference product are licensed under Section 351(k) of the PHS Act. Section 351 of the Public Health Service (PHS) Act is codified as 42 U.S.C. 262. Other biologics may be approved under Section 505(b)(1) of the Federal Food and Cosmetic Act, or as abbreviated applications under Sections 505(b)(2) and 505(j) of the Hatch Waxman Act, wherein Section 505 is codified as 21 U.S.C. 355.

[0046] As used herein, the term “isoform” (singular or plural) refers to any of several different forms of the same protein, arising from either single nucleotide polymorphisms, differential splicing of mRNA, or post-translational modifications (e.g., sulfation, glycosylation, etc.).

[0047] As used herein, the term “antibody” (singular or plural) refers, in the broadest sense, to monoclonal antibodies (including full length monoclonal antibodies) of any of the classes IgG, IgM, IgD, IgA, and IgE, as well as antibody fragments that exhibit a desired biological activity. The phrase “antibody fragments” refers to a portion of a full-length antibody, generally the antigen binding or variable region thereof. Examples of antibody fragments include Fab, Fab',

F(ab')₂, and Fv fragments; diabodies; linear antibodies; single-chain antibody molecules; and multi-specific antibodies formed from antibody fragments.

[0048] As used herein, the term “monoclonal antibody” (singular or plural) refers to antibodies that are highly specific, being directed against a single antigenic epitope. Alternatively, the term “monoclonal antibody” refers to an antibody produced from a single spleen cell clone. In a non-limiting example, a monoclonal antibody can be a fully humanized antibody, i.e., both its variable and constant region are derived from a human source.

[0049] As used herein, the term “approval” refers to the procedure by which a regulatory entity, e.g., the USFDA, approves a candidate for therapeutic or diagnostic use in humans or animals. As used herein, a primary approval process is an approval process which does not refer to a previously approved protein, e.g., it does not require that the protein being approved have structural or functional similarity to a previously approved protein, e.g., a previously approved protein having the same primary amino acid sequence or a primary amino acid sequence. In embodiments, the primary approval process is one in which the applicant does not rely, for approval, on data, e.g., clinical data, from a previously approved product. Exemplary primary approval processes include, in the United States, a Biologics License Application (BLA), or supplemental Biologics License Application (sBLA), a new drug application (NDA) under Section 505(b)(1) of the Federal Food and Cosmetic Act, and, in Europe, an approval in accordance with the provisions of Article 8(3) of the European Directive 2001/83/EC, or an analogous proceeding in other countries or jurisdictions.

[0050] As used herein, the term “glycoprotein” refers to an amino acid sequence that includes one or more oligosaccharide chains (e.g., glycans) covalently attached thereto. Exemplary amino acid sequences include peptides, polypeptides, and proteins. Exemplary glycoproteins include glycosylated antibodies and antibody-like molecules (e.g., Fc fusion proteins). Exemplary antibodies include monoclonal antibodies and/or fragments thereof, polyclonal antibodies and/or fragments thereof, and Fc domain containing fusion proteins (e.g., fusion proteins containing the Fc region of IgG1, or a glycosylated portion thereof). A glycoprotein preparation is a composition or mixture that includes at least one glycoprotein.

[0051] As used herein, the phrase “target biologic”, e.g., target protein, refers to a commercially available, or approved, biologic which defines or provides the basis against which a test biologic is measured or evaluated. In embodiments, a target biologic is commercially available for

therapeutic use in humans or animals. In other embodiments, the target biologic is approved for use in humans or animals by a primary approval process. In further embodiments, the target biologic is a reference listed drug for a secondary approval process. An exemplary target protein is an antibody, e.g., humanized or human antibody. Other target proteins include glycoproteins, cytokines, hematopoietic proteins, soluble receptor fragments, and growth factors.

[0052] As used herein, the term “evaluating” refers to reviewing, considering, determining, assessing, measuring, and/or detecting the presence, absence, level, and/or ratio of one or more parameters in a test protein and/or target biologic to provide information pertaining to the one or more parameters. In some instances, evaluating a glycoprotein preparation includes detecting the presence, absence, level, or ratio of one or more points of similarity between a test protein and a target biologic.

[0053] As used herein, the term “analyzing” refers to performing a process that involves a physical change in a sample or another substance, e.g., a starting material. Exemplary changes include making a physical entity from two or more starting materials, shearing or fragmenting a substance, separating or purifying a substance, combining two or more separate entities into a mixture, or performing a chemical reaction that includes breaking or forming a covalent or non-covalent bond. Analyzing a sample can include performing an analytical process which includes a physical change in a substance, e.g., sample, analyte, or reagent (sometimes referred to herein as “physical analysis”), performing an analytical method, e.g., a method which includes one or more of the following: separating or purifying a substance, e.g., an analyte, or a fragment or other derivative thereof, from another substance; combining an analyte, or fragment or other derivative thereof, with another substance, e.g., a buffer, solvent, or reactant; or changing the structure of an analyte, or a fragment or other derivative thereof, e.g., by breaking or forming a covalent or non-covalent bond, between a first and a second atom of the analyte; or by changing the structure of a reagent, or a fragment or other derivative thereof, e.g., by breaking or forming a covalent or non-covalent bond, between a first and a second atom of the reagent.

[0054] As used herein, the phrase “input value” refers to a value associated with a parameter of a test biologic. The value can be qualitative, e.g., present, absent, intermediate, or the value can be quantitative, e.g., it can be a numerical value such as a single number, or a range, for a parameter.

General Method of the Disclosure

[0055] The methods of the disclosure can be used for analytically determining similarity of a recombinant protein (e.g., test protein) to a parent innovator biologic product (e.g., target protein) throughout the development and manufacture of biosimilar therapeutic molecules.

[0056] Non-limiting applications of the method include use in determining the similarity of recombinant proteins to biologic products including, but not limited to, Adalimumab (Humira®), Bevacizumab (Avastin®), Denosumab (Xgeva®), Cetuximab (Erbix®); Rituximab (Rituxan®); Mabthera®; Campath®; Herceptin®; Xolair®; Prolia®; Vectibix®; ReoPro®; Zenapax®; Simulect®; Synagis®; Remicade®; Mylotarg®; Campath®; Raptiva®; Zevalin®; Erbix®; Tysabri®; Lucentis®; Soliris®; Cimzia®; Ilaris®; Arzerra®; Bexxar®; Simponi®; Actemra®; Benlysta®; Adcetris®; and Yervoy®, as well as other biologics.

[0057] The method provides an analysis for evaluating the primary structure of a test protein, either for analyzing a test protein or a target protein, and/or for analyzing the test protein in comparison to a target protein. This method provides up to 100% amino acid sequence coverage and accuracy.

[0058] In an embodiment, a test protein, such as a recombinant protein that can include variants, can be analyzed. In a non-limiting alternative embodiment, a test protein can optionally be initially purified by column chromatography, e.g., HPLC, to separate the biologic from basic and acidic variants of the biologic and/or any other by-products of manufacture of the biologic, e.g., enzymes, cells, and cellular debris, etc. The biologic, its variants, and related manufacturing by-products can be processed through a chromatographic system, e.g., cation-exchange column, that is capable of high-efficiency, high resolution separation of closely eluting proteins.

[0059] In a further non-limiting example, the disclosure provides methods for identifying and confirming the primary structure of the test protein – a monoclonal antibody (e.g., ONS-3010 a biosimilar to the monoclonal antibody *Adalimumab* (Humira®)) - for characterization.

[0060] In accordance with the general methods of the disclosure, the test protein and/or target protein, e.g., monoclonal antibody (mAb), is denatured, reduced, alkylated, and spun down through a 10 kDa centrifuge filter. This involves optimal digestion and complete sequence coverage by solubilization of the test protein, denaturation of the test protein, and disulphide bond reduction. Discrete peptides are selectively fragmented by using trypsin (Try) and chymotrypsin (Chy). This peptide mixture is injected onto a reverse-phase ultra-high

performance liquid chromatography (RP-UPLC) system to obtain a unique profile (peptide map). The exact mass charge ratios (m/z) of the peptides are determined by full scan on a high resolution mass spectrometer. The peptide is then broken into ion fragments for MS/MS amino acid sequence analysis. The MS/MS data can be analyzed by using, for example, Proteome Discoverer software against an *Adalimumab* amino acid sequence database to identify peptide sequence. The similarity of amino acid sequences between reference product or target product, and test product is reported.

[0061] With its application of identifying sequences, the Proteome Discoverer software extracts relevant MS/MS spectra from the “.raw” file and determines the precursor charge state and the quality of the fragmentation spectrum. The SEQUEST search algorithm correlates experimental MS/MS spectra through comparisons to theoretical MS/MS spectra from protein databases. The Proteome Discoverer uses a probability-based scoring system to rate the relevance of the best matches found by the SEQUEST algorithm. The algorithm color codes the amino acid table to show the portion of the corresponding peptide sequence that is identified. Green, yellow, and pink indicate high, medium, and low confidence, respectively. No color means no hit on the peptide. The Protein Results View highlights the fragment ions in a peptide MS/MS spectrum that match predicted fragment masses. Specifically in Figures 2A-2D, 4A-4D, and 6A-6D of this disclosure, high, medium and low confidence hits are indicated by a solid line (____), a long broken line (-----), and a short broken line (.....) instead of color, respectively.

[0062] In a specific example, two separate aliquots of the monoclonal antibody (target protein, e.g., *Adalimumab* and/or bioequivalents) are prepared at a concentration of ≥ 3.0 mg/mL by transferring water into a 1.5 mL polypropylene centrifuge tube and adding 300 μ g of sample into the tube. Negative controls (e.g., formulation buffer or HPLC-grade water) and positive controls (e.g., reference standard) are also prepared as a reference. The peptide standard mixture used as an instrument system suitability control is a 20 μ g/mL HPLC peptide standard mixture prepared by adding 2.5 mL of Mobile Phase A (*see below*) to one vial of standard mixture (*Sigma, Cat # H2016-1VL*).

[0063] Each aliquot of monoclonal antibodies is denatured by adding a mixture of 500 μ L of 8N Guanidine HCl (Fisher, Cat # 24115), 40 μ L of 2.5 M Tris base (3.03 g of Tris base in HPLC water to a final volume of 10 mL), and 20 μ L of 1N HCl (Fisher Scientific, Cat # SA48-1 or equivalent) into each tube.

[0064] A stabilizing reagent, e.g., Dithiothreitol (DTT), is added to each sample of the denatured protein under conditions that promote the disruption of disulfide bonds of the denatured protein. Each sample of the denatured monoclonal antibodies is then reduced by adding 20 μ L of 25 mg/mL of Dithiothreitol (DTT) (Bio-Rad, Cat # 161-0611) (e.g., 25 mg of DTT in HPLC Grade water (Fisher Scientific, Cat # W5-4) to a final volume of 1.0 mL) to each sample. The samples are incubated separately at about $37 \pm 2^\circ$ C for about 0.5 hour.

[0065] At the end of the incubation, each sample undergoes alkylation by adding 8 μ L of 200 mg/mL sodium iodoacetate (Sigma, Cat # I2512) (e.g., 200 mg of sodium iodoacetate mixed with HPLC water to a final volume of 1 mL) to each sample, and then the samples are incubated in the dark at ambient temperature for about 15 minutes.

[0066] At the end of the alkylation incubation, each sample is desalted. The desalting process involves washing each sample in a Millipore Biomax-10 kDa Ultrafree 0.5 Centrifuge filter. The filter is initially wetted by centrifuging about 300 μ L ammonium bicarbonate centrifuged at 10,000 rpm for 5 minutes. Each sample is transferred to the surface of a pre-wetted filter and is then washed with 300 μ L of 0.1 M ammonium bicarbonate and centrifuged at 10,000 rpm for 10 minutes. The wash step can be repeated up to 2 or more times. The final wash involves centrifugation for about 10-13 minutes at 10,000 rpm so that each sample has a final volume of about 100 μ L.

[0067] The desalted samples are then enzymatically digested by adding a different protease, e.g., trypsin and chymotrypsin, to the samples at optimized incubation conditions that include a reduced time frame for digestion, e.g., up to 0.5 hour for trypsin and up to 1.5 hours for chymotrypsin. The reduced incubation times are shorter in duration than traditional incubations times, e.g., 2-4 hours and even up to 18 hours. The shorter digestion time period provides more instances of specific miscleavage of amino acids so that the glycoprotein produced by digestion comprises longer peptides, rather than shorter peptides produced by longer digestion time. The digestion occurs in two desalted samples individually, namely, one sample is digested using trypsin that is quenched after passage of a first incubation time period, e.g., 0.5 hour, and then a second sample is digested using chymotrypsin that is quenched after passage of a second incubation time, e.g., 1.5 hours. This use of chymotrypsin protease digestion supports the coverage for a small peptide EAK in light chain and a small peptide SLR in heavy chain to archive 100% sequence coverage. During the digestion, the polypeptide chain of the denatured

protein is cut into shorter fragments as the enzymes split peptide bonds that link amino acid residues in the denatured protein.

[0068] The first sample undergoes proteolytic digestion with trypsin. For example, trypsin (sequence grade modified, Promega, Cat # V5111, 20 µg) can be reconstituted in 20 µL of reconstitution buffer. 15 µL of reconstituted trypsin is added to each sample to reach a ratio of trypsin-to-sample ratio of about 1:20. The sample with trypsin added is incubated at 37°C for 0.5 hour and then about 5 µL of 10% Formic acid (v/v) (Thermo, Product # 28905 or equivalent, e.g., *100 µl of formic acid in 900 µl of HPLC grade water*) is added to each sample to quench the enzymatic digestion in preparation for the second digestion.

[0069] The second sample undergoes proteolytic digestion with chymotrypsin. For example, chymotrypsin (Promega Chymotrypsin Sequencing Grade, 25 µg, Cat # V1062) can be reconstituted into 20 µL of 1 mM of HCl in HPLC-grade water (Fisher Scientific, Cat # SA48-1 or equivalent) reconstitution buffer. About 15 µL of reconstituted chymotrypsin is added to each sample at a chymotrypsin-to-sample ratio of about 1:16. The sample with chymotrypsin added is incubated at 37°C for 1.5 hours and then about 5 µL of 10 % Formic acid (v/v) is added to each sample to quench the enzymatic digestion.

[0070] The samples of digested monoclonal antibodies are then run separately through UPLC-MS/MS for analysis under conditions that promote adsorption of the peptides to the column including smaller chain peptides that are less likely to be bound under normal conditions. For example, the UPLC column can be an UPLC column (Waters BEH300 C18, 2.1x100mm, 1.7 µm, Cat# 186003555) having a pre-column (VanGuard, BEH300 C18, 2.1x5mm, 1.7 µm, Cat# 186003975). Samples are injected into the column at a volume of about 10 µL having a protein concentration of about 3 µg/µL at a temperature of about 45°C. After the sample is loaded on the column, a gradient consisting of Mobile Phase A (0.1% TFA in water (*Optima LC/MS*, *Fisher Scientific*, Cat# *LS119*)) and Mobile Phase B (0.085% TFA in 95% *Acetonitrile (ACN)* (v/v)) (*HPLC grade*, *Fisher Scientific*, Cat # *A-998-4 or equivalent*), and are passed through the column at a flow rate of about 400 µL/min. The UPLC system parameters are as follows: the autosampler is set at about 4° C; data rate 20 pts/sec; PMT gain at 1; and PDA wavelength 210 nm and 280 nm.

[0071] The gradient for the peptide standard mixture is run as follows:

Time (min)	% Mobile Phase A	% Mobile Phase B
0	100	0
7	67	33
7.5	59	41
8	30	70
8.5	30	70
9	100	0
10	100	0

[0072] The gradient for a sample is run as follows:

Time (min)	% Mobile Phase A	% Mobile Phase B
0	100	0
10	100	0
15	98	2
75	67	33
85	60	40
90	30	70
95	30	70
96	100	0
98	100	0

[0073] The sample batch may be set as follows:

Description	Method	# Injections
Equilibration Run	Equilibration	1
Water	Peptide std.	2
Peptide Standard	Peptide std.	2
Water	Peptide std.	2
Trypsin-Buffer Blank	Peptide 98 min run-MS/MS	1
Water	Peptide std.	2
Trypsin-ONS3010- Ref Std.	Peptide 98 min run-MS/MS	1
Water	Peptide std.	2

Trypsin-ONS3010-Sample	Peptide 98 min run-MS/MS	1
Water	Peptide std.	2
Chymotrypsin-Buffer Blank	Peptide 98 min run-MS/MS	1
Water	Peptide std.	2
Chymotrypsin-ONS3010-Ref Std.	Peptide 98 min run-MS/MS	1
Water	Peptide std.	2
Chymotrypsin-ONS3010- Sample	Peptide 98 min run-MS/MS	1
Water	Peptide std.	2
Peptide Standard	Peptide std.	1
Water	Peptide std.	2
Water	Shutdown	1

[0074] Amino Acid Sequence Coverage Search Proteome Discoverer 1.3 was used for data analysis of each UPLC run.

[0075] Data acceptance, record, and report included exact mass calibration followed by application of the peptide standard layout of Xcalibur onto peptide standard injections. Record 5 peaks of RT of peptide standard injection on MS chromatograms, calculate % RSD of each peak RT individually. Record exact mass of selected peptide at m/z 532 and calculate the mass accuracy for all injections.

[0076] The peptide standard layout was as follows:

[0077]
$$\text{Mass Accuracy (ppm)} = \frac{|\text{Exact Mass} - \text{Theoretical Mass}|}{\text{Exact Mass}} \times 10^6$$

EXAMPLES

Example 1: ONS-3010/ONS-1045

[0078] ONS-3010 (lot #X1302-BDS-O) and ONS-1045 (lot #1407104501) were prepared in accordance with the disclosed methods by implementing the reduced digestion time (0.5 hour for trypsin) and by testing the digested peptides with mass spectroscopy using the 98 minute methodology to allow time for peptides to bind to the column. Purified samples of ONS-3010

and ONS-1045, which were analyzed using an 88 minute methodology, were analyzed using the 98 minute methodology, and the results were compared against results obtained using the 88 minute methodology. A summary of the improved sequence coverage is provided in **Table 1**.

[0079] Table 1: Sequence Coverage Improvements with 0.5 hour Trypsin Digestion and 98 Minute UPLC Method

ONS-3010	Heavy Chain	Light Chain
0.5 hour Trypsin digestion + 98 min instr method	100.00 %	98.60 %
1 hour Trypsin digestion + 88 min instr method	95.79 %	97.20 %
ONS-1045	Heavy Chain	Light Chain
0.5 hour Trypsin digestion + 98 min instr method	100.00 %	98.60 %
1 hour Trypsin digestion + 88 min instr method	96.91 %	98.60 %

[0080] Trypsin digested ONS-3010 heavy chain sequence coverage increased from 95.79% to 100% and trypsin digested ONS-1045 heavy chain sequence coverage increased from 96.91% to 100%.

[0081] ONS-1045 samples #1407104501 and #1408104502 were analyzed according to the methods of the disclosure using mass spectroscopy implementing the 98 minute UPLC methodology. Samples were frozen at -80°C and re-injected on the same instrument with the same mobile phases, but with the 98 minute UPLC method, which includes an extra 10 minutes of 0.4 mL/min 100% mobile phase A at the beginning of the gradient. The prior method implemented an 88 minute run without an extra 10 minutes of mobile phase A at the start of the run. According to the results of reinjections, the 98 minute UPLC method increased amino acid sequence coverage of the trypsin digested samples. Furthermore, the peptide CK was detected as cleaved peptide CKVSNK, but was missed in the 88 minute method.

[0082] To evaluate the specificity of the assay, matrix, positive control (*Humira*), and negative control (*Rituxan*) were used for specificity determination. Matrix showed no hit from target

sequence. Positive control had 100% sequence coverage against *Adalimumab* sequence on both heavy chain (HC) and light chain (LC) after combining the results of the analysis of the trypsin and chymotrypsin digested samples. Negative control showed no sequence coverage on Fab region against *Adalimumab* sequence. Some peptides were detected in negative control because the constant region amino acid sequence of different IgG1 is the same. Certain figures herein show either sequence coverage results from Proteome Discoverer or total ion chromatograms. The general sequence coverage of each sample is recorded in **Table 2**.

[0083] Table 2: Sequence Coverage Summary for Specificity

	Heavy Chain			Light Chain		
	Try	Chy	Overall	Try	Chy	Overall
Matrix	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
ONS-3010 Ref Std.	100.0%	100.0%	100.0%	98.6%	89.3%	100.0%
Positive Control – Humira (H35)	98.9%	84.1%	100.0%	98.6%	89.3%	100.0%
ONS-3010 Sample	98.9%	87.8%	100.0%	98.6%	89.3%	100.0%
Negative Control – Rituxan (M6)	69.4%	50.8%	72.1%	48.6%	45.8%	50.0%

[0084] To evaluate the reproducibility of the assay, an ONS-3010 reference standard was tested in each run and repeated for 3 runs. The heavy chain and light chain sequence coverage of the reference standards and samples from both trypsin and chymotrypsin digestions are recorded in **Table 3**.

[0085] Table 3: Reproducibility for ONS-3010 Sequence Method

	Heavy Chain			Light Chain		
	Try	Chy	Overall	Try	Chy	Overall
#1 Ref. Std.	100.0%	86.3%	100.0%	98.6%	89.3%	100.0%
#1 Sample	98.9%	87.8%	100.0%	98.6%	89.3%	100.0%
#2 Ref. Std.	99.3%	81.2%	100.0%	98.6%	89.3%	100.0%
#2 Sample	99.3%	86.3%	100.0%	98.6%	89.3%	100.0%
#3 Ref. Std.	99.3%	81.2%	100.0%	98.6%	89.3%	100.0%
#3 Sample	99.3%	81.2%	100.0%	98.6%	89.3%	100.0%

[0086] While the disclosure has been described above in conjunction with specific embodiments, alternatives, modifications, permutations, and variations will become apparent to a person of skill in the art in light of the foregoing description. Accordingly, it is intended that the present invention embraces all such alternatives, modifications, and variations as falling within the scope of the claims below.

[0087] What is claimed is:

CLAIMS

1. A method for determining the biosimilarity of a test protein in relation to a target biologic, the method comprising the steps of:

(a) digesting a first sample of a test protein for a first incubation time using a first protease and digesting a second sample of the test protein for a second incubation time using a second protease, wherein the first sample and the second sample are physically separated;

(b) applying column chromatography and tandem mass spectroscopy to the first sample under conditions sufficient to enhance binding of small peptides to the column, and generating a sequence of the test protein in the first sample;

(c) applying column chromatography and tandem mass spectroscopy to the second sample under conditions sufficient to enhance binding of small peptides to the column, and generating the sequence of the test protein in the second sample, wherein the first sample and second sample are physically separated;

(d) identifying the test protein as biosimilar to the target biologic when the test protein comprises 100% sequence identity to the target biologic; and

(e) identifying the test protein as not biosimilar to the target biologic when the test protein does not comprise 100% sequence identity to the target biologic.

2. The method of claim 1, wherein the monoclonal antibody comprises Adalimumab.

3. The method of any claim 1 or 2, wherein the first protease is Trypsin.

4. The method of any one of claims 1-3, wherein the second protease is Chymotrypsin.

5. The method of any one of claims 1-4, wherein the first digestion period is about 0.1 to about 1.0 hour.

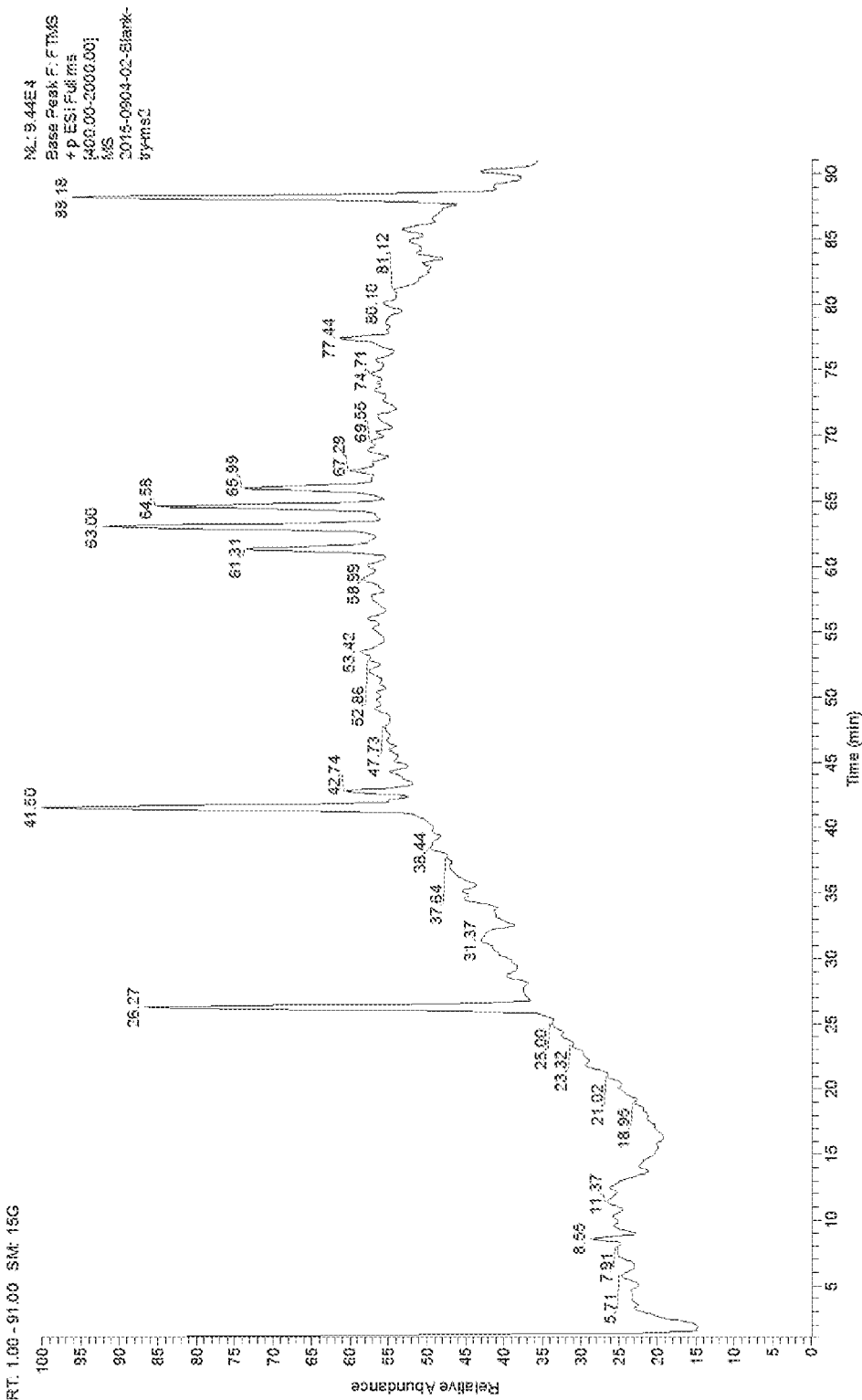
6. The method of claim 5, wherein the first digestion period is about 0.1 to about 0.5 hour.

7. The method of claim 5, wherein the first digestion period is about 0.6 to about 1.0 hour.

8. The method of claim 5, wherein the first digestion period is about 0.5 hours.
9. The method of any one of claims 1-8, wherein the second digestion period is about 0.1 to about 2.0 hours.
10. The method of claim 9, wherein the second digestion period is about 0.1 to about 1.5 hours.
11. The method of claim 9, wherein the second digestion period is about 1.5 to about 2.0 hours.
12. The method of claim 9, wherein the second digestion period is about 1.5 hours.

Trypsin Digested

Fig. 1a



Chymotrypsin Digested

Fig. 1b

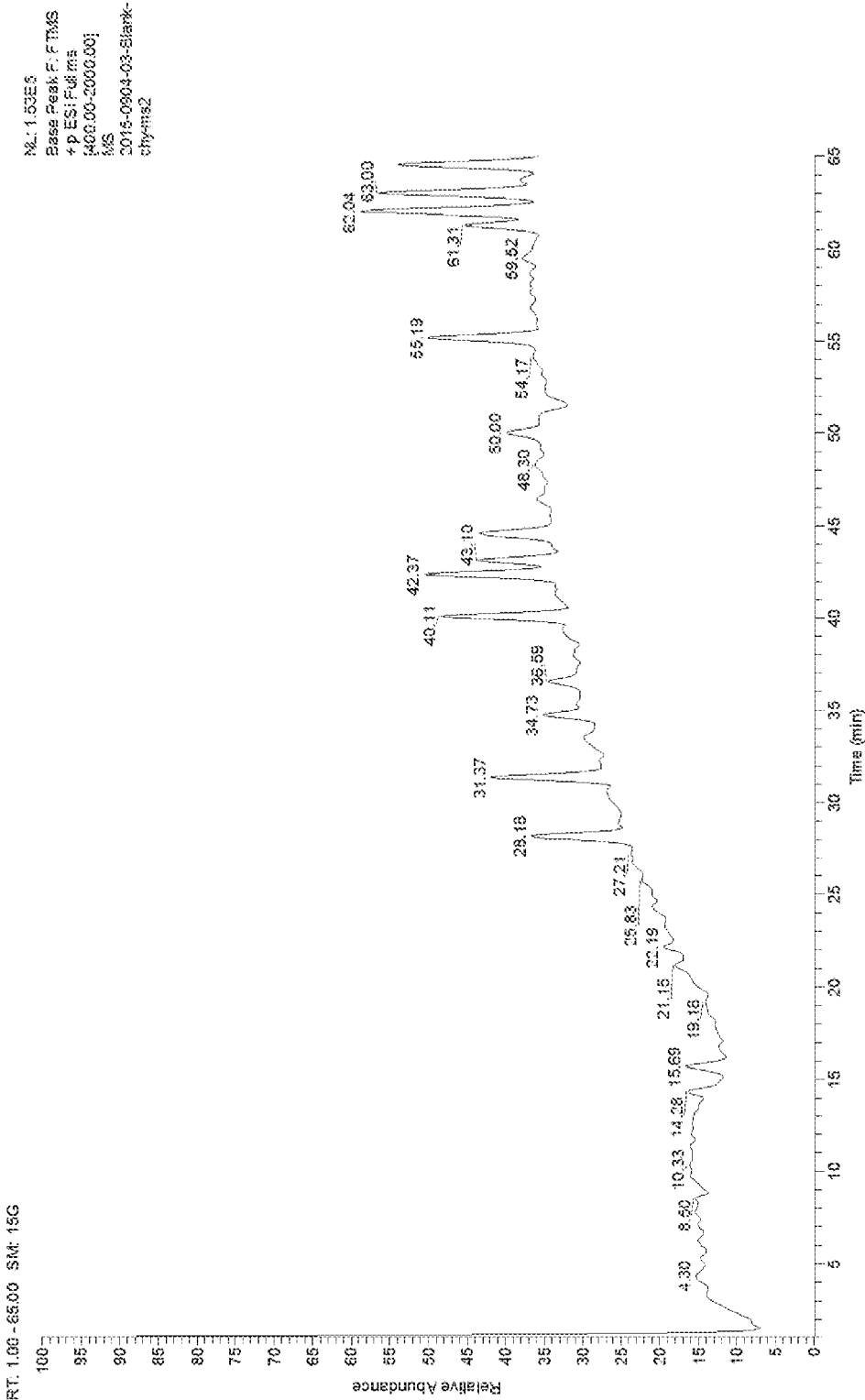


Fig. 2a

Coverage: 100.00%

Found Modifications:

- C Carboxymethyl (C)
- L Lysine-loss (C-term)
- N N-linked glycosylation (N)

Sequence		Modification List									
Modifications	Humira	1	11	21	31	41	51	61	71	81	91
				C							C
		EVLVEGGG LVQPCSLRL SCASGETED DYAMVTRQA PCKGLEWISA ITWNSCHIDY ADSVEGEFTI SDNNKNSLY LQNSLNEED TAVTTCKEYS									
Modifications	Humira	101					C				
		YLSTASLSHY WQGEPLTPVVS SASIKPEPVE PLAPSSKETS GCTRALGLN KYTEPFTTV SONSCHLTSC WHTEPRLQS SGLYSLSSTVV TWPSSSLGTQ									
Modifications	Humira	201	C	L	C	C		C			
		TYICNVNHP SNTKVKRVE PKSCVNTIC PFCPAPELLG GPSVLEPPK PKDTLMISRT PEVTCVVIVV SHEDEPEKEN WTVGCTEVAN AEIKEDREY									
Modifications	Humira	301	N			C				C	
		NSTYRWSTL TVLQWELNG KEYCKEYSK ALPADIETI SKAKQPRP QVYTLPPSD ELTKQVSLT CLNKCTPSP IANVEENQ PENNYKTFP									
Modifications	Humira	401			C		L				
		VLDSGSEFL YSKLTWIKSR WQGNVESCS VDEHRLNNY TQKSLSLSPG K									

High Possibility

Medium Possibility

Low Possibility

Fig. 2b

Coverage: 86.25%

Found Modifications:

- C Carboxymethyl (C)
- L Lysine-loss (C-term)

Sequence		Modification List									
Modifications	1	1	11	21	31	41	51	61	71	81	91
Humira				C							C
EVQLVESGGG LVQPGRSRLA SCRASTED DYAMHWQAQ PKGKLEWNA ITNSGCHDY ADSTEGEFTI SDMAKNSLY LQMSLERED TAVTYCAKYS											
Modifications	101						C				
Humira											
YLETSSELDY WQQTITPTTS SRTNPSSTP PNPSSKSTP GSTRMGLV KYVEDEPTV FNSGSLTSG HTTFDAVLQS SGLYSLSSTV TPTSSSLGTQ											
Modifications	201	C		C	C	C			C		
Humira											
TYICNNHNP SWKTVKAVE PSCNTNHC PFCPRRLIG GFSVLELPK PKDTLMISNI DEVTGVVNV SHEDPEKEN WYVNGVQVN AKTQDEEYV											
Modifications	301										
Humira											
NSTYRWVSU TLNQWELNG KEYKCVSNK ALPAIENI SKAKQPEEP QUTTLPPSD ELTENQVSLT CLVNEPTPD IAVVESNGQ PENNTATIPP											
Modifications	401				C						L
Humira											
VLDSSESEL YSKLVVKKR WQCAVTSLS VMEHLNNY TQKSLSLNG K											

High Possibility

Medium Possibility

Low Possibility

Fig. 2c

Coverage: 98.60%

Found Modifications:

C Carboxymethyl (C)

Sequence		Modification List
Modifications	1	
Humira		DIQMTQSPSS LRSRWEDRVT ITCRASQELR NYLADYQQKP GKPRLITYR ASTLQSCWPS KESGSGSESD PTLITSSLP EDVATYYIQR YNRAPYFEGQ
Modifications	101	
Humira		GTKVEIKETV ARPSTVELEPP SHEQLKGETR SWCLLNNTY PREARQWNV DNALQSGNSQ ESWTRDSKD STYLSLSLT LSKRDYENK WYACEWTRG
Modifications	201	
Humira		LSSPTKSEN KGEK

High Possibility

Medium Possibility

Low Possibility

Fig. 2d

Coverage: 89.25%

Found Modifications:

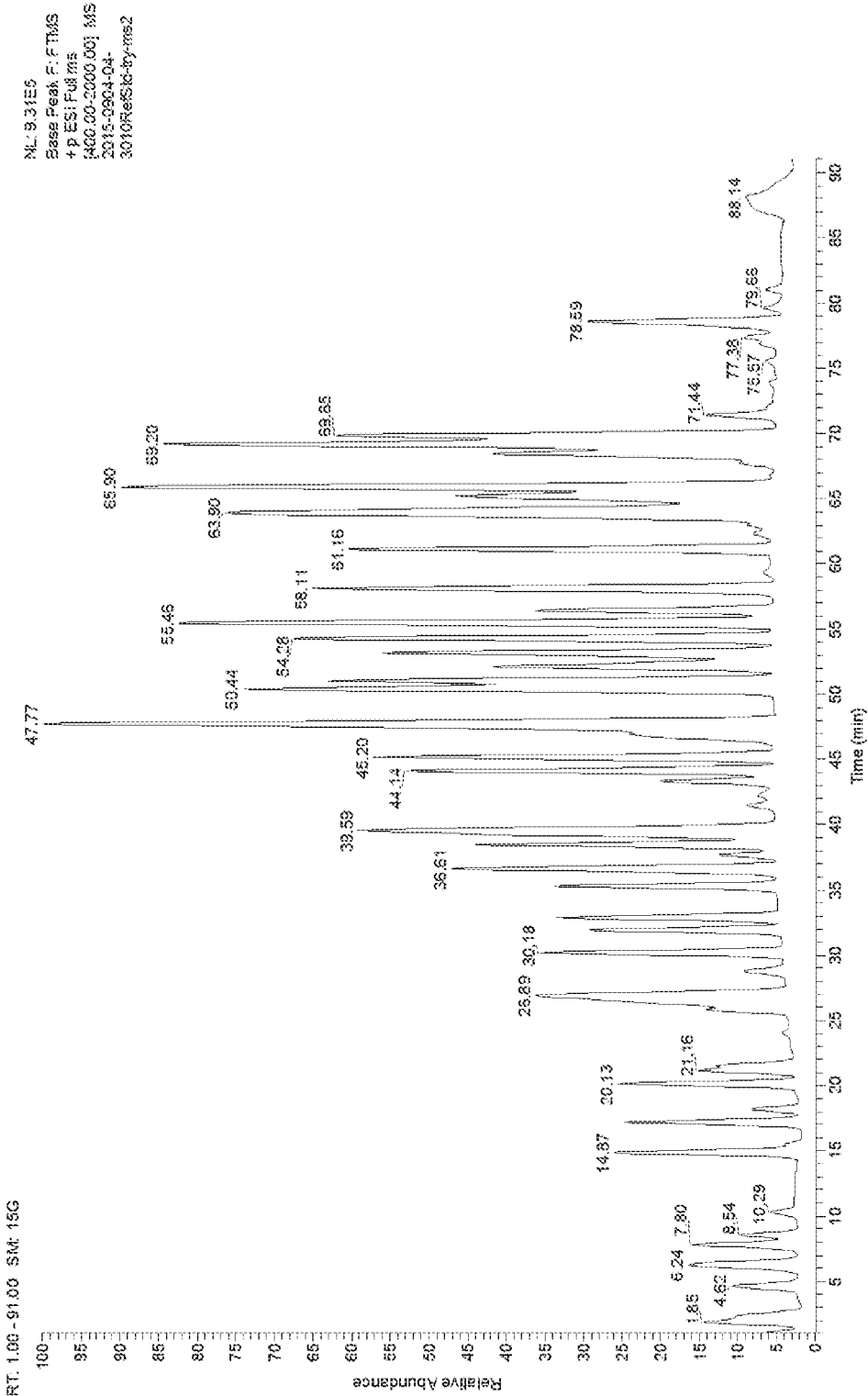
C Carboxymethyl (C)

Sequence	Modification List
Modifications Humira	1 NIDMTQ595S LSASVGDVDT ITCPASQGIR NYLNMVQKRP GKAKMLLIYA KSTHQSCHPS KESGSGSGFD EFTLSSSLQP KDVATYYCQR YNRPPTYEQ C
Modifications Humira	101 GTVKELKTV HAPSVETIEPP SHQIKSGTA SVULLNNEY PREMMNKKV DNMQSCHSQ ESTPEQDSKD STYSLSSSTLT LSKADYEMKK VYACEVTNQG C
Modifications Humira	201 LSSPVTKSEN NCRC C

High Possibility
Medium Possibility
Low Possibility

Trypsin Digested

Fig. 3a



Chymotrypsin Digested

Fig. 3b

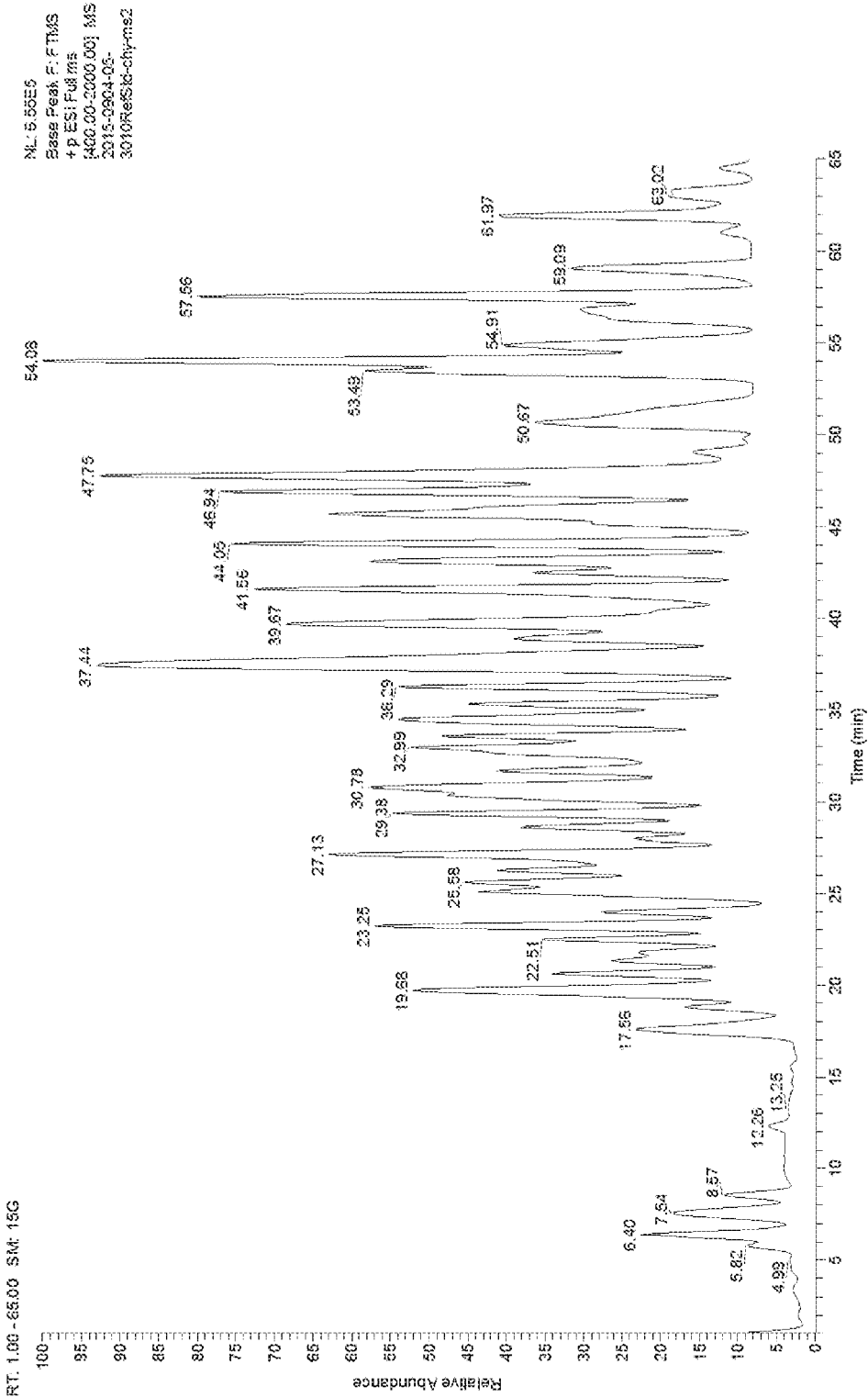


Fig. 4a

Coverage: 98.89%

Found Modifications:

- C Carboxymethyl (C)
- L Lysine-loss (C-term)
- N N-linked glycosylation (N)

Sequence		Modification List									
Modifications	1	11	21	31	41	51	61	71	81	91	
Humira	EVQLVESGGG LVQPGSRSLRL SCARSGTEED DYAGAWYDQA PEGCLEWWSA ITWNGCHIDY ADSVEGRET I SEDNKNKSLY LQNSCLRAED TATYTCARVS		C							C	
Modifications	101										
Humira	YLSTASSLDY WGGTGLTVS SASINRPSVE PLAPSSKSLT GGTALGGLV KYTTEPTTV SWNSGLRLSE WITHEPAVLAS SCGLYSLSSVY TYPSSSLGTD										
Modifications	201	C	L C	C C				C			
Humira	TYICNWHKP SNTKYDKKTE PASCDATHTC PFCPPRELLG GPEVELEPPK PAKTLMISKT PFTTCUTTVW SHEDPEVEN WTUDGVEHEN AKTKREEDY										
Modifications	301N		C						C		
Humira	NSTYAVVGVL TULHQLWLC KYTEKVSNK ALPADIENTI SKRKGQPER QYTLPPSRD ELTEKQVSLT CLVKEPYPSE IAVVESNSQ PEENTETTPD										
Modifications	401			C							
Humira	VLDSSGSEEL YSKLTPDKSA QWQGVESCS VMEHLLNHY TQKSLSLSPG K										

High Possibility

Medium Possibility

Low Possibility

Fig. 4b

Coverage: 85.14%

Found Modifications:

- C Carboxymethyl (C)
- L Lysine-loss (C-term)

Sequence		Modification List									
Modifications	1	11	21	31	41	51	61	71	81	91	
Humira			C							C	
Modifications	101										
Humira											
Modifications	201	C		C	C	C					
Humira											
Modifications	301										
Humira											
Modifications	401				C						
Humira											

- High Possibility
- Medium Possibility
- Low Possibility

Fig. 4c

Coverage: 98.60%

Found Modifications:

C Carboxymethyl (C)

Sequence		Modification List									
Modifications	1	1	11	21	31	41	51	61	71	81	91
Humira		DIGNIQSPSS LRSUGCHRTT ITERRSQGIA NYLNNYQKP GKRPKLIYA ASTLQSEVPS RESGSGCTD YLTLSISIQP EDVATTTQQR YNRAPFTTCQ									
		C									
Modifications	101										
Humira		GTRVELKRTV RAPSVEIEPP SDEQLKSGTA SWCLLNNEY PRALKWQKV DNLQSENSQ ESTEQDSKD STYLSLSLT LSKNPTTEK YTRAEVTEQ									
		C									
Modifications	201										
Humira		LSPPTKSN KEEC									

High Possibility

Medium Possibility

Low Possibility

Fig. 4d

Coverage: 89.25%

Found Modifications:

C Carboxymethyl (C)

Sequence		Modification List	
Modifications	1		
Humira		1	11 21 31 41 51 61 71 81 91
			MIQMTQSPSS LSAEVDNVT ITCASQCIR NTLNWTQKP CARPKLLIYA RETLQSGTFS RESGSESGTB ELTITSELOP EDWNTYYQOR YNKRPTTHCQ
Modifications	101		
Humira			101
			GYKVEIKRTV KAPSVETFPF SBEQLKSTPA SVVCLLNNTY PREKNQWNV DNLQSGNSQ ESWTRQDSKB STYSLSSTLT LSKRNTYEMK VTACEVTHQC
Modifications	201		
Humira			201
			LSSPTKSEN EEEC

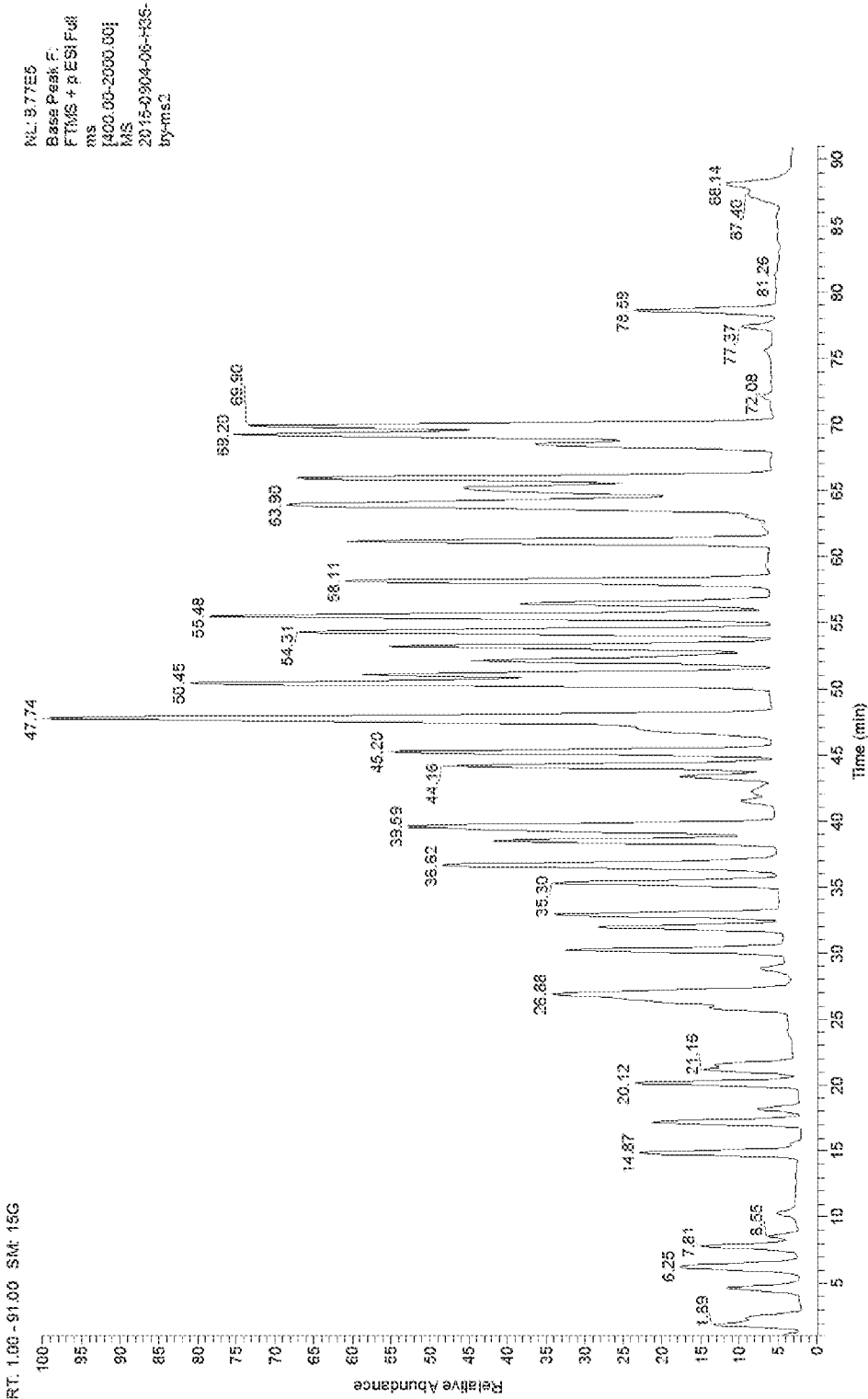
High Possibility

Medium Possibility

Low Possibility

Trypsin Digested

Fig. 5a



Chymotrypsin Digested

Fig. 5b

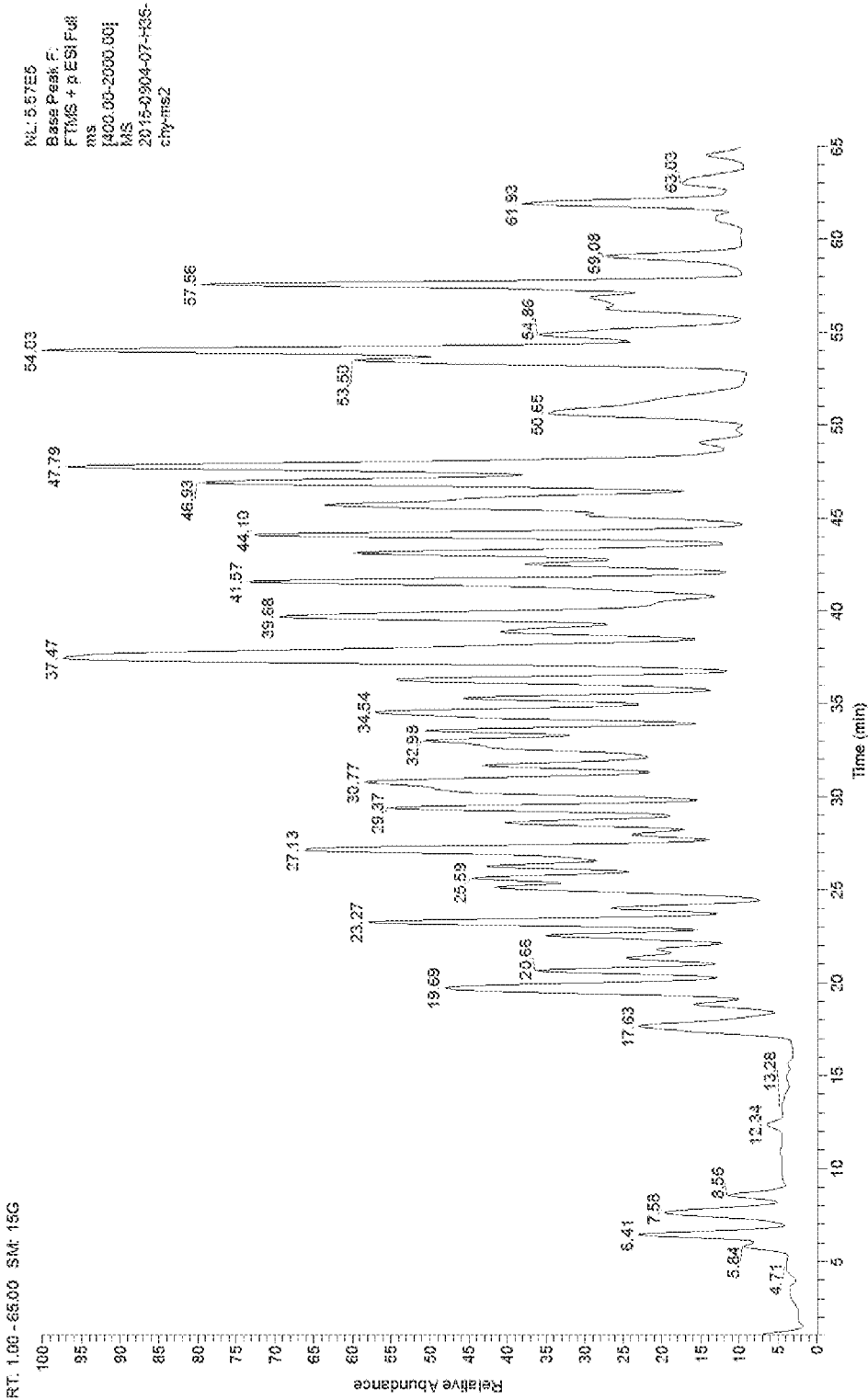


Fig. 6b

Coverage: 50.78%

Found Modifications:

- C Carboxymethyl (C)
- L Lysine-loss (C-term)

Sequence		Modification List									
Modifications	Humira	1	11	21	31	41	51	61	71	81	91
		C									
Modifications	Humira	101	L								
		C									
Modifications	Humira	201	C								
		L									
Modifications	Humira	301	L								
		C									
Modifications	Humira	401	L								
		C									

High Possibility
Medium Possibility
Low Possibility

Fig. 6c

Coverage: 48.60%

Found Modifications:

C Carboxymethyl (C)

Sequence	Modification List
Modifications	1
Humira	DIQNTQSPSS LSASVEDRVT ITCRASQGITR NYLIANTQKRP GRAPKLLIYA ASTLQSGVPS EPSSGSGCID FTLTSSSLQP EDVATYTCQR YNRAPYTFGQ
Modifications	101
Humira	CTKVEIKRTV MAPGVETIEPP SDEQAKSGTA SYVCLINNEY PEAKVQWAV ENRLQSENSQ ESVTEQDSKD STYSLSSTLF LSKRNYENK YINCEVTNGC
Modifications	201
Humira	LSSPTKSEN DEEC

High Possibility

Medium Possibility

Low Possibility

Fig. 6d

Coverage: 45.79%

Found Modifications:

C Carboxymethyl (C)

Sequence	Modification List
Modifications	1
Humira	DICHTQSPSS LSSSVDEPVT ITCPASQELR NYLAWYQKRP GKADKLLIYA ASTLOSQVPS DESGSGSETD PTLTSSSQP EDVATYYCQR YNRAPYTFGQ
Modifications	101
Humira	GTIVETIKRTV AAPSVFIEPP SUBQAKSETA GWCLLNNEY PREAKYQMKV ENALQSENSQ ESWTEQDSKD STYSLSSTLT LSKNDYTEMK VYACEVTFMG C
Modifications	201
Humira	LSPFYTKSEN RGECC C

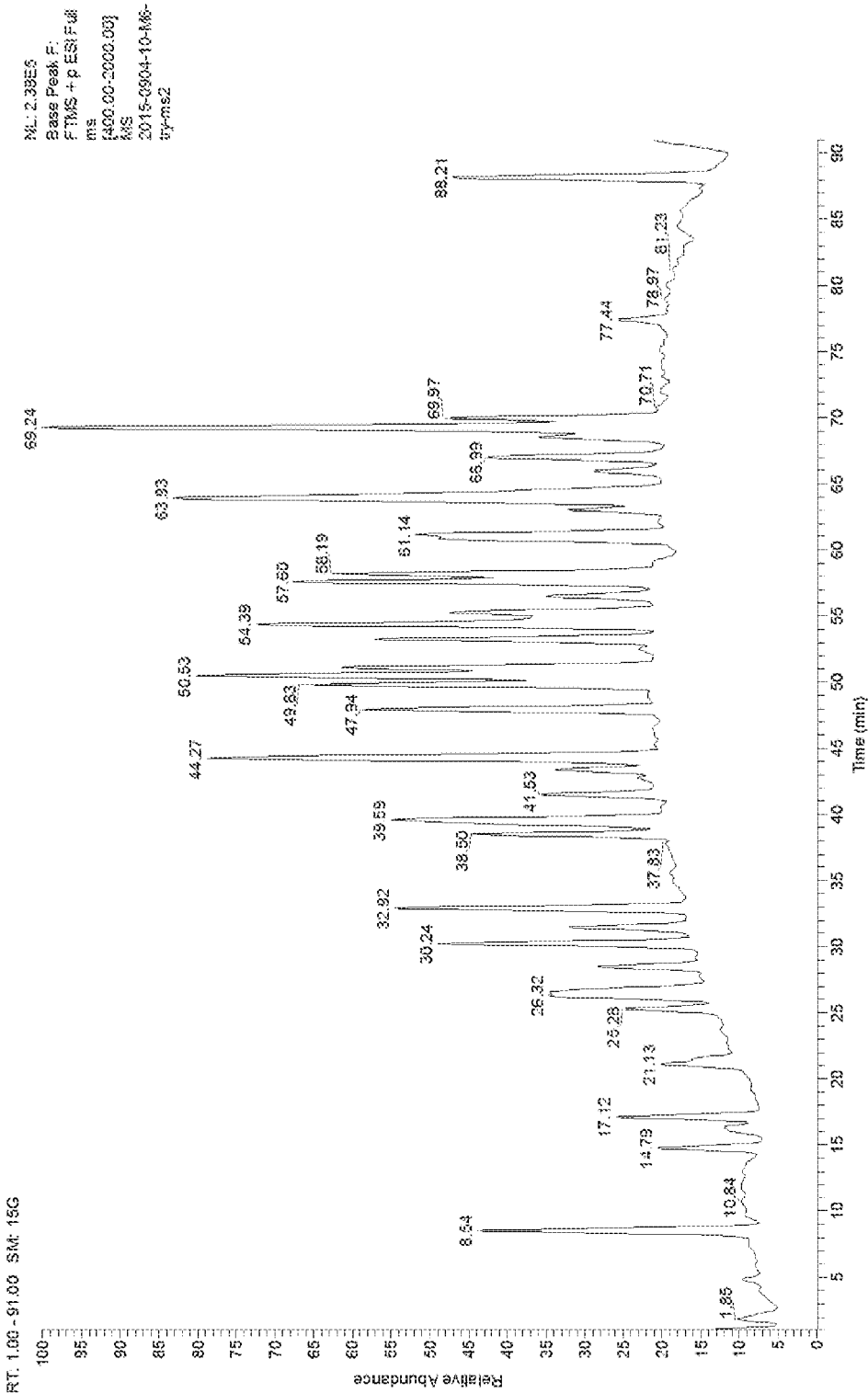
High Possibility

Medium Possibility

Low Possibility

Trypsin Digested

Fig. 7a



Chymotrypsin Digested

Fig. 7b

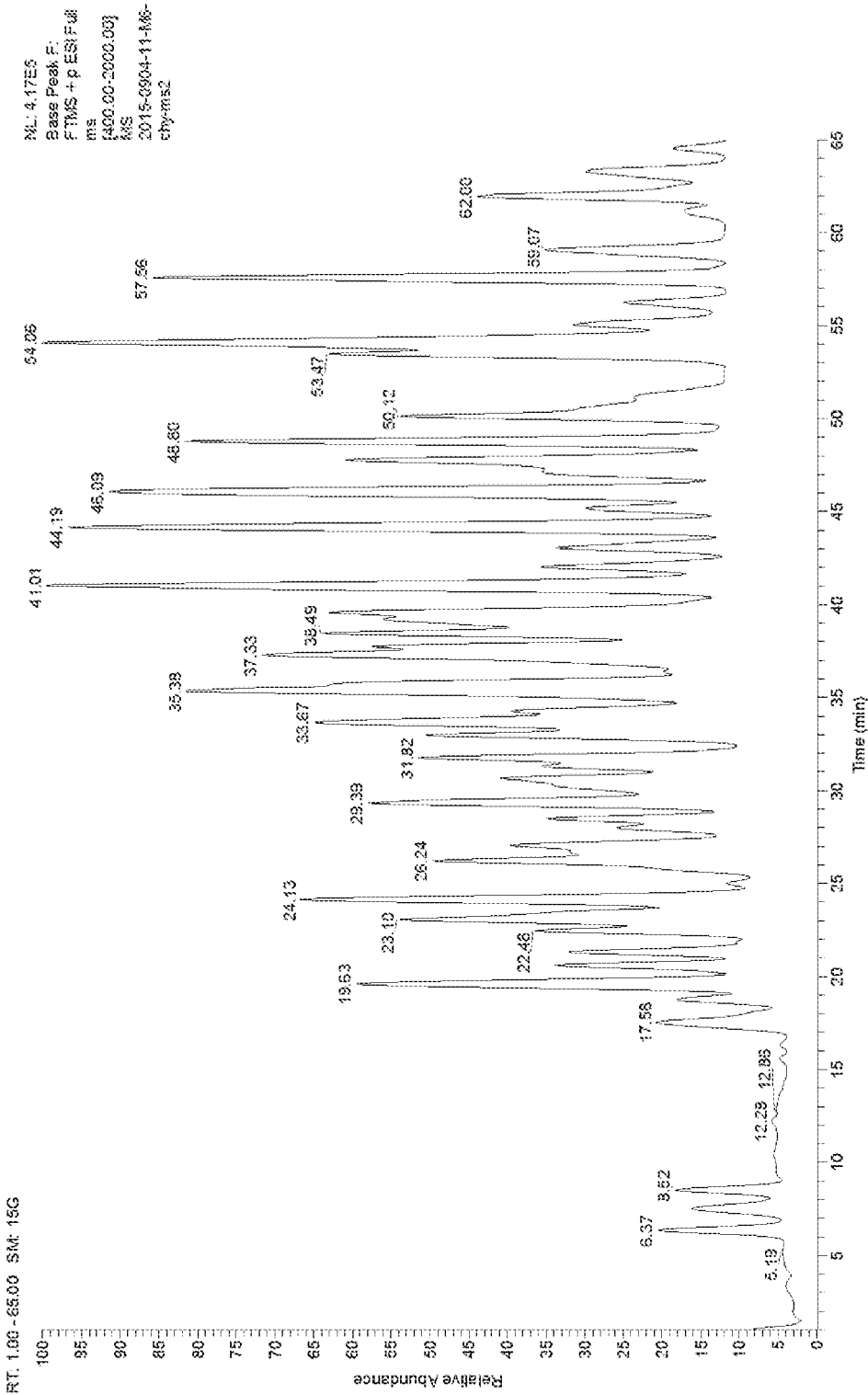


Figure 8 illustrates the theoretical amino acid sequences of the light chain and heavy chain of Humira

Light Chain:

10	20	30	40	50	60
DIQMTQSPSS	LSASVGRVT	ITCRASQGI	NYLAWYQQK	GKAPKLLIYA	ASTLQSGVPS
70	80	90	100	110	120
RFSGSGSGTD	FTLTISSLQP	EDVATYYCQR	YNRAFYTFGQ	GTKVEIKRTV	AAPSVEIFTFP
130	140	150	160	170	180
SDEQLKSGTA	SVVCLLNFF	PREAKVQMKV	DNALQSENQ	ESVTEQDSKD	STYSLSSTLT
190	200	210			
LSKADYEKHK	YYACEVTHQG	ISSPVTKSFN	RGEC		

Heavy Chain:

10	20	30	40	50	60
EVQLVESGGG	LVQPGKSLRL	SCAASGFTED	DYANHWVRQA	PGKGLEWVSA	ITWNSGHIDY
70	80	90	100	110	120
ADSEVEGRFTI	SRDNAKNSLY	LQMNSLRAED	TAVYYCAKVS	YLSTASSLDY	WGQGTLLVTVS
130	140	150	160	170	180
SASTRGPSPVF	FLAPSSKSTS	GGTAALGCI	KDYFPEPVTV	SWNSGALTS	VHTFPAVLQS
190	200	210	220	230	240
SGLYSLSSVV	TVPSSSLGTFQ	TYICNVNHKE	SNTRKVDKKVE	PKSCDKHTHC	PPCPAPELLG
250	260	270	280	290	300
GPEVELFPFK	PKDTLMISRT	PEYTCVVVDV	SHEDPEVKFN	WYVDGVEVHN	AKTKPREEQY
310	320	330	340	350	360
NSTYRVVSVL	TVLHQDWLNG	KEYAKCKVSNK	ALPAPIEKTI	SKAKGQPREP	QVYTIIPPSRD
370	380	390	400	410	420
ELITKNQVSLT	CLVKGFTPSD	IAVEWESNGQ	PENNYKTTTP	VLDSDGSFFL	YSKLTIVDKSR
430	440	450			
WCCGNVFCS	VMHEALHNHY	TOKSLSLSPCK			

ONBI-008001WO_SeqList
SEQUENCE LISTING

<110> Oncobiologics, Inc.
 <120> METHODS FOR IDENTIFYING AND ANALYZING AMINO ACID SEQUENCES OF PROTEINS
 <130> ONBI-008/001WO 325849-2029
 <150> 62/291,216
 <151> 2016-02-04
 <160> 26
 <170> PatentIn version 3.5
 <210> 1
 <211> 451
 <212> PRT
 <213> Artificial Sequence
 <220>
 <223> Synthetic Construct
 <400> 1

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
 1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
 20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
 35 40 45

Ser Ala Ile Thr Trp Asn Ser Gly His Ile Asp Tyr Ala Asp Ser Val
 50 55 60

Glu Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
 65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
 85 90 95

Ala Lys Val Ser Tyr Leu Ser Thr Ala Ser Ser Leu Asp Tyr Trp Gly
 100 105 110

Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys Gly Pro Ser
 115 120 125

Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly Gly Thr Ala
 130 135 140

Ala Leu Gly Cys Leu Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val
 145 150 155 160

Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr Phe Pro Ala
 165 170 175

ONBI-008001W0_SeqList

Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val Val Thr Val
180 185 190

Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Ile Cys Asn Val Asn His
195 200 205

Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Glu Pro Lys Ser Cys
210 215 220

Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly
225 230 235 240

Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met
245 250 255

Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His
260 265 270

Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val
275 280 285

His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr
290 295 300

Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly
305 310 315 320

Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile
325 330 335

Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val
340 345 350

Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser
355 360 365

Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu
370 375 380

Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro
385 390 395 400

Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val
405 410 415

Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met
420 425 430

His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser
435 440 445

Pro Gly Lys
450

<210> 2
<211> 451
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> MOD_RES
<222> (22)..(22)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (96)..(96)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (148)..(148)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (204)..(204)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (222)..(222)
<223> Xaa = lysine-loss

<220>
<221> MOD_RES
<222> (224)..(224)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (230)..(230)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (233)..(233)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (265)..(265)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (301)..(301)
<223> Xaa = N-linked glycosylation

<220>
<221> MOD_RES
<222> (325)..(325)

ONBI-008001W0_SeqList

<223> Xaa = carboxymethyl

<220>

<221> MOD_RES

<222> (371)..(371)

<223> Xaa = carboxymethyl

<220>

<221> MOD_RES

<222> (429)..(429)

<223> Xaa = carboxymethyl

<220>

<221> MOD_RES

<222> (451)..(451)

<223> Xaa = lysine-loss

<400> 2

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Xaa Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Ala Ile Thr Trp Asn Ser Gly His Ile Asp Tyr Ala Asp Ser Val
50 55 60

Glu Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Xaa
85 90 95

Ala Lys Val Ser Tyr Leu Ser Thr Ala Ser Ser Leu Asp Tyr Trp Gly
100 105 110

Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys Gly Pro Ser
115 120 125

Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly Gly Thr Ala
130 135 140

Ala Leu Gly Xaa Leu Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val
145 150 155 160

Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr Phe Pro Ala
165 170 175

Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val Val Thr Val
180 185 190

Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Ile Xaa Asn Val Asn His
Page 4

ONBI-008001W0_SeqList
200 205

195

Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Glu Pro Xaa Ser Xaa
210 215 220

Asp Lys Thr His Thr Xaa Pro Pro Xaa Pro Ala Pro Glu Leu Leu Gly
225 230 235 240

Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met
245 250 255

Ile Ser Arg Thr Pro Glu Val Thr Xaa Val Val Val Asp Val Ser His
260 265 270

Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val
275 280 285

His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Xaa Ser Thr Tyr
290 295 300

Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly
305 310 315 320

Lys Glu Tyr Lys Xaa Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile
325 330 335

Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val
340 345 350

Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser
355 360 365

Leu Thr Xaa Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu
370 375 380

Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro
385 390 395 400

Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val
405 410 415

Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Xaa Ser Val Met
420 425 430

His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser
435 440 445

Pro Gly Xaa
450

<210> 3

ONBI-008001W0_SeqList

<211> 451
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic Construct

<400> 3

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
 1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
 20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
 35 40 45

Ser Ala Ile Thr Trp Asn Ser Gly His Ile Asp Tyr Ala Asp Ser Val
 50 55 60

Glu Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
 65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
 85 90 95

Ala Lys Val Ser Tyr Leu Ser Thr Ala Ser Ser Leu Asp Tyr Trp Gly
 100 105 110

Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys Gly Pro Ser
 115 120 125

Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly Gly Thr Ala
 130 135 140

Ala Leu Gly Cys Leu Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val
 145 150 155 160

Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr Phe Pro Ala
 165 170 175

Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val Val Thr Val
 180 185 190

Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Ile Cys Asn Val Asn His
 195 200 205

Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Glu Pro Lys Ser Cys
 210 215 220

Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly
 225 230 235 240

ONBI-008001W0_SeqList

Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met
245 250 255

Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His
260 265 270

Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val
275 280 285

His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr
290 295 300

Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly
305 310 315 320

Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile
325 330 335

Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val
340 345 350

Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser
355 360 365

Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu
370 375 380

Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro
385 390 395 400

Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val
405 410 415

Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met
420 425 430

His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser
435 440 445

Pro Gly Lys
450

<210> 4
<211> 451
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> MOD_RES

ONBI-008001W0_SeqList

<222> (22)..(22)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (96)..(96)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (148)..(148)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (174)..(174)
<223> Xaa = lysine-loss

<220>
<221> MOD_RES
<222> (204)..(204)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (224)..(224)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (230)..(230)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (233)..(233)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (265)..(265)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (429)..(429)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (451)..(451)
<223> Xaa = lysine-loss

<400> 4

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Xaa Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Ala Ile Thr Trp Asn Ser Gly His Ile Asp Tyr Ala Asp Ser Val

ONBI-008001W0_SeqList

50

55

60

Glu Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Xaa
85 90 95

Ala Lys Val Ser Tyr Leu Ser Thr Ala Ser Ser Leu Asp Tyr Trp Gly
100 105 110

Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys Gly Pro Ser
115 120 125

Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly Gly Thr Ala
130 135 140

Ala Leu Gly Xaa Leu Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val
145 150 155 160

Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr Xaa Pro Ala
165 170 175

Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val Val Thr Val
180 185 190

Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Ile Xaa Asn Val Asn His
195 200 205

Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Glu Pro Lys Ser Xaa
210 215 220

Asp Lys Thr His Thr Xaa Pro Pro Xaa Pro Ala Pro Glu Leu Leu Gly
225 230 235 240

Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met
245 250 255

Ile Ser Arg Thr Pro Glu Val Thr Xaa Val Val Val Asp Val Ser His
260 265 270

Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val
275 280 285

His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr
290 295 300

Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly
305 310 315 320

Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile

ONBI-008001W0_SeqList

325

330

335

Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val
340 345 350

Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser
355 360 365

Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu
370 375 380

Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro
385 390 395 400

Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val
405 410 415

Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Xaa Ser Val Met
420 425 430

His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser
435 440 445

Pro Gly Xaa
450

<210> 5
<211> 214
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 5

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Arg Asn Tyr
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Ala Ala Ser Thr Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Val Ala Thr Tyr Tyr Cys Gln Arg Tyr Asn Arg Ala Pro Tyr
85 90 95

ONBI-008001W0_SeqList

Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Thr Val Ala Ala
100 105 110

Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
115 120 125

Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
180 185 190

Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
195 200 205

Phe Asn Arg Gly Glu Cys
210

<210> 6
<211> 214
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> MOD_RES
<222> (23)..(23)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (88)..(88)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (134)..(134)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (194)..(194)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (214)..(214)
<223> Xaa = carboxymethyl

<400> 6

ONBI-008001W0_SeqList

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Xaa Arg Ala Ser Gln Gly Ile Arg Asn Tyr
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Ala Ala Ser Thr Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Val Ala Thr Tyr Tyr Xaa Gln Arg Tyr Asn Arg Ala Pro Tyr
85 90 95

Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Thr Val Ala Ala
100 105 110

Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
115 120 125

Thr Ala Ser Val Val Xaa Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
180 185 190

Ala Xaa Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
195 200 205

Phe Asn Arg Gly Glu Xaa
210

<210> 7
<211> 214
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 7

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
Page 12

ONBI-008001W0_SeqList

1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Arg Asn Tyr
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Ala Ala Ser Thr Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Val Ala Thr Tyr Tyr Cys Gln Arg Tyr Asn Arg Ala Pro Tyr
85 90 95

Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Thr Val Ala Ala
100 105 110

Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
115 120 125

Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
180 185 190

Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
195 200 205

Phe Asn Arg Gly Glu Cys
210

<210> 8
<211> 214
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> MOD_RES
<222> (88)..(88)
<223> Xaa = carboxymethyl

ONBI-008001W0_SeqList

<220>
 <221> MOD_RES
 <222> (134)..(134)
 <223> Xaa = carboxymethyl

<220>
 <221> MOD_RES
 <222> (194)..(194)
 <223> Xaa = carboxymethyl

<220>
 <221> MOD_RES
 <222> (214)..(214)
 <223> Xaa = carboxymethyl

<400> 8

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
 1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Arg Asn Tyr
 20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
 35 40 45

Tyr Ala Ala Ser Thr Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
 50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
 65 70 75 80

Glu Asp Val Ala Thr Tyr Tyr Xaa Gln Arg Tyr Asn Arg Ala Pro Tyr
 85 90 95

Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Thr Val Ala Ala
 100 105 110

Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
 115 120 125

Thr Ala Ser Val Val Xaa Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
 130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
 145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
 165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
 180 185 190

Ala Xaa Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
 195 200 205

ONBI-008001W0_SeqList

Phe Asn Arg Gly Glu Xaa
210

<210> 9
<211> 451
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 9

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Ala Ile Thr Trp Asn Ser Gly His Ile Asp Tyr Ala Asp Ser Val
50 55 60

Glu Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

Ala Lys Val Ser Tyr Leu Ser Thr Ala Ser Ser Leu Asp Tyr Trp Gly
100 105 110

Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys Gly Pro Ser
115 120 125

Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly Gly Thr Ala
130 135 140

Ala Leu Gly Cys Leu Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val
145 150 155 160

Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr Phe Pro Ala
165 170 175

Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val Val Thr Val
180 185 190

Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Ile Cys Asn Val Asn His
195 200 205

ONBI-008001W0_SeqList

Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Glu Pro Lys Ser Cys
 210 215 220

Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly
 225 230 235

Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met
 245 250 255

Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His
 260 265 270

Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val
 275 280 285

His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr
 290 295 300

Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly
 305 310 315 320

Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile
 325 330 335

Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val
 340 345 350

Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser
 355 360 365

Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu
 370 375 380

Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro
 385 390 395 400

Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val
 405 410 415

Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met
 420 425 430

His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser
 435 440 445

Pro Gly Lys
 450

<210> 10
 <211> 451
 <212> PRT
 <213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> MOD_RES
<222> (22)..(22)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (96)..(96)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (148)..(148)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (204)..(204)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (222)..(222)
<223> Xaa = lysine-loss

<220>
<221> MOD_RES
<222> (224)..(224)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (230)..(230)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (233)..(233)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (265)..(265)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (301)..(301)
<223> Xaa = N-linked glycosylation

<220>
<221> MOD_RES
<222> (325)..(325)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (371)..(371)
<223> Xaa = carboxymethyl

<220>
<221> misc_feature
<222> (429)..(429)

ONBI-008001WO_SeqList

<223> Xaa can be any naturally occurring amino acid

<220>

<221> MOD_RES

<222> (439)..(439)

<223> Xaa = carboxymethyl

<220>

<221> MOD_RES

<222> (451)..(451)

<223> Xaa = lysine-loss

<400> 10

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Xaa Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Ala Ile Thr Trp Asn Ser Gly His Ile Asp Tyr Ala Asp Ser Val
50 55 60

Glu Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Xaa
85 90 95

Ala Lys Val Ser Tyr Leu Ser Thr Ala Ser Ser Leu Asp Tyr Trp Gly
100 105 110

Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys Gly Pro Ser
115 120 125

Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly Gly Thr Ala
130 135 140

Ala Leu Gly Xaa Leu Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val
145 150 155 160

Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr Phe Pro Ala
165 170 175

Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val Val Thr Val
180 185 190

Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Ile Xaa Asn Val Asn His
195 200 205

Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Glu Pro Xaa Ser Xaa
210 215 220

ONBI-008001W0_SeqList

Asp Lys Thr His Thr Xaa Pro Pro Xaa Pro Ala Pro Glu Leu Leu Gly
 225 230 235 240
 Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met
 245 250 255
 Ile Ser Arg Thr Pro Glu Val Thr Xaa Val Val Val Asp Val Ser His
 260 265 270
 Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val
 275 280 285
 His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Xaa Ser Thr Tyr
 290 295 300
 Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly
 305 310 315 320
 Lys Glu Tyr Lys Xaa Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile
 325 330 335
 Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val
 340 345 350
 Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser
 355 360 365
 Leu Thr Xaa Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu
 370 375 380
 Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro
 385 390 395 400
 Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val
 405 410 415
 Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Xaa Ser Val Met
 420 425 430
 His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser
 435 440 445
 Pro Gly Xaa
 450

<210> 11
 <211> 451
 <212> PRT
 <213> Artificial Sequence

<220>

<223> Synthetic Construct

<400> 11

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
 1 5 10 15
 Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
 20 25 30
 Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
 35 40 45
 Ser Ala Ile Thr Trp Asn Ser Gly His Ile Asp Tyr Ala Asp Ser Val
 50 55 60
 Glu Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
 65 70 75 80
 Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
 85 90 95
 Ala Lys Val Ser Tyr Leu Ser Thr Ala Ser Ser Leu Asp Tyr Trp Gly
 100 105 110
 Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys Gly Pro Ser
 115 120 125
 Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly Gly Thr Ala
 130 135 140
 Ala Leu Gly Cys Leu Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val
 145 150 155 160
 Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr Phe Pro Ala
 165 170 175
 Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val Val Thr Val
 180 185 190
 Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Ile Cys Asn Val Asn His
 195 200 205
 Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Glu Pro Lys Ser Cys
 210 215 220
 Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly
 225 230 235 240
 Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met
 245 250 255

ONBI-008001W0_SeqList

Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His
260 265 270

Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val
275 280 285

His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr
290 295 300

Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly
305 310 315 320

Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile
325 330 335

Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val
340 345 350

Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser
355 360 365

Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu
370 375 380

Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro
385 390 395 400

Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val
405 410 415

Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met
420 425 430

His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser
435 440 445

Pro Gly Lys
450

<210> 12
<211> 451
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> MOD_RES
<222> (22)..(22)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES

ONBI-008001W0_SeqList

<222> (96)..(96)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (148)..(148)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (174)..(174)
<223> Xaa = lysine-loss

<220>
<221> MOD_RES
<222> (204)..(204)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (224)..(224)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (230)..(230)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (233)..(233)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (265)..(265)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (429)..(429)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (451)..(451)
<223> Xaa = lysine-loss

<400> 12

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Xaa Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Ala Ile Thr Trp Asn Ser Gly His Ile Asp Tyr Ala Asp Ser Val
50 55 60

Glu Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

ONBI-008001W0_SeqList

Leu Gln Met Asn Ser₈₅ Leu Arg Ala Glu Asp₉₀ Thr Ala Val Tyr Tyr₉₅ Xaa
 Ala Lys Val Ser₁₀₀ Tyr Leu Ser Thr Ala₁₀₅ Ser Ser Leu Asp Tyr₁₁₀ Trp Gly
 Gln Gly Thr₁₁₅ Leu Val Thr Val Ser₁₂₀ Ser Ala Ser Thr Lys₁₂₅ Gly Pro Ser
 Val Phe₁₃₀ Pro Leu Ala Pro Ser₁₃₅ Ser Lys Ser Thr Ser₁₄₀ Gly Gly Thr Ala
 Ala₁₄₅ Leu Gly Xaa Leu Val₁₅₀ Lys Asp Tyr Phe Pro₁₅₅ Glu Pro Val Thr Val₁₆₀
 Ser Trp Asn Ser Gly₁₆₅ Ala Leu Thr Ser Gly₁₇₀ Val His Thr Xaa Pro₁₇₅ Ala
 Val Leu Gln Ser₁₈₀ Ser Gly Leu Tyr Ser₁₈₅ Leu Ser Ser Val Val₁₉₀ Thr Val
 Pro Ser Ser₁₉₅ Ser Leu Gly Thr Gln₂₀₀ Thr Tyr Ile Xaa Asn₂₀₅ Val Asn His
 Lys Pro₂₁₀ Ser Asn Thr Lys Val₂₁₅ Asp Lys Lys Val Glu₂₂₀ Pro Lys Ser Xaa
 Asp₂₂₅ Lys Thr His Thr Xaa₂₃₀ Pro Pro Xaa Pro Ala₂₃₅ Pro Glu Leu Leu Gly₂₄₀
 Gly Pro Ser Val Phe₂₄₅ Leu Phe Pro Pro Lys₂₅₀ Pro Lys Asp Thr Leu₂₅₅ Met
 Ile Ser Arg Thr₂₆₀ Pro Glu Val Thr Xaa₂₆₅ Val Val Val Asp Val₂₇₀ Ser His
 Glu Asp Pro₂₇₅ Glu Val Lys Phe Asn₂₈₀ Trp Tyr Val Asp Gly₂₈₅ Val Glu Val
 His Asn Ala Lys Thr Lys Pro₂₉₅ Arg Glu Glu Gln Tyr₃₀₀ Asn Ser Thr Tyr
 Arg Val Val Ser Val Leu₃₁₀ Thr Val Leu His Gln₃₁₅ Asp Trp Leu Asn Gly₃₂₀
 Lys Glu Tyr Lys Cys₃₂₅ Lys Val Ser Asn Lys₃₃₀ Ala Leu Pro Ala Pro₃₃₅ Ile
 Glu Lys Thr Ile₃₄₀ Ser Lys Ala Lys Gly₃₄₅ Gln Pro Arg Glu Pro₃₅₀ Gln Val

ONBI-008001W0_SeqList

Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser
355 360 365

Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu
370 375 380

Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro
385 390 395 400

Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val
405 410 415

Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Xaa Ser Val Met
420 425 430

His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser
435 440 445

Pro Gly Xaa
450

<210> 13
<211> 214
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 13

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Arg Asn Tyr
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Ala Ala Ser Thr Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Val Ala Thr Tyr Tyr Cys Gln Arg Tyr Asn Arg Ala Pro Tyr
85 90 95

Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Thr Val Ala Ala
100 105 110

ONBI-008001WO_SeqList

Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
115 120 125

Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
180 185 190

Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
195 200 205

Phe Asn Arg Gly Glu Cys
210

<210> 14
<211> 214
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> MOD_RES
<222> (23)..(23)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (88)..(88)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (134)..(134)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (194)..(194)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (214)..(214)
<223> Xaa = carboxymethyl

<400> 14

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1 5 10 15

ONBI-008001WO_SeqList

Asp Arg Val Thr Ile Thr Xaa Arg Ala Ser Gln Gly Ile Arg Asn Tyr
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Ala Ala Ser Thr Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Val Ala Thr Tyr Tyr Xaa Gln Arg Tyr Asn Arg Ala Pro Tyr
85 90 95

Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Thr Val Ala Ala
100 105 110

Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
115 120 125

Thr Ala Ser Val Val Xaa Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
180 185 190

Ala Xaa Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
195 200 205

Phe Asn Arg Gly Glu Xaa
210

<210> 15
<211> 214
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 15

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Arg Asn Tyr
20 25 30

ONBI-008001W0_SeqList

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
 35 40 45
 Tyr Ala Ala Ser Thr Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
 50 55 60
 Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
 65 70 75 80
 Glu Asp Val Ala Thr Tyr Tyr Cys Gln Arg Tyr Asn Arg Ala Pro Tyr
 85 90 95
 Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Thr Val Ala Ala
 100 105 110
 Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
 115 120 125
 Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
 130 135 140
 Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
 145 150 155 160
 Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
 165 170 175
 Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
 180 185 190
 Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
 195 200 205
 Phe Asn Arg Gly Glu Cys
 210

<210> 16
 <211> 214
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic Construct

<220>
 <221> MOD_RES
 <222> (88)..(88)
 <223> Xaa = carboxymethyl

<220>
 <221> MOD_RES
 <222> (134)..(134)
 <223> Xaa = carboxymethyl

ONBI-008001W0_SeqList

<220>
 <221> MOD_RES
 <222> (194)..(194)
 <223> Xaa = carboxymethyl

<220>
 <221> MOD_RES
 <222> (214)..(214)
 <223> Xaa = carboxymethyl

<400> 16

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
 1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Arg Asn Tyr
 20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
 35 40 45

Tyr Ala Ala Ser Thr Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
 50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
 65 70 75 80

Glu Asp Val Ala Thr Tyr Tyr Xaa Gln Arg Tyr Asn Arg Ala Pro Tyr
 85 90 95

Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Thr Val Ala Ala
 100 105 110

Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
 115 120 125

Thr Ala Ser Val Val Xaa Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
 130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
 145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
 165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
 180 185 190

Ala Xaa Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
 195 200 205

Phe Asn Arg Gly Glu Xaa
 210

ONBI-008001W0_SeqList

<210> 17
 <211> 451
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic Construct

<400> 17

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
 1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
 20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
 35 40 45

Ser Ala Ile Thr Trp Asn Ser Gly His Ile Asp Tyr Ala Asp Ser Val
 50 55 60

Glu Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
 65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
 85 90 95

Ala Lys Val Ser Tyr Leu Ser Thr Ala Ser Ser Leu Asp Tyr Trp Gly
 100 105 110

Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr His Gly Pro Ser
 115 120 125

Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly Gly Thr Ala
 130 135 140

Ala Leu Gly Cys Leu Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val
 145 150 155 160

Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr Phe Pro Ala
 165 170 175

Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val Val Thr Val
 180 185 190

Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Ile Cys Asn Val Asn His
 195 200 205

Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Glu Pro Lys Ser Cys
 210 215 220

Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly

```

      ONBI-008001WQ_SeqList
225              230             235            240

```

Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His
260 265 270

Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val
275 280 285

His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr
290 295 300

Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly
305 310 315 320

Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile
325 330 335

Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val
340 345 350

Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser
355 360 365

Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu
370 375 380

Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro
385 390 395 400

Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val
405 410 415

Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Pro Ser Cys Ser Val Met
420 425 430

His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser
435 440 445

Pro Gly Lys
450

```
<210> 18
<211> 451
<212> PRT
<213> Artificial Sequence
```

<220>
<223> Synthetic Construct

ONBI-008001W0_SeqList

<220>
 <221> MOD_RES
 <222> (148)..(148)
 <223> Xaa = carboxymethyl

<220>
 <221> MOD_RES
 <222> (204)..(204)
 <223> Xaa = carboxymethyl

<220>
 <221> MOD_RES
 <222> (224)..(224)
 <223> Xaa = carboxymethyl

<220>
 <221> MOD_RES
 <222> (230)..(230)
 <223> Xaa = carboxymethyl

<220>
 <221> MOD_RES
 <222> (233)..(233)
 <223> Xaa = carboxymethyl

<220>
 <221> MOD_RES
 <222> (265)..(265)
 <223> Xaa = carboxymethyl

<220>
 <221> MOD_RES
 <222> (325)..(325)
 <223> Xaa = carboxymethyl

<220>
 <221> MOD_RES
 <222> (371)..(371)
 <223> Xaa = carboxymethyl

<220>
 <221> MOD_RES
 <222> (429)..(429)
 <223> Xaa = carboxymethyl

<220>
 <221> MOD_RES
 <222> (451)..(451)
 <223> Xaa = lysine-loss

<400> 18

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
 1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
 20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
 35 40 45

Ser Ala Ile Thr Trp Asn Ser Gly His Ile Asp Tyr Ala Asp Ser Val
 50 55 60

ONBI-008001W0_SeqList

Glu Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
 65 70 75 80
 Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
 85 90 95
 Ala Lys Val Ser Tyr Leu Ser Thr Ala Ser Ser Leu Asp Tyr Trp Gly
 100 105 110
 Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr His Gly Pro Ser
 115 120 125
 Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly Gly Thr Ala
 130 135 140
 Ala Leu Gly Xaa Leu Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val
 145 150 155 160
 Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr Phe Pro Ala
 165 170 175
 Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val Val Thr Val
 180 185 190
 Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Ile Xaa Asn Val Asn His
 195 200 205
 Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Glu Pro Lys Ser Xaa
 210 215 220
 Asp Lys Thr His Thr Xaa Pro Pro Xaa Pro Ala Pro Glu Leu Leu Gly
 225 230 235 240
 Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met
 245 250 255
 Ile Ser Arg Thr Pro Glu Val Thr Xaa Val Val Val Asp Val Ser His
 260 265 270
 Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val
 275 280 285
 His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr
 290 295 300
 Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly
 305 310 315 320
 Lys Glu Tyr Lys Xaa Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile
 325 330 335

ONBI-008001WO_SeqList

Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val
340 345 350

Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser
355 360 365

Leu Thr Xaa Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu
370 375 380

Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro
385 390 395 400

Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val
405 410 415

Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Pro Ser Xaa Ser Val Met
420 425 430

His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser
435 440 445

Pro Gly Xaa
450

<210> 19
<211> 451
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 19

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Ala Ile Thr Trp Asn Ser Gly His Ile Asp Tyr Ala Asp Ser Val
50 55 60

Glu Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

Ala Lys Val Ser Tyr Leu Ser Thr Ala Ser Ser Leu Asp Tyr Trp Gly
100 105 110

ONBI-008001W0_SeqList

Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys Gly Pro Ser
115 120 125

Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly Gly Thr Ala
130 135 140

Ala Leu Gly Cys Leu Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val
145 150 155 160

Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr Phe Pro Ala
165 170 175

Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val Val Thr Val
180 185 190

Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Ile Cys Asn Val Asn His
195 200 205

Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Glu Pro Lys Ser Cys
210 215 220

Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly
225 230 235 240

Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met
245 250 255

Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His
260 265 270

Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val
275 280 285

His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr
290 295 300

Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly
305 310 315 320

Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile
325 330 335

Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val
340 345 350

Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser
355 360 365

Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu
370 375 380

ONBI-008001W0_SeqList

Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro
385 390 395 400

Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val
405 410 415

Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Pro Ser Cys Ser Val Met
420 425 430

His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser
435 440 445

Pro Gly Lys
450

<210> 20
<211> 451
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> MOD_RES
<222> (22)..(22)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (148)..(148)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (174)..(174)
<223> Xaa = lysine-loss

<220>
<221> MOD_RES
<222> (265)..(265)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (429)..(429)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (451)..(451)
<223> Xaa = lysine-loss

<400> 20

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Xaa Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
Page 35

ONBI-008001W0_SeqList

20

25

30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Ala Ile Thr Trp Asn Ser Gly His Ile Asp Tyr Ala Asp Ser Val
50 55 60

Glu Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

Ala Lys Val Ser Tyr Leu Ser Thr Ala Ser Ser Leu Asp Tyr Trp Gly
100 105 110

Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys Gly Pro Ser
115 120 125

Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly Gly Thr Ala
130 135 140

Ala Leu Gly Xaa Leu Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val
145 150 155 160

Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr Xaa Pro Ala
165 170 175

Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val Val Thr Val
180 185 190

Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Ile Cys Asn Val Asn His
195 200 205

Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Glu Pro Lys Ser Cys
210 215 220

Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly
225 230 235 240

Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met
245 250 255

Ile Ser Arg Thr Pro Glu Val Thr Xaa Val Val Val Asp Val Ser His
260 265 270

Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val
275 280 285

His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr

ONBI-008001WO_SeqList

290

295

300

Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly
305 310 315 320

Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile
325 330 335

Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val
340 345 350

Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser
355 360 365

Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu
370 375 380

Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro
385 390 395 400

Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val
405 410 415

Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Pro Ser Xaa Ser Val Met
420 425 430

His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser
435 440 445

Pro Gly Xaa
450

<210> 21
<211> 214
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 21

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Arg Asn Tyr
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Ala Ala Ser Thr Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

ONBI-008001W0_SeqList

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Val Ala Thr Tyr Tyr Cys Gln Arg Tyr Asn Arg Ala Pro Tyr
85 90 95

Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Thr Val Ala Ala
100 105 110

Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
115 120 125

Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
180 185 190

Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
195 200 205

Phe Asn Arg Gly Glu Cys
210

<210> 22
<211> 214
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> MOD_RES
<222> (134)..(134)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (194)..(194)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (214)..(214)
<223> Xaa = carboxymethyl

<400> 22

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly

ONBI-008001W0_SeqList

1	5	10	15
Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Arg Asn Tyr	20	25	30
Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile	35	40	45
Tyr Ala Ala Ser Thr Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly	50	55	60
Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro	65	70	75
Glu Asp Val Ala Thr Tyr Tyr Cys Gln Arg Tyr Asn Arg Ala Pro Tyr	85	90	95
Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Thr Val Ala Ala	100	105	110
Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly	115	120	125
Thr Ala Ser Val Val Xaa Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala	130	135	140
Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln	145	150	155
Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser	165	170	175
Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr	180	185	190
Ala Xaa Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser	195	200	205
Phe Asn Arg Gly Glu Xaa	210		

<210> 23
 <211> 214
 <212> PRT
 <213> Artificial Sequence
 <220>
 <223> Synthetic Construct
 <400> 23

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly	5	10	15
1			

ONBI-008001W0_SeqList

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Arg Asn Tyr
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Ala Ala Ser Thr Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Val Ala Thr Tyr Tyr Cys Gln Arg Tyr Asn Arg Ala Pro Tyr
85 90 95

Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Thr Val Ala Ala
100 105 110

Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
115 120 125

Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
180 185 190

Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
195 200 205

Phe Asn Arg Gly Glu Cys
210

<210> 24
<211> 214
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> MOD_RES
<222> (134)..(134)
<223> Xaa = carboxymethyl

<220>

ONBI-008001W0_SeqList

<221> MOD_RES
<222> (194)..(194)
<223> Xaa = carboxymethyl

<220>
<221> MOD_RES
<222> (214)..(214)
<223> Xaa = carboxymethyl

<400> 24

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Arg Asn Tyr
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Ala Ala Ser Thr Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Val Ala Thr Tyr Tyr Cys Gln Arg Tyr Asn Arg Ala Pro Tyr
85 90 95

Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Thr Val Ala Ala
100 105 110

Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
115 120 125

Thr Ala Ser Val Val Xaa Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
180 185 190

Ala Xaa Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
195 200 205

Phe Asn Arg Gly Glu Xaa
210

<210> 25

ONBI-008001W0_SeqList

<211> 214
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic Construct

<400> 25

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
 1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Arg Asn Tyr
 20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
 35 40 45

Tyr Ala Ala Ser Thr Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
 50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
 65 70 75 80

Glu Asp Val Ala Thr Tyr Tyr Cys Gln Arg Tyr Asn Arg Ala Pro Tyr
 85 90 95

Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Thr Val Ala Ala
 100 105 110

Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
 115 120 125

Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
 130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
 145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
 165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
 180 185 190

Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
 195 200 205

Phe Asn Arg Gly Glu Cys
 210

<210> 26
 <211> 451
 <212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic Construct

<400> 26

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Ala Ile Thr Trp Asn Ser Gly His Ile Asp Tyr Ala Asp Ser Val
50 55 60

Glu Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

Ala Lys Val Ser Tyr Leu Ser Thr Ala Ser Ser Leu Asp Tyr Trp Gly
100 105 110

Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys Gly Pro Ser
115 120 125

Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly Gly Thr Ala
130 135 140

Ala Leu Gly Cys Leu Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val
145 150 155 160

Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr Phe Pro Ala
165 170 175

Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val Val Thr Val
180 185 190

Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Ile Cys Asn Val Asn His
195 200 205

Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Glu Pro Lys Ser Cys
210 215 220

Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly
225 230 235 240

Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met

ONBI-008001W0_SeqList

245

250

255

Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His
260 265 270

Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val
275 280 285

His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr
290 295 300

Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly
305 310 315 320

Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile
325 330 335

Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val
340 345 350

Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser
355 360 365

Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu
370 375 380

Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro
385 390 395 400

Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val
405 410 415

Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met
420 425 430

His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser
435 440 445

Pro Gly Lys
450