

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
10 February 2005 (10.02.2005)

PCT

(10) International Publication Number
WO 2005/011576 A2

(51) International Patent Classification⁷: **A61K**
(21) International Application Number:
PCT/US2004/023908
(22) International Filing Date: 23 July 2004 (23.07.2004)
(25) Filing Language: English
(26) Publication Language: English
(30) Priority Data:
60/490,149 25 July 2003 (25.07.2003) US

(71) Applicant (for all designated States except US): **BEYOND GENOMICS, INC.** [US/US] (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **REGNIER, Fred, E.** [US/US]; 1219 Tuckahoe Lane, West Lafayette, IN 47906-2334 (US). **ADEMEC, Jiri** [CZ/US]; 13 Fordham Drive, Milford, MA 01701 (US). **ZHANG, Xiang** [CN/US]; 1 Village Hill Lane, Apt. 7, Natick, 01760-5723 (US).

(74) Agents: **BASTIAN, Michael, J.** et al.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DIGITAL CHROMATOGRAPHY

(57) **Abstract:** The invention provides peptide selection approaches that facilitate the high throughput screening and characterization of protein mixtures. In various aspects, the invention provides methods for analyzing a mixture of proteins. In various embodiments, the methods generate a plurality of peptides from a mixture of proteins; select sets of peptides from the plurality of peptides; sequence at least a portion of the selected peptides to produce data sets of sequence information; and determine the likelihood of the presence of two or more proteins in the mixture of proteins. In one embodiment, the peptide selection approach selects peptides such that two or more sets of peptides are each individually non-representative of the protein mixture. In another embodiment, peptides are selected such that two or more sets of peptides are each individually representative of less than about 70% but greater than about 30% of the proteins in the protein mixture.

WO 2005/011576 A2

5

DIGITAL CHROMATOGRAPHY**RELATED APPLICATIONS**

This patent application claims the benefit of U.S. Provisional Patent Application Serial No. 60/490,149, entitled "Digital Chromatography", filed July 25, 2003. The
10 entire contents of the above-referenced provisional patent application are incorporated herein by this reference.

BACKGROUND OF THE INVENTION

Proteomics is a field of science relating to the study of the proteome. The proteome refers to the entire complement of proteins in a given sample. Sample complexity, however, can vary from a few thousand proteins to tens of thousands of
15 proteins or more. Accordingly, proteomics has accelerated the need for high throughput screening and characterization of complex biological samples, such as, e.g., serum.

Applications of proteomics include investigating differences in the abundance of certain proteins between two or more sample sets (e.g., disease vs. control), and protein state differences, such as, post-translation modification, splice variants, amino acid
20 polymorphisms, etc.

Typically, a separation stage precedes the measurement stage of a proteome analysis. Two of the most widely used separation strategies are two dimensional gel ("2-D gel") electrophoresis and multi-dimensional liquid chromatography ("MDLC"). These strategies can be performed on either intact proteins, proteins which have been
25 subsequently modified in the laboratory, or proteolytic fragments of proteins known as peptides. Peptides can be produced by a number of methods, such as, e.g., by chemical, enzymatic, or physical means. The measurement stage of a proteome analysis typically involves a single mass spectrometric analysis or multiple mass spectrometric analyses such as tandem mass spectrometric analysis, often referred to as MS/MS.

30 Protein identification today is frequently based on the comparison of experimentally derived tryptic peptide sequences with peptide sequences generated by an *in silico* analysis of genomic data. For example, in a typical MDLC approach, proteins are converted to a mixture of tryptic peptides that are then fractionated

5 chromatographically and analyzed by mass spectrometry, typically by sequencing peptides in the tandem mode of analysis (MS/MS). Analysis in the MS/MS mode is typically achieved by selecting the molecular ion of a peptide (often referred to as “the parent ion”) with a first mass spectrometer (often referred to as the first dimension of mass spectrometry) and directing the parent ion into a collision cell where it collides with
10 an inert gas such as helium. The parent ion is fragmented in the collision cell to a series of fragment ions (often referred to as “daughter ions”), among which are a ladder of ions with sequentially decreasing numbers of amino acids. Two prominent sets of ions are produced. One is a sequence ladder with amino acid deletions from the C-terminal end of the peptide. Another is a ladder with amino acid deletions from the N-terminal end.
15 The daughter ions are then typically directed into a second mass spectrometer (often referred to as the second dimension of mass spectrometry) to resolve the fragmentation pattern of the parent ion, i.e., the selected peptide.

The problem in digesting proteins and then analyzing the peptide cleavage fragments is that too many peptides are generated for direct mass spectral analysis. For
20 example, a mammalian cell at any given time may express as many as 20,000 proteins, and a tryptic digest, for example, generally yields an average of 50 tryptic peptides per protein. As a result, approximately 1,000,000 peptides would be generated from the mammalian cell sample. This means that each fraction from, e.g., a reversed phase chromatography (“RPC”) column capable of producing 100 peaks would contain an
25 average of 10,000 peptides. This is far too many peptides for current mass spectrometers. Going to a higher resolution RPC column with a peak capacity of 200 wouldn’t solve the problem; the fractions from the RPC column would still contain an average of 5,000 peptides each.

Further, when sample complexity increases to the point that multiple peptides
30 have the same molecular weight (isobaric peptides), a series of problems begin to occur when the number of peptides exceeds the analytic capacity of the analysis system. One way the problem of exceeding the analytical capacity manifests itself is the decreased ability to determine whether a mass signal arises from isobaric peptides and, if so, the number of such peptides. There is generally no way to know the number of isobaric
35 peptides or even if isobaric peptides are present unless the peptides are fragmented (e.g., by collision induced dissociation (CID)) and a second dimension of mass spectrometry is

5 used to analyze the fragments. Only when the fragmentation spectra are interpreted can the presence and possibly the number of isobaric peptides be determined. However, the greater the number of isobaric peptides there are for a given mass, the greater the difficulty in interpreting the associated fragmentation spectra. Consequently, the analysis system is presented with a significant problem in producing spectra that can be
10 interpreted when its analytical capacity is exceeded.

Another way the problem of exceeding the analytical capacity of the analysis system can manifest itself is in the swamping of mass signals originating from low abundance peptides. Typically, some peptides in a digest are present in substantially higher concentration than others. When the number of peptides in the sample provided
15 to a mass spectrometer is large, the most abundant peptides can mask those of lower abundance, even in MS/MS. In addition, when there are large numbers of isobaric peptides that vary by orders of magnitude in concentration, spectral peaks from low abundance peptides can look like background noise. Consequently, the analysis system is presented with a significant problem in differentiating between signals from low
20 abundance species and those from high abundance species.

For example, knowledge of the sites of protein phosphorylation and the correlation of phosphate residency with metabolic changes is important to understanding cellular regulation. One problem in determining the state of protein phosphorylation is that potentially important phosphoproteins may only present in low concentrations in the
25 cell, meaning that the phosphopeptides generated from these proteins can be swamped by more abundant species.

For instance, when the total number of peptides in the sample presented to the analysis system is very large, there can be a peptide peak at every mass; however, only the highly abundant peptides will be discernable on what appears as a very high level of
30 background noise. This will be the case in both first and second dimensions of mass spectrometry. As a result, when the total number of peptides presented exceeds the analytical capacity of the analysis system, the dynamic range of the analysis system can be significantly limited (for example, in the case of poorly fractionated tryptic digests of human serum, the dynamic range can be less than two orders of magnitude).

35 One approach to the above problems associated with the large number of peptides in complex samples, is to capture a subset of the peptides from a tryptic digest

5 of a proteome that is representative of essentially all the proteins in the proteome. This representative selection approach focuses on selecting peptides containing specific amino acids that are present in essentially all of the proteins in the proteome. However, any subset of peptides from a complex sample that is representative of the proteome is also likely to contain enough peptides to exceed the analytical capacity of liquid
10 chromatography-mass spectrometry ("LC-MS") systems.

Another approach is the so-called "multi-dimensional protein identification technology" ("MUDPIT"). The MUDPIT approach is designed to capture all proteins on a cation exchange chromatography column, the column is then eluted with step gradients of increasing salt concentration to produce a series of fractions. These fractions are
15 further fractionated by gradient elution RPC. The typical objective in the MUDPIT approach is to analyze and determine the sequence of as many of the peptides in the tryptic digest of the proteome as possible. For example, in the mammalian cell example above, if 10 fractions were selected and each transferred to an RPC column with a peak capacity of 200, the RPC fractions would contain approximately between 500-1,000
20 peptides each. This number of peptides can still exceed the analytical capacity of current mass spectrometry systems. In addition, such MUDPIT approaches are not suited for high throughput screening because it typically takes a long time, many hours to many days, to analyze serially all the fractions from the cation exchange column.

In practice, both the representative selection approach and the MUDPIT approach
25 are insufficient for peptide-based high throughput screening and characterization of mixtures of proteins obtained from complex samples. Accordingly, what is needed is a peptide selection approach that facilitates the high throughput screening and characterization of mixtures of proteins obtained from complex samples.

30 SUMMARY OF THE INVENTION

The present invention provides methods and systems that facilitate the high throughput screening and characterization of mixtures of proteins by providing novel peptide selection approaches. According to various aspects, the invention provides methods for analyzing a mixture of proteins utilizing novel peptide selection approaches.
35 The methods generate a plurality of peptides from a mixture of proteins. The methods select sets of peptides from the plurality of peptides and sequence at least a portion of the

5 selected peptides to produce data sets of, for example, sequence information. These data sets are then used to determine, for example, the likelihood of the presence of two or more proteins in the mixture of proteins.

10 In one aspect, the peptide selection approach of the invention selects peptides such that two or more sets of peptides are each individually non-representative of the protein mixture. In one embodiment, peptides are selected such that two or more sets of peptides are each individually representative of less than about 75% of the proteins in the protein mixture. In another embodiment, peptides are selected such that two or more sets of peptides are each individually representative of less than about 70% but greater than about 30% of the proteins in the protein mixture.

15 In another aspect, the peptide selection approach of the invention selects peptides such that two or more sets of peptides contains less than about 15% of the peptides that are known or predicted to be found in the plurality of peptides generated from the mixture of proteins. In one embodiment, the number of peptides in each of the two or more sets of peptides is in the range from about 0.1% to about 1% of the number of peptides known or predicted to be found in the plurality of peptides.

20 According to another aspect, the peptide selection approach of the invention selects peptides such that the number of peptides in two or more sets of peptides is less than about a peptide number capacity of the sequencing method used to sequence selected peptides to produce data sets. In one embodiment, the sequencing method
25 comprises a MALDI-MS method having a peptide number capacity in the range from about 100 to about 200. In another embodiment, the peptide sequencing method comprises an ESI-MS method having a peptide number capacity in the range from about 30 to about 60. In another embodiment, the sequencing method comprises a RPC-MS method having a peptide number capacity in the range from about 3,000 to about 40,000.
30 In one embodiment, the sequencing method comprises a RPC-(MALDI-MS) method having a peptide number capacity in the range from about 10,000 to about 40,000. In another embodiment, the sequencing method comprises a RPC-(ESI-MS) method having a peptide number capacity in the range from about 3,000 to about 12,000. In one embodiment, the number of peptides in each of the two or more sets of peptides is less
35 than about 150. In another embodiment, the number of peptides in each of the two or more sets of peptides is less than about 50.

5 In another aspect, the methods of the present invention utilize a peptide selection approach that comprises at least three different chromatographic selection techniques used in series to produce the sets of peptides from which data sets are derived to determine the likelihood of the presence of two or more proteins in the protein mixture. In one embodiment, one chromatographic selection technique selects peptides in the
10 plurality of peptides based on peptide acidity, another selects peptides based on the presence or absence of cysteine, and yet another selects peptides based on the presence or absence of histidine.

In the various aspects of the invention, peptides are selected through a plurality of selection techniques that select peptides on the basis of, for example, the presence or
15 absence of specific amino acids, the average number of specific amino acids, the ratio of specific types of amino acids (such as, e.g., the ratio of the number of acidic to basic amino acids), posttranslational modifications (such as, e.g., glycosylation and phosphorylation), or combinations of thereof. Preferably, the selection techniques produce fractionated samples suitable for sequencing by, for example, a RPC-MS method.
20 It is to be understood that in the present invention the sets of selected peptides need not contain a set of signature peptides that are representative of substantially all the proteins in the proteome or mixture of proteins being analyzed.

Preferred selection techniques select peptides from an input analyte to produce sets of peptides such that the component of the input analyte selected for and present in
25 one set of peptides is not substantially present in any other set of peptides. For example, in one approach, a preferred selection technique comprises a LC affinity capture technique that captures Z⁺ components in an input analyte and allows Z⁻ components to flow-through, where Z⁺ components are peptides containing Z (where Z, for example, is an amino acid) and Z⁻ components are peptides not containing Z. Such selection
30 techniques are often referred to as orthogonal selection techniques by analogy to the concept of orthogonal vectors in mathematics. In practice, it is to be understood that even in an orthogonal selection technique the component selected for and present in one set of peptides may nevertheless also be present in very small amounts in another set of peptides because of, for example, well known thermodynamic phenomena that create a
35 distribution of input analyte components between mobile and stationary phases.

5 Chromatographic selection techniques are also preferred that produce a small number of peaks. In one embodiment, a preferred chromatographic selection technique produces only two peaks, each peak corresponding to a set of peptides. For example, in one embodiment, the first peak contains the peptides not retained by the chromatography column, and the second peak contains the peptides initially retained by the column and
10 later desorped from the column by changing, for example, the mobile phase.

In one embodiment, the selection techniques include multiple amino acid and/or posttranslational modification specific LC selection schemes (e.g., $M_a^{\rightarrow}$, $M_{a\rightarrow}$, $M_b^{\rightarrow}$, $M_{b\rightarrow}$, $M_c^{\rightarrow}$, $M_{c\rightarrow}$, $M_d^{\rightarrow}$ and $M_{d\rightarrow}$, where the superscript and subscript arrows represent different analyte sets from a column, e.g., one set may be analytes that flow-through and
15 the other set those that are retained by the column and that require a specific mobile phase for desorption) which are coupled in tandem to produce sets of peptides (e.g., $M_a^{\rightarrow} \Rightarrow M_{b\rightarrow} \Rightarrow M_{c\rightarrow} \Rightarrow S_1$; $M_{a\rightarrow} \Rightarrow M_b^{\rightarrow} \Rightarrow M_c^{\rightarrow} = S_2$; $M_a^{\rightarrow} \Rightarrow M_{b\rightarrow} \Rightarrow M_{d\rightarrow} = S_3$, etc.). In one embodiment the step of selecting sets of peptides further includes using a reversed phase chromatography (RPC) technique.

20 In various embodiments, the step of sequencing includes ionizing peptides in a set of peptides and selecting ionized peptides that have a mass-to-charge ratio in a first mass-to-charge ratio range (the parent ions) with a first dimension of mass spectrometry (such as, e.g., a time-of-flight mass spectrometer (TOF-MS), or quadrapole mass spectrometer), fragmenting the parent ions to produce daughter ions, and generating with
25 a second dimension of mass spectrometry a data set comprising a fragmentation spectrum of the daughter ions. In other embodiments, the step of sequencing includes ionizing peptides in a set of peptides and selecting ionized peptides that have an ion mobility in a first ion range of mobilities (the parent ions) with a first dimension of mass spectrometry (such as, e.g., an ion mobility spectrometer (IMS)), fragmenting the parent
30 ions to produce daughter ions, and generating with a second dimension of mass spectrometry a data set comprising a fragmentation spectrum of the daughter ions. Suitable ionization techniques for ionizing peptides include, but are not limited, electrospray ionization (ESI), matrix-assisted laser desorption/ionization (MALDI), fast atom bombardment (FAB), electron ionization (EI), chemical ionization, desorption
35 chemical ionization (DCI), Negative-ion chemical ionization, field desorption (FD), field

- 5 ionization (FI), Secondary Ion Mass Spectrometry (SIMS), and Atmospheric Pressure Chemical Ionization (APCI).

Preferably, the step of sequencing includes subjecting a set of peptides to reverse phase chromatography to further fractionate the set of peptides prior to analysis by mass spectrometry. For example, in one embodiment, the step of sequencing comprises

- 10 loading a set of peptides into an RPC column to produce a series of fractions from the set of peptides and analyzing the fractions as they elute with a mass spectrometry instrument to produce a data set for the set of peptides.

- In various embodiments, the step of determining the likelihood of the presence of two or more proteins in a mixture of proteins includes comparing at least two or more of
- 15 the data sets of the sequencing step to at least one of a deoxyribonucleic acid ("DNA") database and a protein database. In one embodiment, the data sets are compared to peptide sequences derived from a DNA library. For example, the DNA sequences from either a DNA or Expressed Sequence Tag ("EST") database are converted into a protein and subsequently tryptic peptide sequences *in silico* and then compared with the
- 20 sequence of experimentally derived peptides. When a match is found, it is possible to identify potential parent protein(s) and determine the likelihood that the potential protein(s) are present in the mixture or proteome based, for example, on the matching criteria. In another embodiment, the data sets are compared to peptide sequences derived from a protein database. In one version, tryptic peptide sequences are generated *in silico*
- 25 and then compared with the experimentally derived peptides sequences. In another version, the data sets are compared to experimentally determined tryptic peptide sequences of the proteins in a protein database.

- In one embodiment, the step of determining the likelihood of the presence of two or more proteins in a mixture of proteins preferably includes selecting two or more data
- 30 sets such that in combination the sequence information of the selected data sets is representative of greater than about 80% of the proteins in the mixture of proteins. In another embodiment, preferably data sets are selected such that in combination the sequence information of the selected data sets is representative of greater than about 90% of the proteins in the mixture of proteins.

35

5 As used herein, the term "data set" refers to the spectrometric data associated with one or more multi-dimensional mass spectrometric measurements. For example, where the multi-dimensional mass spectrometric measurement is an MS/MS measurement, a data set can comprise one or more parent ion fragmentation spectra. It should be realized that the term "data set" includes both raw spectrometric data and data
10 that has been processed (e.g., to remove noise, to remove baseline, to detect peaks, to normalize peaks, etc.).

 In another aspect, the present invention provides an article of manufacture where the functionality of a method of the present invention is embedded as computer-readable instructions on a computer-readable medium, such as, but not limited to, a floppy disk, a
15 hard disk, an optical disk, a magnetic tape, a PROM, an EPROM, CD-ROM, or DVD-ROM.

BRIEF DESCRIPTION OF THE DRAWINGS

 The foregoing and other objects, features and advantages of the invention will be
20 more fully understood from the following descriptions of various embodiments of the invention and the accompanying drawings. In the drawings like reference characters generally refer to the same parts throughout the different views. The drawings are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention, where:

25 Figure 1 is a flow diagram of analyzing a mixture of proteins in accordance with various embodiments of the present invention;

 Figure 2 is a graphical representation of a relationship between the number of peptides identifiable by an analytical system and the total number of peptides in a sample;

30 Figure 3 is a flow diagram of selecting proteins from a protein mixture in accordance with various embodiments of the present invention;

 Figures 4A and 4B are flow diagrams of selecting sets of peptides from a plurality of peptides according to various embodiments of the present invention;

 Figures 5A-5C are plots of the average number of proteins containing certain
35 amino acids or combinations of certain amino acids as a function of protein molecular weight based on an *in silico* analysis of the *E. Coli* genome;

5 Figures 5D-5G are plots of the average number of tryptic peptides containing certain amino acids or combinations of certain amino acids as a function of protein molecular weight based on an *in silico* analysis of the *E. Coli* genome;

 Figures 6A and 6B are schematic representations of the series of chromatographic selection techniques of Example 1 used for selecting sets of peptides
10 according one embodiment of the present invention;

 Figure 7A is a schematic representations of the series of chromatographic selection techniques of Example 2 used for selecting sets of peptides from a plurality of peptides generated from human serum albumin ("HSA");

 Figure 7B is a plot of molecular weight versus predicted retention time for all the
15 peptides of Figure 7A analyzed together;

 Figure 8A is a schematic representations of the series of chromatographic selection techniques of Example 3 used for selecting sets of peptides from a plurality of peptides generated from human serum albumin ("HSA") and pregnancy zone protein ("PZP");

20 Figure 8B is a plot of molecular weight versus predicted retention time for all the peptides of Figure 8A analyzed together;

 Figure 8C is a plot of molecular weight versus predicted retention time for the set of peptides of sample S₃ of Figure 8A analyzed separately from the other sets of peptides;

25 Figure 9A schematic representations of the series of chromatographic selection techniques of Example 4 used for selecting sets of peptides from a plurality of peptides generated from a mixture of proteins including PZP;

 Figure 9B is a plot of molecular weight versus predicted retention time for all the peptides of Figure 9A analyzed together; and

30 Figures 9C-9J are plots of molecular weight versus predicted retention time for the separate analysis of the sets of peptides of, respectively, samples S₁-S₈ of Figure 9A.

5 DETAILED DESCRIPTION OF THE INVENTION

Referring to Figure 1, a flow chart of various embodiments of a method of analyzing a mixture of proteins according to the present invention are shown. A mixture of proteins **101** is subjected to a peptide generating step **103** to produce a plurality of peptides **104**. Suitable techniques for generating peptides from proteins include any
10 sequence specific cleavage process. Generation of peptides from proteins can be achieved by chemical, enzymatic or physical means, including, for example, sonication or shearing. Preferably, a protease enzyme is used, such as trypsin, chymotrypsin, papain, gluc-C, serine proteases, thiol proteases, enoproteinase Arg C, enoproteinase Asp N, and enoproteinase Lys C, and pepsin. Alternatively, chemical agents such as
15 cyanogen bromide can be used to effect proteolysis. The proteolytic agent can be immobilized in or on a support, or can be free in solution. A preferred technique for generating a plurality of peptides from the protein mixture includes enzymatic hydrolysis of peptide bonds with trypsin to produce the plurality of peptides. The plurality of peptides **104** is then subjected to a selecting step **105** to select sets of peptides from the
20 plurality of peptides **104**. At least a portion of the peptides in two or more sets of peptides are subjected to a sequencing step **107** to produce a data set for the associated set of peptides. Two or more of the data sets are then used in a determining step **109** to determine the likelihood of the presence of two or more proteins in the protein mixture **101**.

25 The steps of selecting sets of peptides and sequencing peptides are preferably performed utilizing a LC-MS approach **110** that combines the selecting step **105** and sequencing steps **107**. In preferred embodiments, RPC **112** is used to separate peptides in the sets of peptides to provide fractions for the sequencing using a mass spectrometer method. For example, RPC can be used to separate peptides in a set of peptides to
30 provide fractions for subsequent analysis by ESI-MS or MALDI-MS.

In various embodiments, the methods of the present invention further include a step of selecting a set of proteins **114** from a biological sample **116** to generate the mixture of proteins **101**. A wide variety of biological samples are suitable for use in the present invention, including, but not limited to, cells, body fluids (such as, e.g., serum,
35 plasma, urine, lymph, cerebrospinal fluid, amniotic fluid, synovial fluid, sebum, and saliva), tissue, and whole organism homogenates. Proteins may be selected from

5 biological samples, for example, to remove contaminants from the sample, to remove abundant proteins not of interest, or to produce a plurality of protein mixtures for analysis. Suitable protein selection techniques include, but are not limited to, size-exclusion chromatography, ion exchange chromatography, hydrophobic interaction chromatography, gel-filtration chromatography, metal affinity chromatography, dye-
10 based affinity chromatography, antibody based affinity chromatography, and combinations thereof. For example, cibicron blue can be used to select proteins with andenosine 5'-monophosphate ("AMP") binding sites.

The step of selecting peptides to produce sets of peptides comprises subjecting the plurality of peptides to a series of selection techniques until at least two or more of
15 the sets of peptides output from the selection techniques meet the analysis sample criteria of an embodiment of the present invention for a set of peptides. For example, in one aspect, the selecting step 105 selects peptides such that two or more sets of peptides meet the analysis sample criteria that a set of peptides is individually non-representative of the protein mixture 101. In another aspect, the selecting step 105 selects peptides
20 such that two or more sets of peptides meet the analysis sample criteria that a set of peptides is individually representative of less than about 25% of the proteins in the protein mixture 101. In another aspect, the selecting step 105 selects peptides such that two or more sets of peptides meet the analysis sample criteria that a set of peptides contain less than about 25% of the peptides that are known or predicted to be found in
25 the plurality of peptides 104. In yet another aspect, the selecting step 105 selects peptides such that the number of peptides in two or more sets of peptides meets the analysis sample criteria that the number of peptides is less than about a peptide number capacity of the peptide sequencing method of the sequencing step 107.

In preferred embodiments, the selection techniques separate peptides in an input
30 analyte into sets of peptides on the basis of the posttranslational modifications and/or amino acids present or absent from the peptides in the input analyte. For example, one set of peptides may contain components of the input analyte that flow-through a column under an initial set of conditions (e.g., temperature, eluent pH); while the another set of peptides may contain the analyte components that were retained under the initial set of
35 conditions, but which are later eluted under different conditions.

5 The combined sets of peptides produced by a selection technique do not need to contain all of the components of the input analyte and a selection technique can produce more than two sets of peptides. For example, the fractions that elute within different retention time ranges in a chromatographic selection technique can comprise a set of peptides (e.g., one set may comprise eluates with a retention time in the range of 1 min. to 5 min., another set those eluates with a retention time in the range of 5 min. to 15 min., yet another set those eluates with a retention time in the range of 15 min. to 40 min., etc.).

 The methods of the present invention do not require that each set of peptides produced by a selection technique meet the analysis sample criteria of an embodiment of the present invention for a set of peptides. For example, a plurality of peptides can be subject to a first selection technique that produces a first set of peptides and a second set of peptides, neither of which meet the analysis sample criteria of an embodiment of the present invention for a set of peptides. In turn, for example, the first set of peptides can serve as the input analyte for a second selection technique, which produces a third set of peptides and a fourth set of peptides. In addition, although one or both of the third and fourth sets of peptides may meet the analysis sample criteria of an embodiment of the present invention for a set of peptides, neither is required to do so. In this example, sets of peptides are subjected selection techniques until two or more of the sets of peptides meet the analysis sample criteria of an embodiment of the present invention for a set of peptides.

 Peptides selected by a selection technique can be described by the posttranslational modifications and/or amino acids substantially present and/or absent from the selected peptides. A set of peptides is representative of about x% of the proteins in a protein mixture when about x% of the proteins in the protein mixture are known or predicted to have the same combination of posttranslational modification and/or amino acid presences and absences as the selected peptides in the set of peptides.

 For example, a plurality of peptides is generated from a protein mixture and an input analyte comprising the plurality of peptides is subjected to a histidine selection technique to produce: (a) a first set of peptides that contains peptides having substantially no histidines ("His") but that does not substantially contain peptides having one or more His; and (b) a second set of peptides that contains peptides having one or

5 more His but that does not substantially contain peptides having no His. If about x% of the proteins in the protein mixture are know or predicted to contain one or more His, then the first set of peptides is representative of about 1-x% of the proteins in the protein mixture (i.e., the fraction of proteins know or predicted to have no His) and the second set of peptides is representative of about x% of the proteins in the protein mixture.

10 Further, for example, the first and second sets of peptides can each be subjected to a lectin based selection technique for selecting glycopeptides. In this example, where the input analyte to the glycopeptide selection technique is the first set of peptides, the glycopeptide selection technique produces: (a1) a third set of peptides that contains peptides having substantially no His and substantially no glycosylation but that does not

15 substantially contain peptides having one or more His or glycosylation; and (a2) a fourth set of peptides that contains peptides having substantially no His but which have glycosylation, but the fourth set of peptides does not substantially contain peptides having one or more His or which have no glycosylation. Similarly, in this example, where the input analyte to the glycopeptide selection technique is the second set of

20 peptides, the glycopeptide selection technique produces: (b1) a fifth set of peptides that contains peptides having one or more His and substantially no glycosylation but that does not substantially contain peptides having no His or peptides having glycosylation; and (b2) a sixth set of peptides that contains peptides having one or more His and glycosylation but that does not substantially contain peptides having no His or no

25 glycosylation. If about x% of the proteins in the protein mixture are know or predicted to contain one or more His and z% of the proteins in the protein mixture are know or predicted to contain glycosylation, then the third set of peptides is representative of about $(1-x\%)(1-z\%)$ of the proteins in the protein mixture, the fourth set of peptides is representative of about $(1-x\%)(z\%)$ of the proteins in the protein mixture, the fifth set of

30 peptides is representative of about $(x\%)(1-z\%)$ of the proteins in the protein mixture, and the sixth set of peptides is representative of about $(x\%)(z\%)$ of the proteins in the protein mixture; assuming there is no correlation between the presence or absence of His and the presence or absence of glycosylation.

Further, for example, databases predict that about 90% of the proteins in yeast,

35 and about 76% of the proteins in humans, have at least one peptide with both cysteine and histidine; but of the peptides produced by a tryptic digest, only about 3% of these

5 peptides in yeast, and about 6% of these peptides in humans, contain both cysteine and histidine. In this example, a set of peptides produced by selecting peptides that contain both Cys and His could be said, for example, to be representative of about 90% of the proteins in a mixture of yeast proteins, or representative of about 76% of the proteins in a mixture of human proteins.

10 In one aspect, the peptide selection approach of the present invention selects peptides such that the number of peptides in two or more sets of peptides is less than about a peptide number capacity of the sequencing method used to sequence selected peptides and produce data sets. Analytical systems that are designed to analyze multiple components generally have a capacity given by a maximum number of components
15 which they can accommodate in a single analysis. As used herein, the term "peptide number capacity" refers to the maximum number of peptides that can be accommodated in a single scan of a sequencing method. Where the sequencing method is, for example, a mass spectrometry method, the peptide number capacity, P_{nc} , can be determined from the relation,

20

$$P_{nc} = P_{ms}/k, \quad (1)$$

where P_{ms} is peak capacity of the mass spectrometer in a single scan (at the resolution at which the peptide is to be sequenced and the scan range used in sequencing the peptide),
25 and k is the number of isotope peaks and multiple charge state peaks appearing in the mass spectrum from each peptide. For example, the peak capacity, P_{ms} , of a mass spectrometer would be 2000 for an instrument with unit mass resolution and set to scan over a 2000 amu range, such as, e.g., the useful peptide mass range from 500-2500 amu. Increases in resolution and scan range increase the mass spectrometer peak capacity, P_{ms} ,
30 which can be expressed as,

$$P_{ms} = \Delta m/r, \quad (2)$$

where Δm is the scan range and r the mass resolution. Values of k can vary with the
35 type of mass spectrometer. For example, a cluster of 3-4 isotope peaks is often seen in MALDI-MS instruments resulting in a k value in the range of 1-10. Whereas, for

5 example, the generation of ions in multiple charge states in ESI-MS leads to a k value in the range of 1-30 because a peptide can appear as 6-12 different mass-to-charge ratios.

Examples of suitable mass spectrometric methods for sequencing peptides include, but are not limited to, ESI-MS methods, MALDI-MS methods, and multi-dimensional (e.g., MS/MS, MSⁿ) versions of these methods. For example, in one
10 embodiment, the sequencing method comprises utilizing an ESI-MS instrument with a peptide number capacity in the range from about 40 to about 50. For example, in another embodiment, the sequencing method comprises utilizing a MALDI-MS instrument with a peptide number capacity in the range from about 100 to about 150. Suitable mass spectrometers include, for example, time-of-flight, quadrupole, RF
15 multipole, magnetic sector, and electrostatic mass spectrometers.

Where the sequencing method is a chromatography-mass spectrometry method (such as, for example, RPC-MS) the peptide number capacity, P_{nc}, can be determined from the relation,

$$20 \quad P_{nc} = [(P_{ms}/k)P_{ch}] \quad (3)$$

where P_{ms} and k are as described previously, and P_{ch} is the peak capacity of the chromatography system, such as, e.g., a RPC. For example, in one embodiment the sequencing method comprises utilizing an RPC-MS sequencing method having a peptide number capacity in the range from about 8,000 to about 10,000 (such as, for example,
25 using a RPC column with a peak capacity of about 200 and an ESI-MS instrument). In another embodiment, the sequencing technique comprises utilizing an RPC-MS sequencing technique having a peptide number capacity in the range from about 10,000 to about 20,000 (such as, for example, using a RPC column with a peak capacity of about 200 and a MALDI-MS instrument).

30 Where a multi-dimensional chromatography system is used, the peak capacity of the chromatography system, P_{ch}, is the combined peak capacity of the system. For example, where the chromatography system comprises a first dimension of chromatography followed by RPC as a second dimension, the peak capacity of the chromatography technique can be expressed as,

5

$$P_{ch} = P_{ch1} \times P_{rpc}, \quad (4)$$

or more generally by,

$$P_{ch} = \prod_{i=1}^n P_i, \quad (5)$$

where P_{ch1} is the peak capacity in the first dimension, P_{rpc} is the peak capacity of the reversed phase chromatography column, and P_i is the peak capacity of the i^{th} dimension of chromatography for an n dimensional chromatography system.

15 In practice, the maximum number of peptides that a mass spectrometer or chromatography-mass spectrometry system can accommodate is typically less than the peptide number capacity P_{nc} because of isobaric peptides even though these isobaric peptides are of different sequence. For a complex biological sample, isobaric peptides are typically present because the probability of isobaric peptides increases with the total
20 number of peptides, N_{sa} , in a sample of peptides. As a result, it is actually counter-productive to have too many peptides in a sample of peptides submitted to, for example, a reverse phase chromatography mass spectrometry or mass spectrometry analysis because at some point the increasing numbers of isobaric peptides cause increasing numbers of isobaric peptides to occur at multiple masses.

25 The decrease in the ability to identify peptides in a sample of peptides as the number of peptides in the peptide sample increases is schematically illustrated in Figure 2. Figure 2 is a plot 200 of the number of components identifiable, N_{id} , by a chromatography-mass spectrometry (such as, e.g., a RPC-MS system) or a mass spectrometry system as a function of with the number of peptides, N_{sa} , in the peptide sample submitted to the system. Initially, the number of components identified, N_{id} ,
30 increases with the number of peptides, N_{sa} , in the peptide sample, as illustrated by Region A 202 of Figure 2. As N_{sa} increases, the rate at which N_{id} increases begins to decrease and the number of identifiable components approaches a maximum, C_{max} , as illustrated by Region B 204 of Figure 2. Beyond some maximum number of peptides, C_{max} , the number of isobaric peptides in the peptide sample has become so large that
35 further increases in N_{sa} start to reduce the number of components being identified, N_{id} , as

5 illustrated by Region C 206 of Figure 2. As illustrated in Figure 2, $C_{\max}$, represents the experimentally achievable peptide number capacity.

The relationship between the number of peptides in a sample that can be identified, N_{id} , and the total number of peptides in the sample, N_{sa} , over Regions A and B in Figure 2 can be represented as,

10

$$N_{id} = C_{\max} N_{sa} / (J + N_{sa}) \quad (6)$$

where J is a constant related to rate with at which N_{id} approaches the maximum $C_{\max}$.

When N_{sa} is low there is substantially no peak overlap due to isobaric peptides
15 and all peptides can be identified. As illustrated in Region A 202 of Figure 2 and by equation 6, when N_{sa} is low N_{id} is a substantially linear function of N_{sa} (e.g., $N_{sa} \ll J$, so $(J + N_{sa}) \cong J$). This is referred to as the N_{lin} region. As N_{sa} increases, so does the number of isobaric peptides and the relationship between N_{id} and N_{sa} is no longer linear, as illustrated by Region B 204. In Region B 204, increasing numbers of peptides are being
20 identified (N_{id} is increasing) but some are not because of mass overlap. As N_{sa} continues to increase the number of isobaric peptides and mass overlap increases and the number of components being identified, N_{id} , begins to decrease, as illustrated by Region C 206.

Typically, the maximum number of peptides that can be accommodated in a
25 single scan and still be in Region B 204 of Figure 2 (i.e., the experimentally achievable peptide number capacity) with an ESI-MS instrument is typically in the range from about 40 to about 50 peptides, and is typically in the range from about 100 to about 150 with a MALDI-MS instrument. Assuming a chromatographic peak capacity of 200 such as, e.g., provided by a RPC column the experimentally achievable peptide number capacity
30 of a chromatography-mass spectrometry system using RPC and an ESI-MS instrument is in the range from about 8,000 to about 10,000, whereas that of a RPC-MS system using a MALDI-MS instrument as the MS component is in the range from about 10,000 to about 20,000.

5 Protein Selection

 In various embodiments, the methods of the present invention further include a step of selecting a set of proteins from a biological sample to generate a mixture of proteins. A wide variety of techniques can be used to select a set of proteins from a biological sample including, but not limited to size-exclusion chromatography, ion
10 exchange chromatography, hydrophobic interaction chromatography, gel-filtration chromatography, metal affinity chromatography, dye-based affinity chromatography, antibody based affinity chromatography, and combinations thereof. In one embodiment, a set of proteins is selected from a sample using a series of chromatographic selection techniques.

15 For example, some of the difficulties associated with analyzing the proteome of human blood are the high concentrations of human serum albumin ("HSA") and transferrins. Referring to Figure 3, a schematic flow diagram 300 of one embodiment of removing one or more abundant proteins from a human blood sample to produce one or more mixtures of proteins is illustrated. The human blood sample 302 is subjected to a
20 series of affinity LC selection techniques 304 including an affinity column to capture an abundant protein (Protein A, which can be any protein) 305, an anti-HSA antibody affinity column to capture HSA 307 and an anti-transferrin antibody affinity column to capture transferrin 309. In one embodiment, the resultant eluate 310 is further fractionated through, for example, a mixed mode ion exchange protein separation
25 technique 312. In one approach, the resultant eluate 310 is input to an anion exchange ("AEX") column 313 to produce a first set of proteins 314 from the components captured by and later eluted from the AEX column, and a second set of proteins 316 from the flow-through. The second set of proteins 316, or flow through, is input to a cation exchange ("CEX") column 317 to produce a third set of proteins 318 from the
30 flow-through and a fourth set of proteins 319 from the components captured by and later eluted from the CEX column. The first set of proteins 314 (comprising primarily positively charged proteins), the third set of proteins 318 (comprising primarily neutral proteins), and the fourth set of proteins 319 (comprising primarily negatively charged proteins) are preferably desalted using reverse-phase columns 321, 323, 325 to produce
35 protein mixtures 326, 327, 328 for analysis.

5 Chromatographic Based Peptide Selection

Preferably, the peptide selecting step comprises a series of chromatographic selection techniques which are coupled in tandem to produce sets of peptides. Preferably, each chromatographic selection technique selects peptides on the basis of the presence or absence of one or more of the following: (1) specific amino acids; (2) the
10 average number of specific amino acids; (3) the ratio of specific types of amino acids (such as, e.g., the ratio of the number of acidic to basic amino acids); and (4) posttranslational modifications (such as, e.g., glycosylation and phosphorylation). Selection of specific amino acids will generally be of broadest utility in studying protein expression and protein transport. One approach to specific amino acid based selection is
15 to select peptides with a typically low abundance amino acid such as, for example, cysteine and histidine. One concept in specific amino selection is that the peptides containing the specific amino acid(s) can serve as signature peptides for the parent protein from which they were derived. The selection of specific amino acids will generally be of less utility in studying post-translational modifications, mutations, and
20 protein degradation unless the particular amino acid(s) being selected is contained or predicted to be contained in a peptide being selected.

A wide variety of chromatographic selection techniques are suitable for use in the present invention, including, but not limited to, liquid-chromatography, electrophoresis, and gas-chromatography. Examples of suitable chromatographic selection techniques
25 include liquid-chromatography separation techniques such as affinity selection of low abundance amino acids (such as, e.g., cysteine and histidine), anion exchange to select acidic peptides, and strong anion exchange in tandem with strong cation exchange to select neutral peptides. Examples of suitable chromatographic selection techniques also include liquid-chromatography separation techniques directed towards posttranslational
30 modifications such as lectin based selection for selecting glycopeptides, and avidin based selection for selecting phosphopeptides.

Cysteine Selection

Cysteine selection techniques separate peptides on the basis of whether they contain one or more cysteines ("Cys"). DNA databases indicate that in yeast, about 92%
35 of all proteins and about 9% of the peptides produced from these proteins by a tryptic digest contain or are predicted to contain Cys. DNA databases indicate that in humans,

5 about 89% of all proteins and about 17% of the peptides produced from these proteins by
a tryptic digest contain or are predicted to contain Cys. One preferred approach to
cysteine selection uses an affinity selection approach. In one embodiment, an isotope
coded affinity tag is used to select peptides containing one or more Cys. In one
10 approach, Cys containing peptides are labeled with a biotin affinity group derivatized
with a sulfhydryl-specific iodoacetate moiety and a preferably having an isotope coded
linker (e.g., linkers with either eight hydrogens ^1H or eight deuteriums ^2H). One
example of a reagent suitable for isotope coded affinity tagging of Cys-containing
peptides includes the ICATTM reagent available from Applied Biosystems, Inc., Foster
City, California. In one embodiment of cysteine selection, peptides containing one or
15 more Cys are initially retained on an avidin affinity column and peptides with no Cys are
eluted to produce a set of peptides. In this embodiment, after substantially all the Cys
absent peptides have been eluted, the column is then washed to elute peptides containing
one or more Cys to produce another set of peptides.

Suitable chromatographic selection techniques also include the use, for example,
20 of chemical modification to increase or decrease the number of peptides retained by a
chromatographic column. For example, the number of Cys containing peptides retained
by an avidin affinity column can be reduced through a sulfhydryl derivatization process.
One common approach is cysteine based selection to first reduce proteins and then
derivatize them with a biotinylated alkylating agent. When this approach is reversed so
25 that proteins are first derivatized with a biotinylated alkylating agent and then reduced,
only free sulfhydryl groups on the proteins are biotinylated. After reduction, the rest of
the cysteine residues are alkylated with some other reagent such as iodoacetic acid or
vinyl pyridine. Avidin selection later will capture only the small number of peptides
whose cysteine was not involved in a disulfide bridge in the parent protein.

30 Another approach to cysteine based selection is a covalent chromatography
procedure for the selection of Cys-containing peptides such as has been reported by S.
Wang, X. Zhang, and F.E. Regnier in *Journal of Chromatography A*, vol. 949, pp. 153-
160 (2002). In this approach, Cys-containing peptides were selected by covalent
chromatography using thiol disulfide exchange.

35

5 Histidine Selection

Histidine selection techniques separate peptides on the basis of whether they contain one or more histidine ("His"). According to DNA and protein databases, His should occur in about 97% of all proteins in yeast, and constitute about 15% of the peptides in a tryptic digest. In humans, databases indicate that His should occur in about 10 83% of all proteins and in about 17% of all peptides in a tryptic digest. One preferred approach to histidine selection uses an affinity selection approach. In one embodiment, immobilized affinity metal chromatography ("IMAC") is used. In one approach, an IMAC column carrying iminodiacetate complexed with copper II ("Cu⁺²" or "Cu(II)") is used to capture His-containing peptides. As a such an IMAC column can also capture 15 Cys, tryptophan ("Trp") and N-terminal amino groups, it is preferred that steps such as, for example, the reduction, alkylation, and acylation steps in Global Isotope Stable Tagging ("GIST") be used to enhance the selectivity of Cu(II) loaded IMAC columns for His-containing peptides.

Alkylation of cysteine can greatly reduce its binding to Cu(II) loaded IMAC 20 columns, whereas, acylation of primary amine groups in peptides reduces cooperative, electrostatic binding to carboxyl groups in IMAC columns. This latter reduction in cooperative electrostatic binding is a significant factor in reducing tryptophan binding. For example, reduction and alkylation steps (such as used, e.g., before tryptic digestion) can be used to substantially eliminate Cys binding to Cu(II) loaded IMAC columns. 25 Further, for example, acylation can be used to introduce acyl groups that substantially block the binding of Trp and N-terminal amino groups to Cu(II) loaded IMAC columns. Other types of binding in Cu (II) loaded IMAC columns arise from chelation on His and hydrophobic interactions. The number of histidine containing peptides selected by a Cu(II)-IMAC column can be reduced by acylation as well. Free amine groups work 30 cooperatively with the binding histidine on Cu(II) to affect adsorption. Removing primary amines by acylation reduces the number of peptides captured by the column.

Acidic, Basic and Neutral Peptide Selections

Selection techniques based on the average number of specific amino acids or the ratio of specific types of amino acids include, but are not limited to, selection by anion 35 exchange chromatography to select acidic peptides, selection by Ga⁺³-IMAC to select acidic peptides and selection of neutral peptides by a combination of strong anion

5 exchange ("SAX") and strong cation exchange ("SCX") chromatography. Peptides from protein digests vary in the presence of acidic amino acids, i.e., aspartic acid ("Asp" or "E") and glutamic acid ("Glu" or "D"). Some peptides have none, others have one, while still others may have more than two. In the case of *E. Coli*, the genome predicts that about 99.7% of all proteins contain either one glutamate or aspartate. Recalling that
10 substantially all tryptic peptides contain a carboxyl group, substantially all Asp and/or Glu containing peptides produced by a tryptic digest will contain two carboxyl groups. Examination of the *E. Coli* genome also predicts that about 98.0% of all proteins have at least two Asp and/or Glu, about 92.2% have at least three Asp and/or Glu, and about 78.9% have at least four Asp and/or Glu.

15 In various embodiments, acidic peptides are selected using an anion exchange chromatography selection process.

In one embodiment, a strong anion exchange column and a 10 mM buffer at pH 7.5 is used to capture acidic peptides. Under these conditions, substantially none of the peptides with only one acidic amino acid are captured, whereas substantially all of the
20 peptides with three or more acidic amino acids (Asp or Glu) will be captured. In one approach, rather than trying to further fractionate the mixture of captured peptides with varying numbers of acidic amino acids, captured peptides are eluted in a single step with the initial buffer made up to 500 mM in NaCl. Advantages of this approach are that it facilitates quickly selecting (in a period of time shorter than it takes to fractionate a
25 sample on an RPC column) a group of peptides with multiple acidic amino acids and then releases them in a single step.

In another embodiment, the anion exchange chromatography process increases or decreases the number of peptides selected by a SAX column by manipulating the selectivity of the sorbent with mobile phase additives. For example, since SAX columns
30 select peptides based on net charge, peptides that are more anionic will be more strongly retained and require higher ionic mobile phases for elution. In one approach, the ionic strength of the mobile phase is increased to decrease the number of acidic peptides retained. As the ionic strength of the mobile phase increases, the number of acidic amino acids necessary for a peptide to be retained by the SAX column also increases.
35 Since the probability of acidic amino acid occurrence decreases with increasing numbers

5 of acidic amino acids there are fewer of these peptides and the number of peptides captured will decrease.

In another approach, the primary amine groups in the peptides of the input analyte are acylated causing them to become more anionic and resulting in more peptides being retained by the SAX column. In yet another approach, the pH of the mobile phase
10 is decreased to increase the number of peptides retained by the SAX column. Accordingly, acylation, lowering of mobile phase pH, or both can be used to reduce the number of peptides in the flow-through from a SAX column.

In one embodiment, the selecting step includes using a selection technique based on the ratio of specific types of amino acids that separates neutral peptides from acidic
15 and basic peptides. In one approach, a strong anion exchange column and a strong cation exchange column are coupled in series and eluted with a low ionic strength buffer at a fixed pH to elute a set of peptides in the flow-through that are near net neutral in charge. Peptides that are either strongly anionic or strongly cationic will be captured and can be later eluted to produce another set of peptides. The highest stringency of
20 selection can be achieved when the buffer concentration is at or below 10 mM; and more peptides will be eluted as the ionic strength of the mobile phase is increased.

In addition, acylation of primary amine groups of the peptides in an input analyte will cause these peptides to be less acidic. Such acylation results in different peptides being selected by the SCX-SAX tandem column set, the selection criteria being
25 established by a number of parameters such as the pH of the buffer being used, and increases the stringency of the neutral peptide selection, e.g., more peptides will be retained by the SCX-SAX tandem column set. For example, at a buffer pH of approximately 8.5, the SCX-SAX tandem column set retains peptides with a net negative charge of approximately three.

30 Posttranslational Modification Based Selections

In various embodiments, the selecting step includes using a selection technique that selects peptides based on posttranslational modifications (such as, e.g., glycosylation and phosphorylation). Posttranslational modifications are known to be involved in, for example, cellular signal transduction and modulation of protein function, transport and
35 lifetime. For example, addition of carbohydrate groups to certain amino acids can affect protein function, and can be associated with a metabolic or disease state of a cell.

5 Further, for example, reversible phosphorylation is known to be involved in the regulation of cellular signal transduction, gene expression, metabolism, and cell growth.

The posttranslational modifications that result from glycosylation can be complex, resulting in a single amino acid sequence being present in multiple glycoforms. For example, saccharides can form both heterogeneous and homogeneous carbohydrate
10 structures, oligosaccharides can assemble stepwise at serine and threonine residues to form O-linked glycosylation sites, and oligosaccharides can preassemble and attach at asparagine residues to form N-linked glycosylation sites.

In one embodiment, peptides are selected based on their glycosylation using a lectin based selection technique. Lectins are plant proteins that bind to carbohydrate
15 groups with differing specificity. In one approach, glycopeptides are initially retained on a lectin column and peptides with substantially no glycosylation are eluted to produce a set of peptides. In this approach, after the substantially non-glycosylated peptides have been eluted, the glycopeptides are then eluted with a suitable displacing agent to produce another set of peptides.

20 A wide variety of lectins are commercially available, some of which are listed in Table 1 together with their binding specificity. Examples of suitable displacing agents include, for example, methyl- α -D-mannopyranoside for a concanavalin A column, L-fucose for AAA or LTA (*Tetragonolobus purpureus*) column, D-mannose for VFA column, lactose for ECorA column, sialic acid for MMA column, and *N*-
25 acetylglucosamine for a BSI-A4 or BSI-B4 column.

Phosphorylation is a posttranslational modification found in roughly a third of all mammalian cellular proteins. Phosphorylated serine accounts for about 90% of these modifications, phosphorylated threonine for about 10% and phosphorylated tyrosine for about 0.1%.

30 In one embodiment, peptides are selected based on their phosphorylation using an IMAC selection technique. Suitable IMAC columns for capturing phosphopeptides include, but are not limited to, columns carrying nitrilotriacetate and/or iminodiacetate complexed with Fe^{3+} , Ga^{3+} , Al^{3+} , or Zr^{3+} . In various embodiments, an IMAC column carrying an iminodiacetate complexed with Ga^{3+} is preferred. In various embodiments
35 where the sequencing method comprises MALDI-MS, IMAC columns carrying

- 5 nitrilotriacetate complexed with Fe^{3+} or Ga^{3+} are preferred because such columns facilitate the direct ionization of phosphopeptides.

In one approach, peptide cleavage products are loaded onto the IMAC column under acidic conditions (pH 2.0-3.5), unbound nonphosphopeptides are selected by removal from the column with an acidic wash solution to produce one set of peptides, and phosphopeptides are selected by elution under alkaline conditions (~ pH 10) to produce another set of peptides. In another approach, Ga^{3+} complexed IMAC columns are preferable for selection of peptides with multiple phosphate groups, and Fe^{3+} complexed IMAC columns are preferable for selection of monophosphorylated peptides because multiphosphorylated peptides are strongly bound.

15

TABLE 1

Lectin	Affinity
AAA	α -L-Fucosyl residues
ABA	Carbohydrate receptors
BPA	<i>N</i> -Acetyl-D-galactose
BSI-A4	Terminal <i>N</i> -acetyl- α -D-galactosyl residues
BSI-B4	Terminal α -D-galactosyl residues
CCA	<i>N</i> -Acetyl-D-galactosamine
Codium fragile	<i>N</i> -Acetyl-D-galactosamine
Concanavalin A	Terminal α -D-mannosyl and α -D-glycosyl residues
ECorA	<i>N</i> -Acetylglactosamine, <i>N</i> -acetyl-D-galactosamine, lactose, D-galactose
EEA	α -D-Gal(1 $\rightarrow$ 3)[α -L-Fuc(1 $\rightarrow$ 2)]- β -D-Gal(1 $\rightarrow$ 3/4)- β -D-ClcNac
GNL	Terminal α -D-mannosyl residues of glycoconjugates
HAA	Terminal <i>N</i> -acetyl- α -D-galactosaminyl residues
HPA	Terminal <i>N</i> -acetyl- α -D-galactosaminyl residues
LcH	Terminal α -D-mannosyl and α -D-glucosyl residues
LEA	<i>N</i> -Acetyl- β -D-glucosamine oligomers
Limulus polyphemus	<i>N</i> -Acetylated D-hexosamines
MMA	Sialic acid-containing glycoconjugates
MPA	α -D-Galactosyl and <i>N</i> -acetyl-D-galactosaminyl residues
NPA	Terminal and internal α -D-mannosyl residues of glycoconjugates
PSA	Terminal α -D-mannosyl residues
Psophocarus tetragonolobus	<i>N</i> -Acetyl-D-galactosamine
PWM	<i>N</i> -Acetyl- β -D-glucosamine oligomers
RCA ₆₀	<i>N</i> -Acetyl-D-galactosamine

RCA ₁₂₀	Terminal α -D-galactosyl residues
SBA	<i>N</i> -Acetyl-D-galactosamine
SNA	α -NeuNAC-[2>6]-Gal; α -NeuNAC [2>6]-GalNAc; α -NeuNAC[2>3]-Gal
STA	<i>N</i> -Acetyl- β -D-glucosamine oligomers
Tetragonolobus purpureas	α -L-Fucosyl residues
UEA	L-Fucose, <i>N,N'</i> -diacetylchitobiose
VAA	Terminal β -D-galactosyl residues
VFA	D-Mannose, D-glucose
VVA	<i>N</i> -Acetyl-D-galactosamine
WFA	<i>N</i> -Acetyl-D-galactosamine
WGA	<i>N</i> -Acetyl- β -D-glucosaminyl residues, <i>N</i> -acetyl- β -D-glucosamine oligomers

5

In another embodiment, peptides are selected based on their phosphorylation using a technique that uses site-specific chemical modification of the phosphoamino acids followed by introduction of an affinity ligand. In one approach, peptides are subjected to strongly alkaline conditions to chemically modify the phosphoamino acids such that phosphate groups on serine residues undergo β -elimination to form a dehydroalanyl residue, and those on threonine residues yield dehydroamino-2-butyric acid. Both of these elimination products react readily with ethanedithiol. The remaining free thiol group of ethanedithiol can then, for example, be linked to a biotin affinity tag for affinity-based selection on an avidin column.

15 In another approach, peptides are selected based on their phosphorylation using a phosphopeptide biotinylation procedure that introduces an isotope-coded affinity tag. For example, in one version of this approach, the ethanedithiol reagent is with labeled with either four hydrogen or four deuterium atoms. The free sulfhydryl of the ethanedithiol addition product is then tagged with iodoacetyl-PEO-biotin for affinity based-selection of peptides.

20 Selection techniques can be combined in a wide variety of ways for selecting sets of peptides in accordance with the present invention. As used herein, the term "selection protocol" refers to a series of selection techniques. In various embodiments, the plurality of peptides (the original analyte) is subjected to a series of selection techniques where the sets of peptides output from a selection technique serve as the input analyte for the next technique in the series. It is to be understood, for example, that one or more

5 analytes of a set of peptides can be labeled (e.g., with an affinity tag, isotope, etc.) or derivatized prior to their use as an input analyte for a subsequent selection technique.

In preferred embodiments, the selecting step selects peptides on the basis of the presence or absence of two or more of the following: (1) specific amino acids; (2) the average number of specific amino acids; (3) the ratio of specific types of amino acids
10 (such as, e.g., the ratio of the number of acidic to basic amino acids); and (4) posttranslational modifications (such as, e.g., glycosylation and phosphorylation). In one embodiment, the selecting step further includes a RPC technique.

The flow diagrams 400, 420 of Figures 4A and 4B illustrate various embodiments of a step of selecting sets of peptides from a plurality of peptides 401, 421.
15 Although Figures 4A and 4B illustrate a series of three selection techniques, it is to be understood that a series of four or more selection techniques can be used to select peptides to produce sets of peptides in accordance with the present invention. The plurality of peptides (the original analyte) 401, 421 is subjected to a series of selection techniques where the sets of peptides of one technique serve as the input analyte for the
20 next technique in the series. One or more peptides of a set of peptides can be labeled (e.g., with an affinity tag, isotope, etc.) or derivatized prior to their use as an input analyte for a subsequent selection technique.

A set of peptides of one selection technique can serve as the input analyte for another selection technique. For example, as illustrated in Figure 4A, the first set of
25 peptides 403 serves as the input analyte for the second selection technique, M_b , 405a, which produces a third set of peptides 406 and a fourth set of peptides 407 from the first set of peptides 403. The third set of peptides 406 and fourth set of peptides 407 that serve as input analytes to a third selection technique, M_c , 410a, 410b.

Referring to Figure 4A, in one embodiment of the step of selecting sets of
30 peptides, a plurality of peptides 401 (the original analyte) is subjected to a series of selection techniques until at least two or more of the sets of peptides meet the analysis sample criteria of an embodiment of the present invention for a set of peptides. In one embodiment, the plurality of peptides is subjected to a first selection technique, M_a , 402 to produce a first set of peptides 403 and a second set of peptides 404. Preferably, the
35 first set of peptides 403 comprises components of the original analyte that are not substantially present in the second set of peptides 404.

5 In one embodiment, the first selection technique **402** comprises a strong anion exchange LC technique such that the flow-through comprises peptides with less than three aspartic acids (“Asp” or “D”) and/or glutamic acids (“Glu” or “E”) and the initially retained peptides contain three or more D and/or E. In this embodiment, the first set of peptides **403** comprises flow-through and will not substantially contain peptides with
10 three or more D and/or E, and the second set of peptides **404** comprises retained analyte and will not substantially contain peptides with less than three D and/or E. The first set of peptides may nevertheless contain some peptides with three or more D and/or E, and the second set of peptides some peptides with less than three D and/or E, because of, for example, well known thermodynamic phenomena that create a distribution of input
15 analyte components between mobile and stationary phases.

 The first and second sets of peptides **403**, **404** are each subjected to a second selection technique, M_b , **405a**, **405b** to produce a third set of peptides **406** and a fourth set of peptides **407** from the first set of peptides **403**, and to produce a fifth set of peptides **408** and a sixth set of peptides **409** from the second set of peptides **404**. In one
20 embodiment, the second selection technique, M_b , **405a**, **405b** comprises a histidine LC selection technique that separates peptides on the basis of whether they contain one or more histidines (“His”). In one embodiment, peptides containing one or more His are initially retained by a LC column of the second selection technique **405a**, **405b** and peptides with substantially no His are eluted to produce the third set of peptides **406**
25 from the first set of peptides **403**, and the fifth set of peptides **408** from the second set of peptides **404**. In this embodiment, after substantially all the His absent peptides have been eluted, the column is then washed to elute peptides containing one or more His to produce the fourth set of peptides **407** from the first set of peptides **403**, and the sixth set of peptides **409** from the second set of peptides **404**.

30 The first and second sets of peptides can be run through the same LC column, or, more preferably, separate LC columns can be used for the first and second sets of peptides to perform the second selection technique. The use of separate columns, for example, allows for processing the first and second sets of peptides substantially in parallel.

5 In one embodiment, where one or more of the third, fourth, fifth and sixth sets of peptides meet the analysis sample criteria of an embodiment of the present invention for a set of peptides, at least a portion of the peptides of the set(s) of peptides meeting the analysis sample criteria are then sequenced to produce a data set for each sequenced set. Preferably, a set of peptides is subjected to a RPC technique prior to sequencing with,
10 for example, a mass spectrometry instrument. In another embodiment, sets of peptides meeting the analysis sample criteria of an embodiment of the present invention for a set of peptides are subjected to additional selection techniques before sequencing.

In one embodiment, one or more of the third, fourth, fifth and sixth sets of peptides are also subjected to a third selection technique, **M_c, 410a, 410b, 410c, 410d** to
15 produce, respectively, seventh and eighth sets of peptides **411, 412**, ninth and tenth sets of peptides **413, 414**, eleventh and twelfth sets of peptides **415, 416**, and thirteenth and fourteenth sets of peptides **417, 418**. In one embodiment, the third selection technique, **M_c, 410a, 410b, 410c, 410d** comprises a cysteine LC selection technique that separates peptides on the basis of whether they contain one or more Cys. Preferably, all the sets of
20 peptides of the seventh to fourteenth sets of peptides that meet the analysis sample criteria of an embodiment of the present invention for a set of peptides, are then sequenced to produce a data set for each sequenced set of peptides.

The third, fourth, fifth and sixth sets of peptides can also be run through the same LC column, or, more preferably, separate LC columns can be used for the different sets
25 of peptides to perform the third selection technique. The use of separate columns, for example, allows for processing the third, fourth, fifth and sixth sets of peptides substantially in parallel.

Referring to Figure 4B, in another embodiment of the step of selecting sets of peptides, a plurality of peptides **421** (the original analyte) is subjected to a series of
30 selection techniques to produce two or more of the sets of peptides that meet the analysis sample criteria of an embodiment of the present invention for a set of peptides. In one embodiment, the plurality of peptides is subjected to a first selection technique **423** to produce a first set of peptides **424** and a second set of peptides **426**. Preferably, the first set of peptides **424** comprises components of the original analyte that are not
35 substantially present in the second set of peptides **426**.

5 In one embodiment, the first selection technique **423** is an LC technique employing a strong anion exchange column such that the flow-through comprises peptides with less than three D and/or E and the initially retained peptides contain three or more D and/or E. In this embodiment, the first set of peptides **424** comprises flow-through and will not substantially contain peptides with three or more D and/or E, and
10 the second set of peptides **426** comprises retained analyte and will not substantially contain peptides with less than three D and/or E. The first set of peptides may nevertheless contain some peptides with three or more D and/or E, and the second set of peptides some peptides with less than three D and/or E, because of, for example, well known thermodynamic phenomena that create a distribution of input analyte components
15 between mobile and stationary phases.

 The first set of peptides **424** and the second sets of peptides **426** are each subjected to a second selection technique **427a**, **427b** to produce a third set of peptides **428** and a fourth set of peptides **430** from the first set of peptides **424**, and to produce a fifth set of peptides **432** and a sixth set of peptides **434** from the second set of peptides
20 **426**. In one embodiment, the second selection technique **427a**, **427b** comprises a LC histidine selection technique that separates peptides on the basis of whether they contain one or more histidines ("His"). In one embodiment, peptides containing one or more H are initially retained by a LC column of the second selection technique **427a**, **427b** and peptides with substantially n His are eluted to produce the third set of peptides **428** from
25 the first set of peptides **424**, and the fifth set of peptides **432** from the second set of peptides **426**. In this embodiment, after substantially all the His absent peptides have been eluted, the column is then washed to elute peptides containing one or more His to produce the fourth set of peptides **430** from the first set of peptides **424**, and the sixth set of peptides **434** from the second set of peptides **426**.

30 The first and second sets of peptides can be run through the same selection device (e.g., GC column, LC column, etc.) or, more preferably, separate selection devices (e.g., separate GC columns, LC columns, etc.) can be used for the first and second sets of peptides to perform the second selection technique. The use of separate columns, for example, allows for processing the first and second sets of peptides
35 substantially in parallel.

5 The fourth set of peptides **430** is then subjected to a third selection technique **435** to produce a seventh set of peptides **436** and an eighth set of peptides **438** from the fourth set of peptides **430**. In one embodiment, the third selection technique **435** comprises a LC cysteine selection technique that separates peptides on the basis of whether they contain one or more Cys.

10 In one embodiment, two or more the third **428**, fifth **432**, sixth **434**, seventh **436** and eighth **438** sets of peptides that meet the analysis sample criteria of an embodiment of the present invention for a set of peptides are then sequenced to produce a data set for each sequenced set of peptides.

Peptide Sequencing

15 A wide variety of methods and instruments are suitable for sequencing peptides in accordance with the present invention. Preferred methods and instruments include, but are not limited to, sequencing by multi-dimensional mass spectrometry with an ESI-MS or MALDI-MS instrument. Suitable mass spectrometers for an ESI-MS or MALDI-MS instrument include, for example, time-of-flight, quadrupole, RF multipole, magnetic
20 sector, and electrostatic mass spectrometers. Preferably, the step of sequencing includes subjecting a set of peptides to reverse phase chromatography prior to analysis by mass spectrometry. For example, in one embodiment, the step of sequencing comprises loading a set of peptides into an RPC column to produce a series of fractions from the set of peptides and analyzing the fractions as they elute with a mass spectrometry instrument
25 to produce one or more data sets for the set of peptides.

 In one embodiment, the step of sequencing comprises ionizing at least a portion of the peptides in a set of peptides with the ionization source of a mass spectrometer and selecting ionized peptides that have a mass-to-charge ratio (the parent ions) in a first mass-to-charge ratio range with a first dimension of mass spectrometry, fragmenting the
30 parent ions to produce daughter ions, and generating with a second dimension of mass spectrometry a data set comprising a mass spectrum of the daughter ions. Suitable techniques for fragmenting ions include, but are not limited to: collision induced dissociation ("CID") with an inert gas collision partner (such as, e.g., helium, argon and xenon); thermally induced dissociation ("TID"), radiation induced dissociation, electron
35 capture dissociation ("ECD"), multiphoton dissociation "MPD"), photon induced

- 5 dissociation ("PID"); and surface induced dissociation ("SID") with a chemically modified surface.

Preferably, the parent ions are fragmented in a collision cell to a series of fragment ions, among which are a ladder of ions with sequentially decreasing numbers of amino acids. Since the fragmentation can occur anywhere along the peptide, a
10 spectrum of mass-to-charge ratios is generated. Typically, two prominent sets of ions are observed in the fragmentation spectrum. One set is a sequence ladder with amino acid deletions from the C-terminal end of the peptide (often referred to as the y series), while the other set is a sequence ladder with amino acid deletions from the N-terminal end (often referred to as the b series). Complete or partial amino acid sequence
15 information for the parent ions is then obtained by interpretation of the fragmentation spectra. As the different amino acids within a peptide each have different masses, the fragmentation spectrum of a peptide is usually characteristic of the peptide sequence.

Protein Determination

- 20 A wide variety of techniques are suitable for determining the likelihood of the presence of two or more proteins in the mixture of proteins in accordance with the present invention. Preferably, the data sets include the masses and sequences of peptides which are used for database searches to identify proteins. In one embodiment, determining the likelihood of the presence of two or more proteins in the mixture of
25 proteins comprises identifying the proteins present in the mixture of proteins. In another embodiment, determining the likelihood of the presence of two or more proteins comprises identifying potential candidate proteins present in the mixture of proteins and ranking the potential candidate proteins. The ranking of proteins can be based, for example, on the number and extent to which a protein candidate meets the matching
30 criteria of the database search. In yet another embodiment, determining the likelihood of the presence of two or more proteins comprises assigning a probability that the two or more proteins are present in the mixture of proteins.

- Preferably, the step of determining the likelihood of the presence of two or more proteins in a mixture preferably includes selecting two or more data sets such that in
35 combination the sequence information of the selected data sets is representative of greater than about 80% of the proteins in the mixture of proteins. In another

5 embodiment, data sets are selected such that in combination the sequence information of the selected data sets is representative of greater than about 90% of the proteins in the mixture of proteins.

 A data set is representative of about x% of the proteins in a protein mixture when about x% of the proteins in the protein mixture are known or predicted to have the same
10 combination of posttranslational modification and/or amino acid presences and absences as the selected peptides in the set of peptides from which the data set was produced. A combination of data sets is representative of about x% of the proteins in a protein mixture when about x% of the proteins in the protein mixture are known or predicted to have the same combination of posttranslational modification and/or amino acid
15 presences and absences as the selected peptides in the combined sets of peptides from which the combination of data sets was produced.

 Preferably, protein identification, protein candidate determination, and assignment of a probability that a protein is present, are based on the comparison of peptide sequence information from one or more data sets with: (1) peptide sequences
20 generated by an *in silico* analysis of genomic data (such as, e.g., a DNA or EST database); (2) peptide sequences generated by an *in silico* digestion of proteins in a protein database; and/or (3) experimentally determined peptide sequences. One advantage of searching an EST database is that EST data comes from mRNA.

 For example, it is typically possible to identify the protein parent from the
25 sequence information from a single peptide because almost all peptides obtained from a tryptic digest are unique to their parent protein. Those peptides that are unique to their parent are often referred to signature peptides because they may be used as a signature for their protein parent. For example, a target protein can be identified with as little as a two or three amino acid sequence from a single peptide when this sequence information
30 is combined with mass data and the specificity of the digest to generate, for example, a peptide sequence tag. Specific sequence tags can facilitate the identification proteins from, for example, incomplete databases, error-prone databases, and facilitate the analysis of novel proteins.

 It is also possible that a peptide appears in more than one protein. When a
35 peptide appears in more than one protein, these proteins are almost always related and such peptides are often referred to as familial peptides. The fact that two proteins carry

5 the same peptide is generally of biological significance and therefore important. The relationship between related proteins can be, for example, that they have a common biological function, that they are splice variants carrying the same exon, or that they are mutation based variants of the same protein. When a peptide appears in more than one protein, identification of the parent protein from which the peptide is derived typically
10 depends on finding another peptide in the digest that is unique to the protein in question.

In various embodiments, the step of determining the likelihood of the presence of two or more proteins in a mixture of proteins includes comparing at least two or more of the data sets of the sequencing step to at least one of a deoxyribonucleic acid ("DNA") database and a protein database. In one embodiment, the data sets are compared to
15 peptide sequences derived from a DNA library. For example, the DNA sequences from either a DNA or Expressed Sequence Tag ("EST") database are converted into a protein and subsequently tryptic peptide sequences *in silico* and then compared with the peptide sequence information obtained from one or more data sets. When a match is found, it is possible to identify potential parent protein(s) and determine the likelihood that the
20 potential protein(s) are present in the mixture or proteome based, for example, on the matching criteria.

In another embodiment, the data sets are compared to peptide sequences derived from a protein database. In one version, tryptic peptide sequences are generated *in silico* and then compared with the peptide sequence information obtained from one or more
25 data sets. In another version, the data sets are compared to experimentally determined tryptic peptide sequences of the proteins in the protein database.

In various embodiments, the present invention provides increased throughput for proteomic studies of mixtures of proteins. In one set of embodiments, the methods of the present invention increase the throughput of peptide sequencing by increasing the
30 efficiency of the sequencing component of the analysis. In one embodiment, the methods increase the efficiency of mass spectral analysis by producing sets of peptides such that the number of peptides does not exceed the peptide number capacity of the sequencing method.

For example, in one embodiment, the method selects a combination of sets of
35 peptides where the combination of sets of peptides is representative of greater than about 80% of the proteins in the mixture of proteins, and sequences the combination of sets of

5 peptides in less than about 10 hours. In another embodiment, the method selects a combination of sets of peptides where the combination of sets of peptides is representative of greater than about 90% of the proteins in the mixture of proteins, and sequences the combination of sets of peptides in less than about 4 hours. Preferably, the method sequences all the sets of peptides in the combination of sets of peptides, such
10 that data sets are produced for all the sets of peptides in the combination in a time that does not exceed four times the time required for an RPC separation of one of the sets of peptides in the combination.

In one embodiment, each set of peptides selected for the combination of sets of peptides is individually representative of less than about 75% of the proteins in the
15 mixture of proteins. In another embodiment, each set of peptides selected for the combination of sets of peptides is individually representative of between about 30% to about 70% of the proteins in the mixture of proteins.

In another embodiment, the method selects two or more sets of peptides wherein the number of peptides in the two or more sets of peptides is such that the sequencing
20 method is capable of at least about 50% of the peptides in the two sets of peptides in less than about 180 minutes. In one version, the method selects two or more sets of peptides wherein the number of peptides in the two or more sets of peptides is such that a MALDI-MS sequencing method is capable of sequencing at least about 50% of the peptides in the two sets of peptides in less than about 90 minutes. In another version, the
25 method selects two or more sets of peptides wherein the number of peptides in the two or more sets of peptides is such that an ESI-MS sequencing method is capable of sequencing at least about 50% of the peptides in the two sets of peptides in less than about 180 minutes. In yet another embodiment, the method selects two or more sets of peptides wherein the number of peptides in the two or more sets of peptides is such that
30 a RPC-MS sequencing method is capable of sequencing at least about 50% of the peptides in the two sets of peptides in less than about 180 minutes.

In another aspect, the functionality of the methods described above may be implemented as computer-readable instructions on a general purpose computer. The computer may be separate from, detachable from, or integrated into a chromatography-
35 mass spectrometry system. The computer-readable instructions may be written in any one of a number of high-level languages, such as, for example, FORTRAN, PASCAL,

5 C, C++, or BASIC. Further, the computer-readable instructions may be written in a script, macro, or functionality embedded in commercially available software, such as EXCEL or VISUAL BASIC. Additionally, the computer-readable instructions could be implemented in an assembly language directed to a microprocessor resident on a computer. For example, the computer-readable instructions could be implemented in
10 Intel 80x86 assembly language if it were configured to run on an IBM PC or PC clone. In one embodiment, the computer-readable instructions be embedded on an article of manufacture including, but not limited to, a computer-readable program medium such as, for example, a floppy disk, a hard disk, an optical disk, a magnetic tape, a PROM, an EPROM, or CD-ROM.

15

EXAMPLES

EXAMPLE 1: Peptide Selection of Tryptic Digest of *E. Coli* Proteins

Example 1 provides an example of an analysis of a mixture of *E. Coli* proteins. The mixture of *E. Coli* proteins was predicted by an *in silico* analysis of the *E. Coli*
20 genome. An *in silico* analysis of the *E. Coli* genome (F. R. Blattner, G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau and Y. Shao (1997). "The complete genome sequence of *Escherichia coli* K-12." *Science* **277**(5331): 1453-1474) predicts about 4,289 proteins assuming that all
25 possible proteins that *E. Coli* can express are expressed but not including posttranslationally modified proteins. Prediction of the number and sequence of proteins from the *E. Coli* genome was accomplished using a computer program which identified open-reading frames from the genomic DNA sequence (author of computer program is Xiang Zhang, 2002). The *in silico* tryptic digest of these proteins yields about 130,000
30 peptides and was a complete digestion (no missed cleavages). Figures 5A-5G provide further information on the average number of peptides and proteins containing certain amino acids or combinations of certain amino acids as a function of protein molecular weight based on an *in silico* analysis of the *E. Coli* genome.

Figures 5A-5C provide information on the average number of proteins containing
35 certain amino acids as a function of protein molecular weight, where the x-axis in Figures 5A-5C is in kiloDaltons (kDa). Figure 5A is a plot 501 of the average number of

5 Cys containing proteins **502** (filled square symbols) as a function of protein molecular weight, Figure 5B is a plot **503** of the average number of His containing proteins **504** (filled triangle symbols) as a function of protein molecular weight, and Figure 5C is a plot **505** of the average number of His containing proteins **506** (filled diamond symbols) as a function of protein molecular weight.

10 Figures 5D-5G provide information on the average number of peptides containing certain amino acids or combinations of certain amino acids as a function of protein molecular weight, where the x-axis in Figures 5D-5G is also in kiloDaltons (kDa). Figure 5D is a plot **507** that compares the average number of Cys containing tryptic peptides **508** (filled square symbols), the average number of tryptic peptides
15 containing both Cys and one or more of Asp or Glu **509** (filled triangle symbols), and the average number of tryptic peptides containing both Cys and three or more of Asp or Glu **510** (filled diamond symbols) as a function of protein molecular weight.

Figure 5E is a plot **511** that compares the average number of His containing tryptic peptides **512** (filled triangle symbols), the average number of Cys containing
20 tryptic peptides **513** (filled square symbols), and the average number of tryptic peptides containing both Cys and His **514** (filled circle symbols) as a function of protein molecular weight.

Figure 5F is a plot **515** that compares the average number of His containing tryptic peptides **516** (filled triangle symbols), the average number of tryptic peptides
25 containing both His and one or more of Asp or Glu **517** (filled circle symbols), and the average number of tryptic peptides containing both His and three or more of Asp or Glu **518** (filled diamond symbols) as a function of protein molecular weight.

Figure 5G is a plot **519** that compares the average number of tryptic peptides containing one or more of Asp or Glu **520** (filled triangle symbols), the average number
30 of Cys containing tryptic peptides **521** (filled square symbols), and the average number of tryptic peptides containing both Cys and one or more of Asp or Glu **522** (filled diamond symbols) as a function of protein molecular weight.

An example of operating selectivity columns in tandem to select sets of peptides in accordance with embodiments of the present invention from the plurality of peptides
35 generated by the above *in silico* tryptic digest of *E. Coli* proteins is illustrated by the

5 schematic diagrams 600, 601 in Figures 6A and 6B. In this example, all cysteine-containing peptides are considered to be derivatized with biotin and thus selectable with an avidin affinity column. The number of peptides selected by various combinations of
10 above *in silico* analysis of the *E. Coli* genome. In this example, selection techniques are run in tandem to form sets of peptides some of which serve as samples S₁– S₆ for subsequent submittal to a RPC-MS system for sequencing by a RPC-MS sequencing method.

15 Although the absolute numbers of peptides indicated in the Figures 6A and 6B are based on perfect orthogonal peptide selection (for example, all peptides with less than three acidic amino acids are in the first set of peptides, and all peptides with three or more acidic amino acids are in the second set of peptides), in practice the selection techniques illustrated in Figures 6A and 6B are likely to be only substantially orthogonal. As discussed above and as understood by those of ordinary skill in the art, even in an
20 orthogonal selection technique the component selected for and present in one set of peptides may nevertheless also be present in very small amounts in another set of peptides due to thermodynamic phenomena. Absolute numbers of peptides are used in this example to give an idea of the relative numbers of peptides selected and to provide, for example, examples of N_{sa} and the percentage of proteins in the protein mixture
25 represented by the various sets of peptides. Although absolute numbers of peptides selected will vary between organisms, relative selection and percent coverage of the proteome will generally be independent of the organism.

Referring to Figures 6A and 6B, the 4,289 proteins predicted from the *E. Coli* genome (the protein mixture 602 in this example) are subjected to an *in silico* tryptic
30 digest to generate a plurality of peptides (the original analyte) 603 containing 130,000 peptides. The plurality of peptides 603 is first subjected to a LC selection technique 604 to select peptides based on their acidity by employing a strong anion exchange (“SAX”) column such that the flow-through 605 comprises peptides with less than three D and/or E and the initially retained peptides contain three or more D and/or E. In this example,
35 the flow-through from the SAX column comprises the first set of peptides 605 and the initially retained peptides are eluted to produce the second set of peptides 606.

5 The first **605** and second **606** sets of peptides are each then subjected to a second LC selection technique **607a**, **607b** to select peptides based on the presence or absence of His by employing a Cu(II)-IMAC column. The flow-through **609** of these Cu(II)-IMAC column comprises peptides that do not contain His, whereas the retained peptides **608** each contain one or more His. In this example, the peptides in the first set of
10 peptides that are retained by the first Cu(II)-IMAC column **607a** are eluted to produce a fourth set of peptides **610** having 16,040 peptides which together are representative of 3,915 proteins in the protein mixture. The flow-through of the first Cu(II)-IMAC column **607a** produces a fifth set of peptides **611** that is subjected to further LC selection **612a** to select peptides based on the presence or absence of Cys by employing an avidin
15 affinity column. The peptides retained by this avidin affinity column **612a** are then eluted to produce a sixth set of peptides **613** having 6,101 peptides which together are representative of 2,864 proteins in the protein mixture. In this example, at least a portion of the peptides of the fourth set of peptides **610** form the first sample S_1 and at least a portion of the peptides of the sixth set of peptides **613** form the second sample S_2 .

20 The peptides in the second set of peptides that are retained by the second Cu(II)IMAC column **607b** are eluted to produce a seventh set of peptides **614** having 7,677 peptides which together are representative of 3,135 proteins in the protein mixture. The flow-through of the second Cu(II)IMAC column **607b** produces an eighth set of peptides **615** having 11,537 peptides which together are representative of 3,537 proteins
25 in the protein mixture. In this example, at least a portion of the peptides of the seventh set of peptides **614** form the third sample S_3 and at least a portion of the peptides of the eighth set of peptides **615** form a fourth sample S_4 .

 Referring to Figure 6B, in this example, a portion of the eluate that forms the eighth set of peptides **615** is subjected to further LC selection **612b** to select peptides
30 from the eighth set of peptides **615** based on the presence or absence of Cys using an avidin affinity column. The peptides retained **616** by this avidin affinity column **612b** are eluted to produce a ninth set of peptides **617** having 2,115 peptides which together are representative of 1,527 proteins in the protein mixture. The flow-through **618** of this avidin affinity column **612b** produces a tenth set of peptides **619** having 9,421 peptides
35 which together are representative of 3,266 proteins in the protein mixture. At least a

- 5 portion of the peptides of the ninth set of peptides **617** then form the fifth sample S_5 and at least a portion of the peptides of the tenth set of peptides **615** form the sixth sample S_6 .

It will be noted in Figures 6A and 6B that in addition to the specific selection of peptides containing certain amino acids, there is value in removing classes of peptides.

- 10 For example, using the SAX column peptides with multiple acidic amino acids can typically be removed from samples in a few minutes. As illustrated, samples S_1 and S_2 come from the flow-through from the SAX column.

- Table 2 provides for each sample S_1 - S_6 of Example 1, and certain combinations of these samples, a summary of: the number of peptides in the sample or combination of
15 samples; the percentage of the plurality of peptides in the sample or combination of samples; the number of proteins in the protein mixture represented by the sample or combination of samples; and the percentage of the proteins in the protein mixture represented by in the sample or combination of samples.

TABLE 2

Sample or Combination of Samples	Number of Peptides	% of Plurality of Peptides	Number of Proteins Represented	% of Proteins Represented
S_1	16,040	12.3	3,915	91.3
S_2	6,101	4.7	2,864	66.8
S_3	7,677	5.9	3,135	73.1
S_4	11,537	8.9	3,537	82.5
S_5	2,115	1.6	1,527	35.6
S_6	9,421	7.3	3,266	76.1
$S_1 + S_3$	23,717	18.2	4,131	96.4
$S_1 + S_4$	27,577	21.2	4,183	97.5
$S_2 + S_3$	13,778	10.5	3,536	82.4
$S_2 + S_4$	17,638	13.5	3,830	89.3
$S_3 + S_4$	19,214	14.8	3,954	92.2
$S_1 + S_3 + S_4$	35,254	27.1	4,242	98.9
$S_1 + S_2 + S_3 + S_4$	41,355	31.8	4,264	99.4

20

Table 2 illustrates that a variety of combinations of sets of peptides and a variety of combinations of their corresponding data sets can be selected that are representative of greater than about 80% and greater than about 90% of the proteins in the protein mixture **602**. It is seen that the combination of samples S_3 and S_4 (i.e., the seventh and

5 eight sets of peptides) is representative of about 92% of the proteins in the mixture of proteins (e.g., the selection techniques that produce samples S₃ and S₄ result in sets of peptides that together contain signature peptides from 92% of the *E. Coli* proteome of this example).

In addition, Table 2 illustrates that the peptide separation approach of the present
10 invention can provide a greater than 90% reduction in original analyte (the plurality of peptides **603**) complexity. For example, sample S₂ retained only 4.7% of the original 130,000 peptides from the tryptic digest whereas samples S₃, S₄ and S₆ retained 5.9%, 8.9%, and 7.3%, respectively. Complexity can be reduced even further, for example, a
15 set of peptides selected by a serial grouping of SAX, avidin, and Cu(II)IMAC columns, to select peptides with three or more Asp or Glu, one or more His, and one or more Cys, would contain only 1% of the peptides in the original analyte.

It will be recognized that the percentage representation of the mixture of proteins obtained with any one of the samples S₁-S₆ or their associated data sets, is low, generally less than about 90%. This also reduces the complexity of the original analyte such that
20 the number of peptides in a sample submitted for sequencing is less than the peptide number capacity of the RPC-MS sequencing method of this example.

EXAMPLES 2-4

Examples 2-4 provide examples of analysis of the increasingly complex mixtures of human proteins in accordance with various embodiments of the present invention.

25 The plurality of peptides used in each example is generated by a complete (no missed cleavages) *in silico* tryptic digest of the protein mixture using a computer program (author of computer program is Xiang Zhang, 2002).

Figures 7A, 8A and 9A provide schematic representations **700, 800, 900** of the series of chromatographic selection techniques used to select peptides in Examples 2, 3
30 and 4, respectively. The chromatographic selection techniques used in Examples 2-4 are the same. In each of Examples 2-4, the original analyte of the example is first subjected to a LC selection technique **704** to select peptides based on their acidity by employing a SAX column such that the flow-through comprises peptides with less than three D and/or E and the initially retained peptides contain three or more D and/or E. The
35 initially retained peptides and the flow-through of the first LC selection technique **704**

5 are then subjected a second LC selection technique **706a, 706b** to select peptides based on the presence or absence of His by employing a Cu(II)-IMAC column.

Both the initially retained peptides and the flow-through of the second selection techniques **706a, 706b** are then subjected to a third LC selection technique **708a, 708b, 708c, 708d** to select peptides based on the presence or absence of Cys by employing an
 10 avidin affinity column. In Examples 2-4, all cysteine-containing peptides are considered to be labeled with a biotin affinity group derivatized with iodoacetate acid and thus selectable with avidin affinity columns.

The third LC selection techniques produce eight sets of peptides in each of Examples 2-4 that each form one of samples S₁-S₈ for submittal to an RPC-MS
 15 sequencing method. Although the contents of these eight sets of peptides can vary between Examples 2-4, they share the same group of peptide selection criteria across these examples. For example, the peptides of sample S₁ share the selection criteria of peptides with less than three Asp or Glu, but with one or more His, and with one or more Cys. Table 3 summarizes the peptide selection criteria met by samples S₁-S₈ in
 20 Examples 2-4.

TABLE 3

Sample of Examples 2-4	Peptide Selection Criteria Met by Sample
S ₁	peptides with less than three Asp or Glu, but with one or more His, and with one or more Cys
S ₂	peptides with less than three Asp or Glu, but with one or more His, and with no Cys
S ₃	peptides with less than three Asp or Glu, but with no His, and with one or more Cys
S ₄	peptides with less than three Asp or Glu, but with no His, and with no Cys
S ₅	peptides with three or more Asp or Glu, and with one or more His, and with one or more Cys
S ₆	peptides with three or more Asp or Glu, and with one or more His, but with no Cys
S ₇	peptides with three or more Asp or Glu, but with no His, and with one or more Cys
S ₈	peptides with three or more Asp or Glu, but with no His, and with no Cys

Figure 7B, 8B, 8C, and 9B-J Figure provide plots of molecular weight in atomic mass units (amu) versus predicted retention time (% ACN) on a RPC column for the

5 plurality of peptides of Examples 2-4, or individual samples S₁-S₈. The retention times were predicted based on the retention coefficients of Chabanet, C., Yvon, M., *Journal of Chromatography*, vol. 599, pp. 211-225 (1992). Although the predicted retention times do not exactly match experimentally determined retention times, the correlation between experimental and predicted retention times is greater than 0.9. In addition, in Examples
 10 2-4, the predicted retention times do not consider the contribution of modified Cys to peptide retention.

EXAMPLE 2: Peptide Selection of Tryptic Digest of HSA

Example 2 provides an example of an analysis of a plurality of peptides **702** (the
 15 original analyte) generated by an *in silico* tryptic digest of human serum albumin ("HSA") **701**. The *in silico* tryptic digest of HSA yielded 45 peptides. The eight sets of peptides produced **721, 722, 723, 724, 725, 726, 727, 728**, serve as samples S₁-S₈ for subsequent submittal to a sequencing method.

Table 4 provides for each sample S₁-S₈ of Example 2 a summary of: the number
 20 of peptides in the sample or combination of samples; the percentage of the plurality of peptides in the sample or combination of samples; the number of proteins in the protein mixture represented by the sample or combination of samples; and the percentage of the proteins in the protein mixture represented by in the sample or combination of samples.

TABLE 4

Samples of Example 2	Number of Peptides	% of Plurality of Peptides	Number of Proteins Represented	% of Proteins Represented
S ₁	3	6.7	1	100
S ₂	4	8.9	1	100
S ₃	6	13.3	1	100
S ₄	16	35.6	1	100
S ₅	5	11.1	1	100
S ₆	0	0	0	100
S ₇	6	13.5	1	100
S ₈	3	6.7	1	100

25 Figure 7B is a plot **750** of molecular weight in atomic mass units (amu) versus predicted retention time (% ACN) on a RPC column for an input analyte comprising all 45 peptides of the plurality of peptides **702** loaded at once. The various retention times for peptides of various molecular weights are represented by the filled circles **752**. The

- 5 amino acid sequence for HSA (Sequence I.D. No. 1) **754** is given in the upper right corner of the plot **750**.

EXAMPLE 3: Peptide Selection of Tryptic Digest of HSA and PZP

- Example 3 provides an example of an analysis of a plurality of peptides **802**
 10 generated by an *in silico* tryptic digest of a mixture of human proteins consisting of HSA and pregnancy zone protein ("PZP") **801**. The mixture of proteins was assumed to contain less than 0.5 mg% of PZP and 3500-5500 mg% of HSA. The *in silico* tryptic digest of the HSA and PZP mixture yielded 111 peptides. The eight sets of peptides produced **821, 822, 823, 824, 825, 826, 827, 828**, serve as samples S₁–S₈ for subsequent
 15 submittal to a sequencing method.

- Table 5 provides for each sample S₁–S₈ of Example 3 a summary of: the number of peptides in the sample or combination of samples; the percentage of the plurality of peptides in the sample or combination of samples; the number of proteins in the protein mixture represented by the sample or combination of samples; and the percentage of the
 20 proteins in the protein mixture represented by in the sample or combination of samples.

TABLE 5

Sample of Example 3	Number of Peptides	% of Plurality of Peptides	Number of Proteins Represented	% of Proteins Represented
S ₁	8	7.2	2	100
S ₂	13	11.7	2	100
S ₃	14	12.6	2	100
S ₄	52	46.8	2	100
S ₅	7	6.3	2	100
S ₆	1	0.9	1	50
S ₇	8	7.2	2	100
S ₈	8	7.2	2	100

- Figure 8B is a plot **830** of molecular weight in atomic mass units (amu) versus predicted retention time (% ACN) on a RPC column for an input analyte comprising all
 25 111 peptides of the plurality of peptides **802** loaded at once. The retention times for the peptides of PZP are represented by the filled circles **832**, and those for the peptides of HSA by unfilled circles **834**.

5 Figure 8C is a plot **850** of molecular weight in atomic mass units (amu) versus predicted retention time (% ACN) on a RPC column for an input analyte comprising sample S₃ of Example 3 **823**. The amino acid sequences for the peptides of sample S₃ are given in the upper right corner of the plot for both sequences of the peptides of PZP **852** (Sequence I.D. Nos. 2-9) and those for the peptides of HSA **854** (Sequence I.D. Nos. 10-15).

EXAMPLE 4: Peptide Selection of Tryptic Digest of Protein Mixture

Example 4 provides an example of an analysis of a plurality of peptides **902** generated by an *in silico* tryptic digest of a mixture of human proteins consisting of five abundant human proteins and PZP **901**. The five abundant proteins were HSA, alpha-1-antitrypsin, transferrin, fibrinogen, and immunoglobulin-G. The mixture of proteins was assumed to contain equal quantities of PZP and of the five abundant proteins. The *in silico* tryptic digest of this mixture of proteins yielded 277 peptides. The eight sets of peptides produced **921, 922, 923, 924, 925, 926, 927, 928**, serve as samples S₁– S₈ for subsequent submittal to a sequencing method.

Table 6 provides for each sample S₁–S₈ of Example 4 a summary of: the number of peptides in the sample or combination of samples; the percentage of the plurality of peptides in the sample or combination of samples; the number of proteins in the protein mixture represented by the sample or combination of samples; and the percentage of the proteins in the protein mixture represented by in the sample or combination of samples.

TABLE 6

Sample or Combination of Samples	Number of Peptides	% of Plurality of Peptides	Number of Proteins Represented	% of Proteins Represented
S ₁	19	8.4	5	83.3
S ₂	33	14.5	6	100
S ₃	33	14.5	5	83.3
S ₄	134	59.0	6	100
S ₅	10	4.4	4	66.7
S ₆	11	4.8	5	83.3
S ₇	17	7.5	6	100
S ₈	19	8.4	6	100

5 Figure 9B is a plot **930** of molecular weight in atomic mass units (amu) versus predicted retention time (% ACN) on a RPC column for an input analyte comprising all 277 peptides of the plurality of peptides **902** loaded at once. The retention times for the peptides of PZP are represented by the filled circles **932**, and those for the peptides of the other proteins by unfilled circles **934**.

10 Figures 9C-9J provide predicted RPC retention time plots for each of samples S₁-S₈ of Example 4. In Figures 9C-9J, The retention times for the peptides of PZP are represented by the filled circles **932**, and those for the peptides of the other proteins by unfilled circles **934**. Figure 9C is a plot **940** of predicted retention time for an input analyte comprising sample S₁ **921** and provides the amino acid sequences of the peptides of PZP **942** (Sequence I.D. Nos. 16-20) in sample S₁ and those for the other proteins **944** (Sequence I.D. Nos. 21-34) in sample S₁.

 Figure 9D is a plot **945** of predicted retention time for an input analyte comprising sample S₂ **922** and provides the amino acid sequences of the peptides of PZP **946** (Sequence I.D. Nos. 35-43) in sample S₂ and those for the other proteins **948** (Sequence I.D. Nos. 44-67) in sample S₂.

 Figure 9E is a plot **950** of predicted retention time for an input analyte comprising sample S₃ **923** and provides the amino acid sequences of the peptides of PZP **952** (Sequence I.D. Nos. 2-9) in sample S₃ and those for the other proteins **954** (Sequence I.D. Nos. 10-15 and 68-86) in sample S₃.

25 Figure 9F is a plot **955** of predicted retention time for an input analyte comprising sample S₄ **924** and provides the amino acid sequences of the peptides of PZP **956** (Sequence I.D. Nos. 87-122) in sample S₄ and those for the other proteins **954** (Sequence I.D. Nos. 123-220) in sample S₄.

 Figure 9G is a plot **960** of predicted retention time for an input analyte comprising sample S₅ **925** and provides the amino acid sequences of the peptides of PZP **962** (Sequence I.D. Nos. 221-222) in sample S₅ and those for the other proteins **964** (Sequence I.D. Nos. 223-230) in sample S₅.

 Figure 9H is a plot **965** of predicted retention time for an input analyte comprising sample S₆ **926** and provides the amino acid sequences of the peptides of PZP

- 5 **966** (Sequence I.D. Nos. 231) in sample S₆ and those for the other proteins **968**
(Sequence I.D. Nos. 232-241) in sample S₆.

Figure 9I is a plot **970** of predicted retention time for an input analyte comprising sample S₇ **927** and provides the amino acid sequences of the peptides of PZP **972** (Sequence I.D. Nos. 242-243) in sample S₇ and those for the other proteins **974**

- 10 (Sequence I.D. Nos. 244-258) in sample S₇.

Figure 9J is a plot **975** of predicted retention time for an input analyte comprising sample S₈ **928** and provides the amino acid sequences of the peptides of PZP **976** (Sequence I.D. Nos. 259-262) in sample S₈ and those for the other proteins **978** (Sequence I.D. Nos. 263-277) in sample S₈.

- 15 The claims should not be read as limited to the described order or elements unless stated to that effect. While the invention has been particularly shown and described with reference to specific illustrative embodiments, it should be understood that various changes in form and detail may be made without departing from the spirit and scope of the invention as defined by the appended claims. By way of example, any
20 of the disclosed features may be combined with any of the other disclosed features to analyze a mixture of proteins in accordance with the invention. Therefore, all embodiments that come within the scope and spirit of the following claims and equivalents thereto are claimed as the invention.

5 CLAIMS

What is claimed is:

1. A method of analyzing a mixture of proteins comprising the steps of:
10 generating a plurality of peptides from a mixture of proteins;
 selecting sets of peptides from the plurality of peptides such that two or
more of the sets of peptides are individually representative of less than about
75% of the proteins in the mixture of proteins;
 sequencing at least a portion of the peptides in at least two of the two or
15 more set of peptides to produce a data set for at least two of the two or more sets
of peptides; and
 determining the likelihood of the presence of two or more proteins in the
mixture of proteins based on two or more of the data sets.
- 20 2. The method of claim 1 further comprising the step of selecting a set of proteins
from a sample to provide the mixture of proteins.
3. The method of claim 1, wherein the step of generating a plurality of peptides
comprises performing a proteolysis on at least a portion of proteins in the mixture
25 of proteins.
4. The method of claim 3, wherein performing a proteolysis comprises contacting at
least a portion of the proteins in the mixture of proteins with a serine protease.
- 30 5. The method of claim 1, wherein the step of selecting comprises selecting
peptides such that the peptides in any one set of peptides of the two or more sets
of peptides contain amino acids that are not substantially present in the peptides
of another set of peptides of the two or more sets of peptides.

- 5 6. The method of claim 1, wherein the step of selecting comprises selecting peptides such that the number of peptides in each of the two or more sets of peptides is in the range from about 1% to 20% of the number of peptides in the plurality of peptides.
- 10 7. The method of claim 1, wherein the step of selecting comprises selecting peptides such that the number of peptides in each of the two or more of the sets of peptides is less than about 150 peptides.
- 15 8. The method of claim 1, wherein the step of selecting comprises selecting peptides such that the number of peptides in each of the two or more sets of peptides is in the range from about 3,000 peptides to 20,000 peptides.
- 20 9. The method of claim 1, wherein the step of selecting comprises selecting peptides such that the two or more sets of peptides are individually representative of more than about 30% and less than about 70% of the proteins in the protein mixture.
- 25 10. The method of claim 1, wherein the step of selecting comprises using a series of chromatographic selection techniques.
- 30 11. The method of claim 10, wherein at least one of the chromatographic selection techniques comprises a histidine selection technique utilizing Cu^{+2} -immobilized ion metal affinity chromatography.
- 35 12. The method of claim 10, wherein at least one of the chromatographic selection techniques comprises an acidic peptide selection technique.
13. The method of claim 12, wherein the acidic peptide selection technique utilizes anion exchange chromatography.

- 5 14. The method of claim 10, wherein at least one of the chromatographic selection techniques comprises a cysteine selection technique.
15. The method of claim 10, wherein at least one of the chromatographic selection techniques comprises a lectin based selection technique.
- 10 16. The method of claim 10, wherein at least one of the chromatographic selection techniques comprises a phosphopeptide selection technique.
17. The method of claim 10, wherein at least one of the chromatographic selection techniques comprises a reverse phase chromatography technique.
- 15 18. The method of claim 1, wherein the step of selecting comprises using:
 an acidic peptide selection technique utilizing anion exchange chromatography;
 a histidine selection technique utilizing Cu^{+2} -ion metal affinity chromatography; and
 a reverse phase chromatography technique.
- 20 19. The method of claim 18, wherein the step of selecting further comprises using a cysteine selection technique preceding the reverse phase chromatography technique.
- 25 20. The method of claim 1, wherein:
 the step of sequencing comprises using a sequencing method having a peptide number capacity; and
 the step of selecting comprises selecting peptides such that the number of peptides in each set of peptides is less than about the peptide number capacity.
- 30 21. The method of claim 1, wherein the number of peptides in the two or more sets of peptides is such that a matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) method is capable of sequencing at least about 50%
- 35

- 5 of the peptides in the two or more sets of peptides in less than about 90 minutes..
22. The method of claim 1, wherein the number of peptides in the two or more sets of peptides is such that an electrospray ionization mass spectrometry (ESI-MS) method is capable of sequencing at least about 50% of the peptides in the two or
10 more sets of peptides in less than about 180 minutes.
23. The method of claim 1, wherein the step of sequencing comprises sequencing at least about 25% of the peptides in each of the two or more sets of peptides.
- 15 24. The method of claim 1, wherein the step of sequencing comprises using at least one of a MALDI-MS method and an ESI-MS method to sequence at least a portion of the peptides in a set of peptides.
- 20 25. The method of claim 1, wherein the step of sequencing comprises using a reverse phase chromatography-mass spectrometry (RPC-MS) method to sequence at least a portion of the peptides in a set of peptides.
- 25 26. The method of claim 1, wherein the step of determining the likelihood of the presence of two or more proteins in the mixture of proteins comprises selecting two or more of the two or more data sets such that in combination the sequence information of the selected data sets is representative of greater than about 80% of the proteins in the mixture of proteins.
- 30 27. The method of claim 1, wherein the step of determining the likelihood of the presence of two or more proteins in the mixture of proteins comprises selecting two or more of the two or more data sets such that in combination the sequence information of the selected data sets is representative of greater than about 90% of the proteins in the mixture of proteins.

- 5 28. The method of claim 1, wherein the step of determining the likelihood of the presence of two or more proteins in the mixture of proteins comprises comparing one or more of the two or more data sets to at least one of a DNA database and a protein database.
- 10 29. An article of manufacture comprising a computer-readable media with computer-readable instructions embodied thereon for performing the method of claim 1.
30. A method of analyzing a mixture of proteins comprising the steps of:
 generating a plurality of peptides from a mixture of proteins;
15 selecting sets of peptides from the plurality of peptides such that the number of peptides in two or more sets of peptides is in the range from about 1% to 30% of the number of peptides in the plurality of peptides;
 sequencing at least a portion of the peptides in each of the two or more sets of peptides to produce a data set for each of the two or more sets of peptides;
20 and
 determining the likelihood of the presence of two or more proteins in the mixture of proteins based on two or more of the data sets.
31. The method of claim 30, wherein the step of generating a plurality of peptides
25 comprises contacting at least a portion of the proteins in the mixture of proteins with a serine protease.
32. The method of claim 30, wherein the step of selecting comprises selecting
30 peptides such that the peptides in any one set of peptides of the two or more sets of peptides contain amino acids that are not substantially present in the peptides of another set of peptides of the two or more sets of peptides.
33. The method of claim 30, wherein the step of selecting comprises selecting
35 peptides such that each of the two or more sets of peptides is individually representative of less than about 75% of the proteins in the mixture of proteins.

- 5 34. The method of claim 30, wherein the step of selecting comprises using a series of chromatographic selection techniques.
35. The method of claim 34, wherein at least one of the chromatographic selection techniques comprises a histidine selection technique utilizing Cu^{+2} -immobilized
10 ion metal affinity chromatography.
36. The method of claim 34, wherein at least one of the chromatographic selection techniques comprises an acidic peptide selection technique.
- 15 37. The method of claim 36, wherein the acidic peptide selection technique utilizes anion exchange chromatography.
38. The method of claim 34, wherein at least one of the chromatographic selection techniques comprises a cysteine selection technique.
20
39. The method of claim 34, wherein at least one of the chromatographic selection techniques comprises a lectin based selection technique.
40. The method of claim 34, wherein at least one of the chromatographic selection
25 techniques comprises a phosphopeptide selection technique.
41. The method of claim 34, wherein at least one of the chromatographic selection techniques comprises a reverse phase chromatography technique.
- 30 42. The method of claim 30, wherein the step of selecting comprises using:
 an acidic peptide selection technique utilizing anion exchange chromatography;
 a histidine selection technique utilizing Cu^{+2} -ion metal affinity chromatography; and
35 a reverse phase chromatography technique.

- 5 43. The method of claim 42, wherein the step of selecting further comprises using a
cysteine selection technique preceding the reverse phase chromatography
technique.
- 10 44. The method of claim 30, wherein the step of sequencing comprises sequencing at
least about 25% of the peptides in each of the two or more sets of peptides.
- 15 45. The method of claim 30, wherein the step of sequencing comprises using at least
one of a MALDI-MS method and an ESI-MS method to sequence at least a
portion of the peptides in a set of peptides.
- 20 46. The method of claim 30, wherein the step of sequencing comprises using a
reverse phase chromatography mass spectrometry (RPC-MS) method to sequence
at least a portion of the peptides in a set of peptides.
- 25 47. The method of claim 30, wherein the step of determining the likelihood of the
presence of two or more a proteins in the mixture of proteins comprises selecting
two or more of the two or more data sets such that in combination the sequence
information of the selected data sets is representative of greater than about 80%
of the proteins in the mixture of proteins.
- 30 48. The method of claim 30, wherein the step of determining the likelihood of the
presence of two or more a proteins in the mixture of proteins comprises selecting
two or more of the two or more data sets such that in combination the sequence
information of the selected data sets is representative of greater than about 90%
of the proteins in the mixture of proteins.
- 35 49. The method of claim 30, wherein the step of determining the likelihood of the
presence of two or more proteins in the mixture of proteins comprises comparing
one or more of the two or more data sets to at least one of a DNA database and a
protein database.

- 5 50. An article of manufacture comprising a computer-readable media with computer-readable instructions embodied thereon for performing the method of claim 30.
51. A method of analyzing a mixture of proteins comprising the steps of:
 generating a plurality of peptides from a mixture of proteins;
10 selecting sets of peptides from the plurality of peptides such that the number of peptides in two or more sets of peptides is less than about a peptide number capacity of a peptide sequencing method;
 sequencing at least a portion of the peptides in at least two of the two or more sets of peptides using the peptide sequencing method to produce a data set
15 for at least two of the two or more sets of peptides; and
 determining the likelihood of the presence of two or more proteins in the mixture of proteins based on two or more of the data sets.
- 20 52. The method of claim 51, wherein the step of generating a plurality of peptides comprises contacting at least a portion of the proteins in the mixture of proteins with a serine protease.
- 25 53. The method of claim 51, wherein the step of selecting comprises selecting peptides such that the peptides in any one set of peptides of the two or more sets of peptides contain amino acids that are not substantially present in the peptides of another set of peptides of the two or more sets of peptides.
- 30 54. The method of claim 51, wherein the step of selecting comprises selecting peptides such that each of the two or more sets of peptides is individually representative of less than about 30% of the proteins in the mixture of proteins.
- 35 55. The method of claim 51, wherein the step of selecting comprises selecting peptides such that the two or more sets of peptides are individually representative of more than about 30% and less than about 70% of the proteins in the protein mixture.

- 5 56. The method of claim 51, wherein the step of selecting comprises using a series of chromatographic selection techniques.
57. The method of claim 56, wherein at least one of the chromatographic selection techniques comprises a histidine selection technique utilizing Cu^{+2} -immobilized ion metal affinity chromatography.
- 10 58. The method of claim 56, wherein at least one of the chromatographic selection techniques comprises an acidic peptide selection technique.
- 15 59 The method of claim 58, wherein the acidic peptide selection technique utilizes anion exchange chromatography.
60. The method of claim 56, wherein at least one of the chromatographic selection techniques comprises a cysteine selection technique.
- 20 61. The method of claim 56, wherein at least one of the chromatographic selection techniques comprises a lectin based selection technique.
62. The method of claim 56, wherein at least one of the chromatographic selection techniques comprises a phosphopeptide selection technique.
- 25 63. The method of claim 56, wherein at least one of the chromatographic selection techniques comprises a reverse phase chromatography technique.
- 30 64. The method of claim 51, wherein the step of selecting comprises using:
 a acidic peptide selection technique utilizing anion exchange chromatography;
 a histidine selection technique utilizing Cu^{+2} -ion metal affinity chromatography; and
35 a reverse phase chromatography technique.

- 5 65. The method of claim 64, wherein the step of selecting further comprises using a
cysteine selection technique preceding the reverse phase chromatography
technique.
- 10 66. The method of claim 51, wherein the step of sequencing comprises sequencing at
least about 25% of the peptides in the two or more sets of peptides.
- 15 67. The method of claim 51, wherein the step of sequencing comprises using a
MALDI-MS method having a peptide number capacity in the range from about
100 to about 200.
- 20 68. The method of claim 51, wherein the step of sequencing comprises using an ESI-
MS method having a peptide number capacity in the range from about 30 to
about 60.
- 25 69. The method of claim 51, wherein the step of sequencing comprises using a
reverse phase chromatography mass spectrometry (RPC-MS) method to sequence
at least a portion of the peptides in a set of peptides.
- 30 70. The method of claim 51, wherein the step of determining the likelihood of the
presence of two or more proteins in the mixture of proteins comprises selecting
two or more of the two or more data sets such that in combination the sequence
information of the selected data sets is representative of greater than about 80%
of the proteins in the mixture of proteins.
- 35 71. The method of claim 51, wherein the step of determining the likelihood of the
presence of two or more proteins in the mixture of proteins comprises selecting
two or more of the two or more data sets such that in combination the sequence
information of the selected data sets is representative of greater than about 90%
of the proteins in the mixture of proteins.

- 5 72. The method of claim 51, wherein the step of determining the likelihood of the presence of two or more proteins in the mixture of proteins comprises comparing one or more of the two or more data sets to at least one of a DNA database and a protein database.
- 10 73. An article of manufacture comprising a computer-readable media with computer-readable instructions embodied thereon for performing the method of claim 51.
74. A method of analyzing a mixture of proteins comprising the steps of:
- 15 providing a plurality of peptides generated from a mixture of proteins;
- selecting a first set of peptides from the plurality of peptides using a first selection protocol such that the first set of peptides is representative of less than about 75% of the proteins in the mixture of proteins;
- selecting a second set of peptides from the plurality of peptides using a second selection protocol such that the second set of peptides is representative of less than about 75% of the proteins in the mixture of proteins;
- 20 selecting a third set of peptides from the plurality of peptides using a third selection protocol such that the second set of peptides is representative of less than about 75% of the proteins in the mixture of proteins;
- sequencing at least 95% of the first set of peptides to produce a first data set comprising sequence information for the sequenced peptides of the first set of peptides;
- 25 sequencing at least 95% of the second set of peptides to produce a second data set comprising sequence information for the sequenced peptides of the second set of peptides;
- 30 sequencing at least 95% of the third set of peptides to produce a third data set comprising sequence information for the sequenced peptides of the third set of peptides; and
- determining the likelihood of the presence of two or more proteins in the mixture of proteins based on the first data set, the second data set and the third data set.
- 35

- 5 75. The method of claim 74, wherein a different selection protocol is used to select each set of peptides.
76. The method of claim 76, wherein each selection protocol selects a substantially different set of peptides from the plurality of peptides.
- 10 77. The method of claim 74, wherein the step of selecting comprises selecting peptides such that each of the first, second and third sets of peptides is individually representative of less than about 70% of the proteins in the mixture of proteins.
- 15 78. The method of claim 74, wherein each selection protocol comprises using a series of chromatographic selection techniques.
- 20 79. The method of claim 74, wherein the step of determining the likelihood of the presence of two or more proteins in the mixture of proteins comprises selecting at least two of the first, second and third data sets such that in combination the sequence information of the selected data sets is representative of greater than about 80% of the proteins in the mixture of proteins.
- 25 80. The method of claim 74, wherein the step of determining the likelihood of the presence of two or more proteins in the mixture of proteins comprises selecting at least two of the first, second and third data sets such that in combination the sequence information of the selected data sets is representative of greater than about 90% of the proteins in the mixture of proteins.
- 30 81. The method of claim 74, wherein the step of determining the likelihood of the presence of two or more proteins in the mixture comprises comparing one or more of the two or more data sets to at least one of a DNA database and a protein database.
- 35

- 5 82. An article of manufacture comprising a computer-readable media with computer readable instructions embodied thereon for performing the method of claim 74.

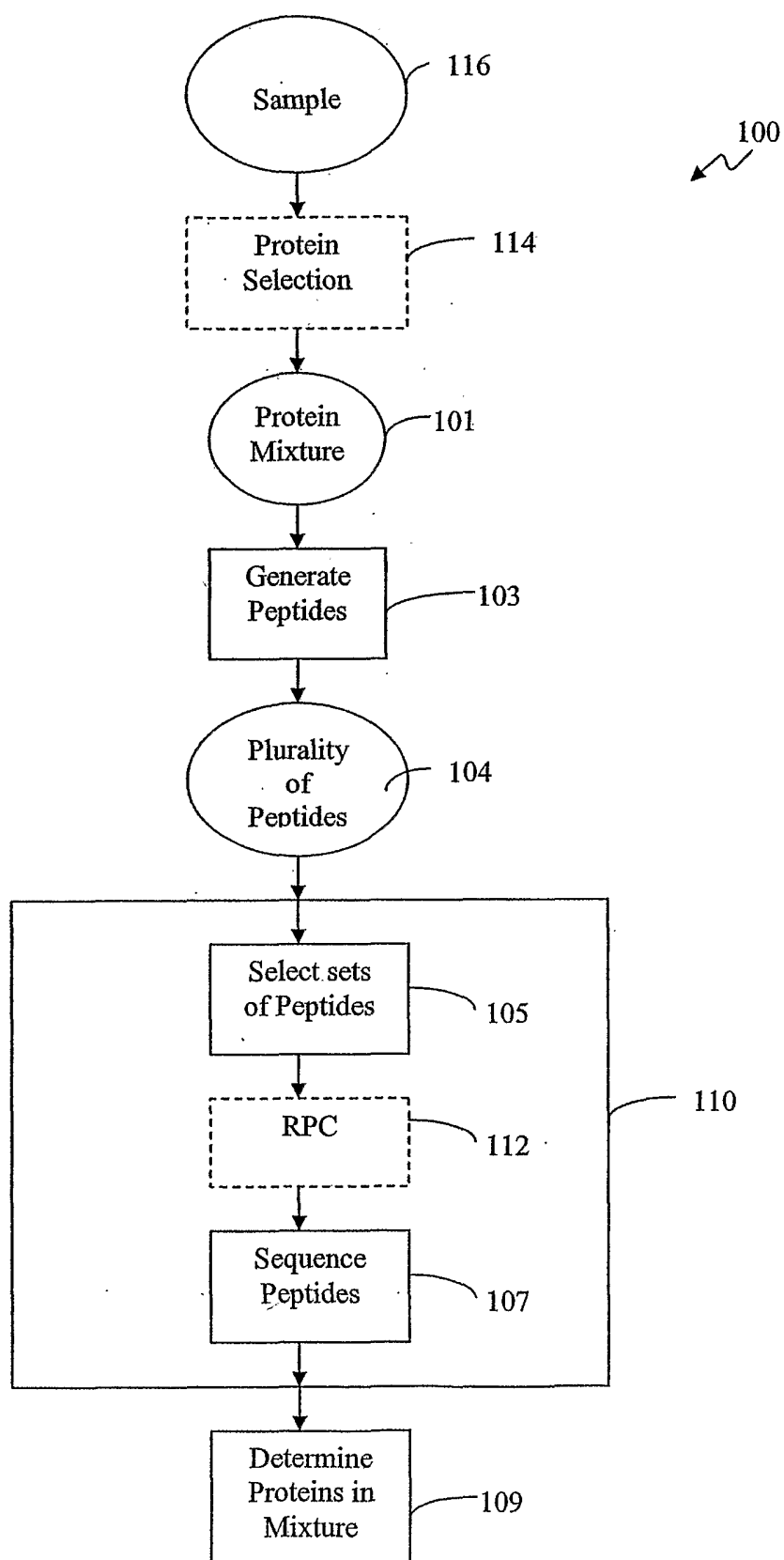


FIG. 1

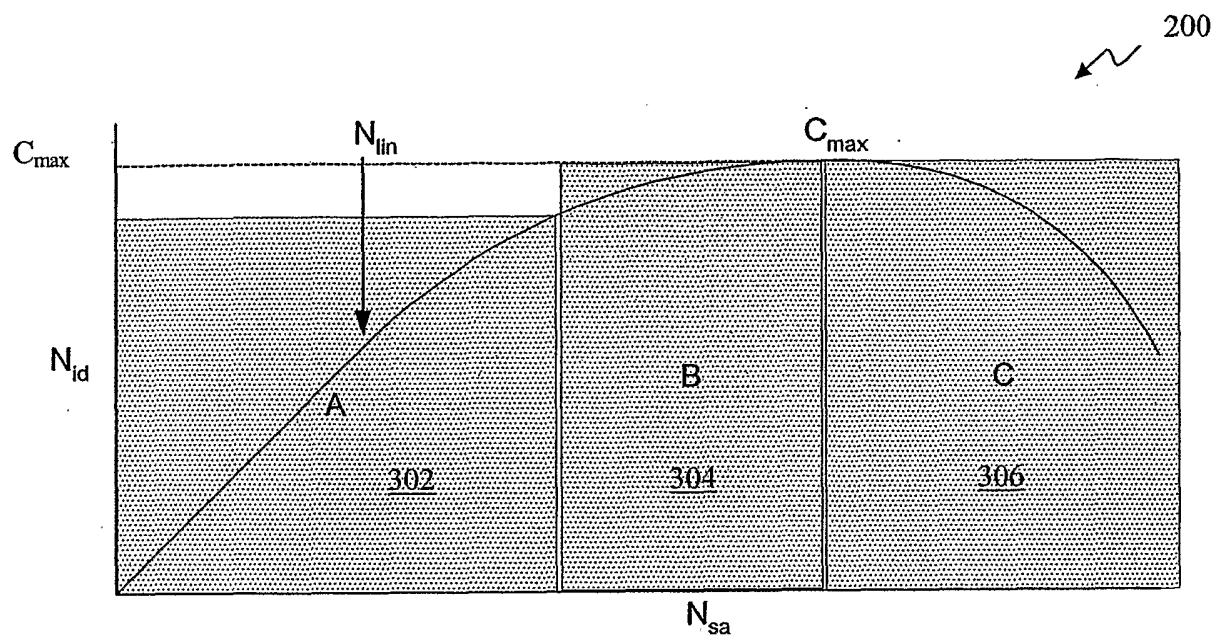


FIG. 2

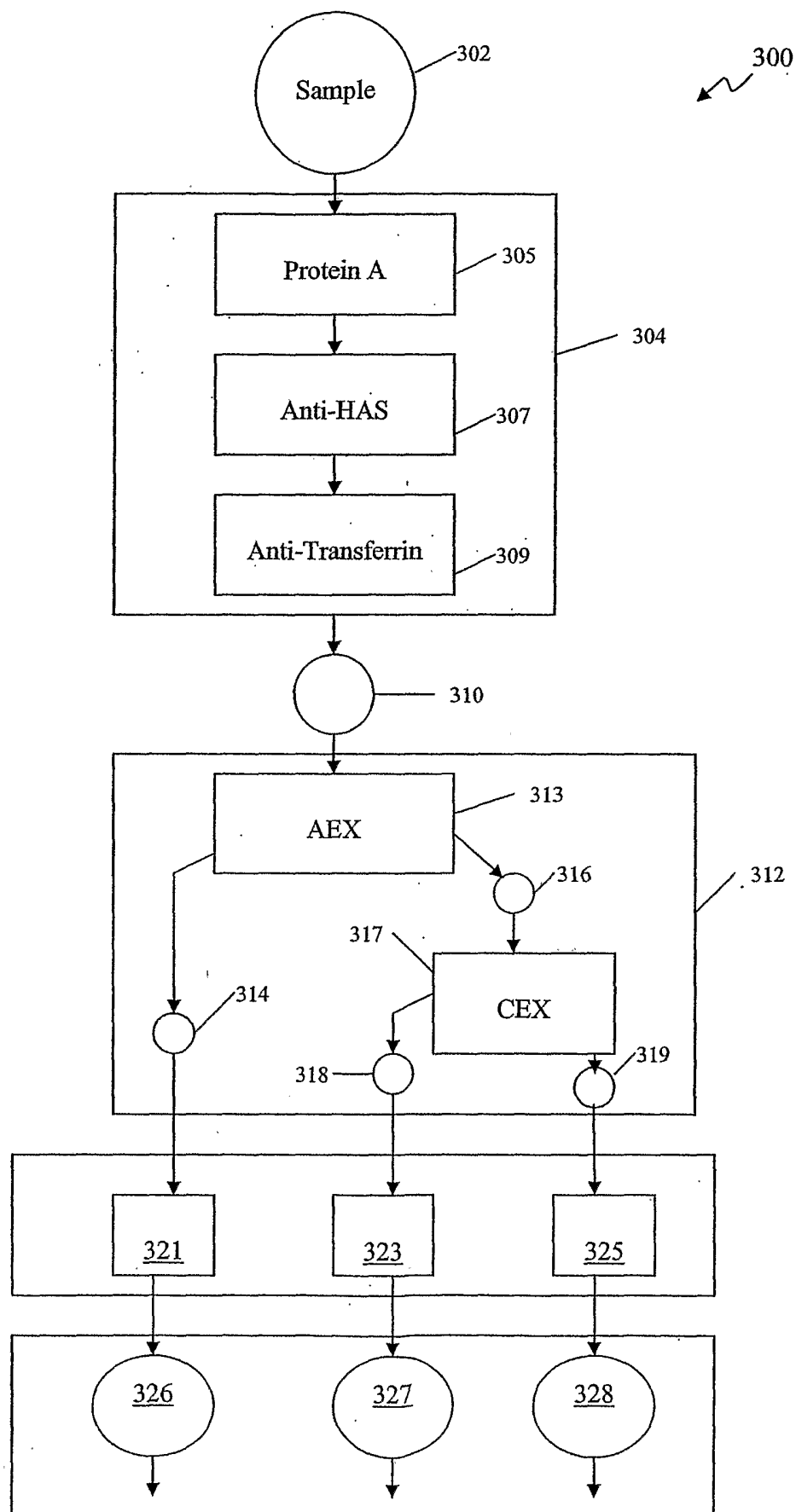


FIG. 3

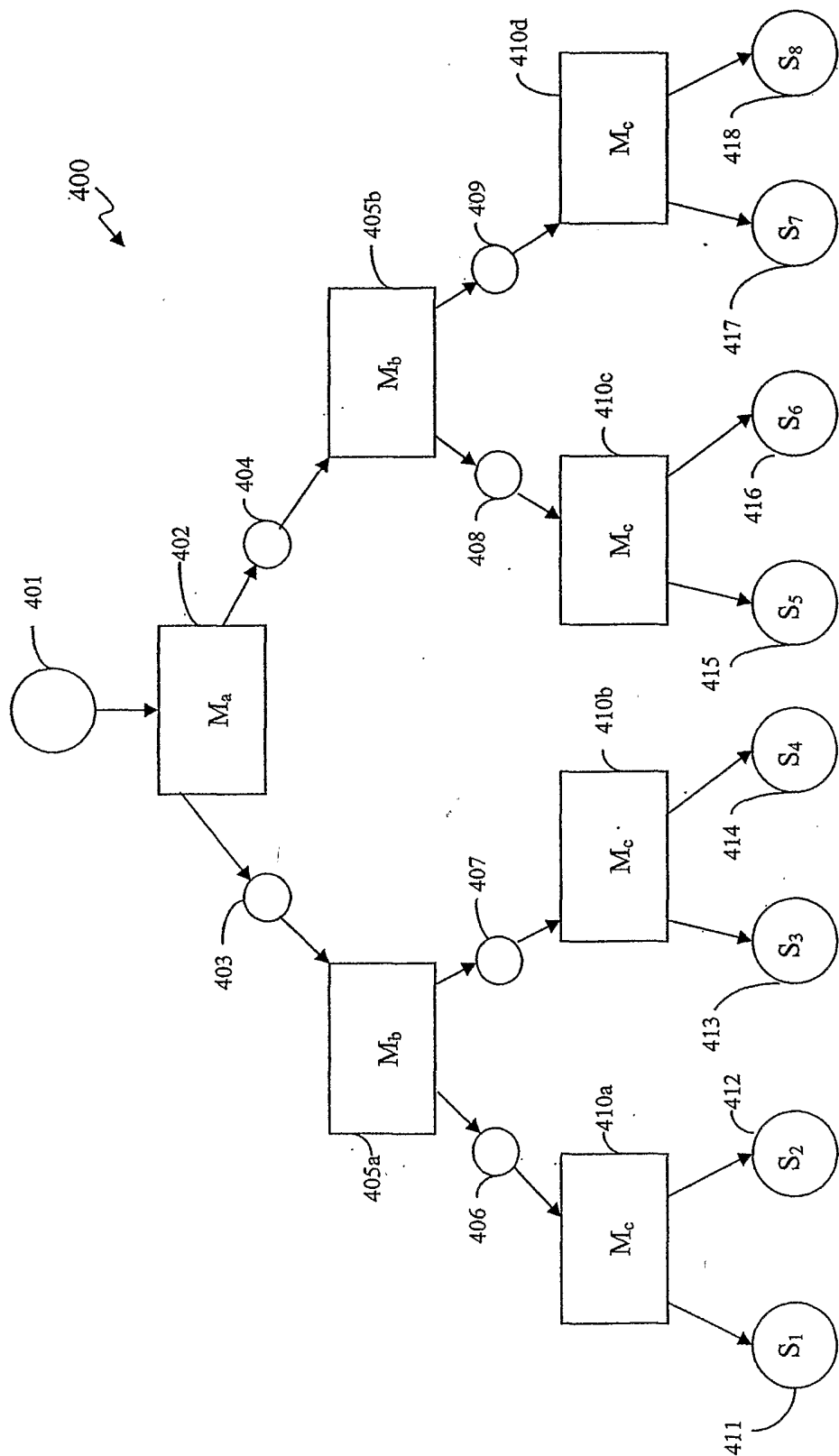


FIG. 4A

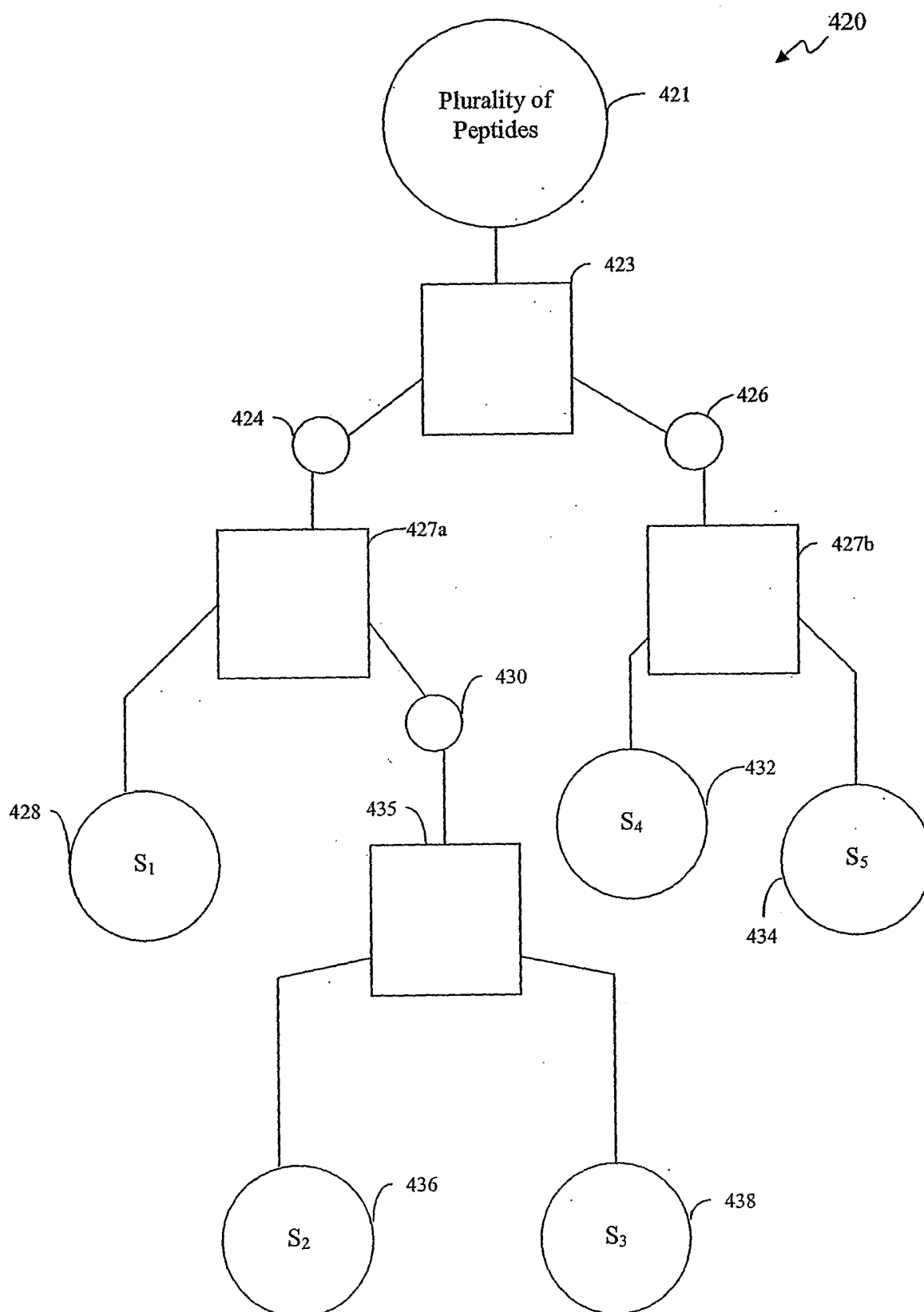


FIG. 4B

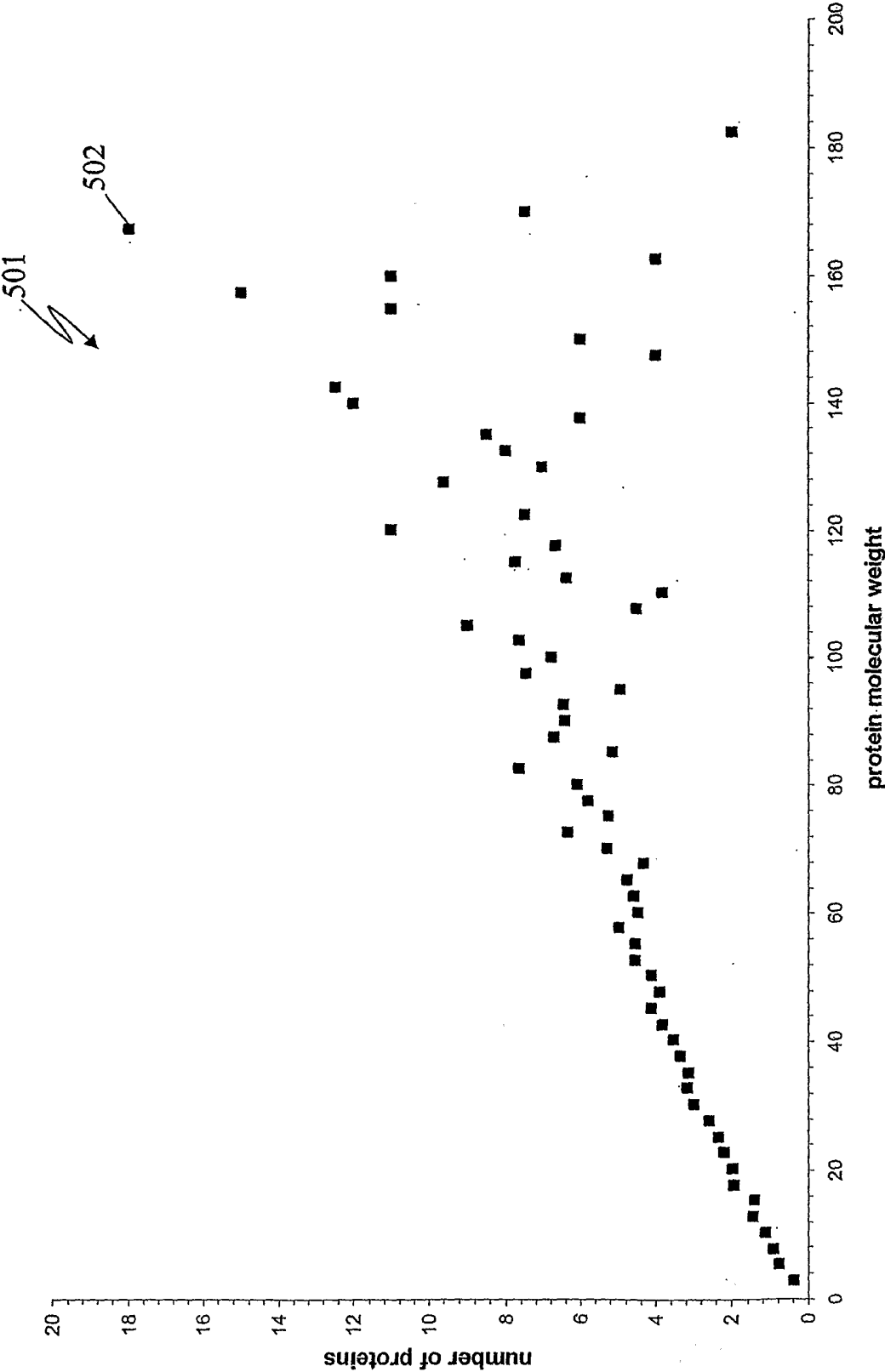


FIG. 5A

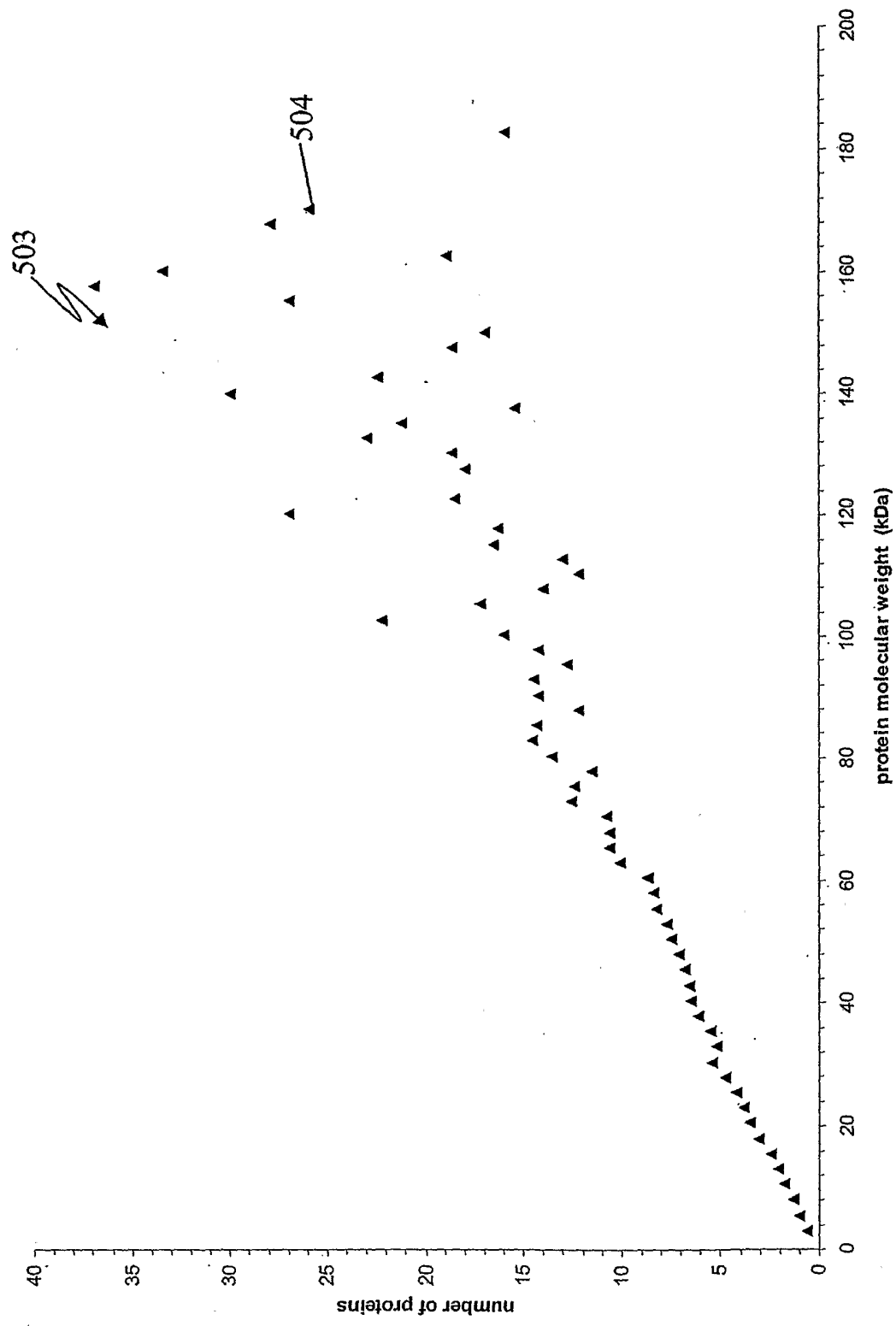


FIG. 5B

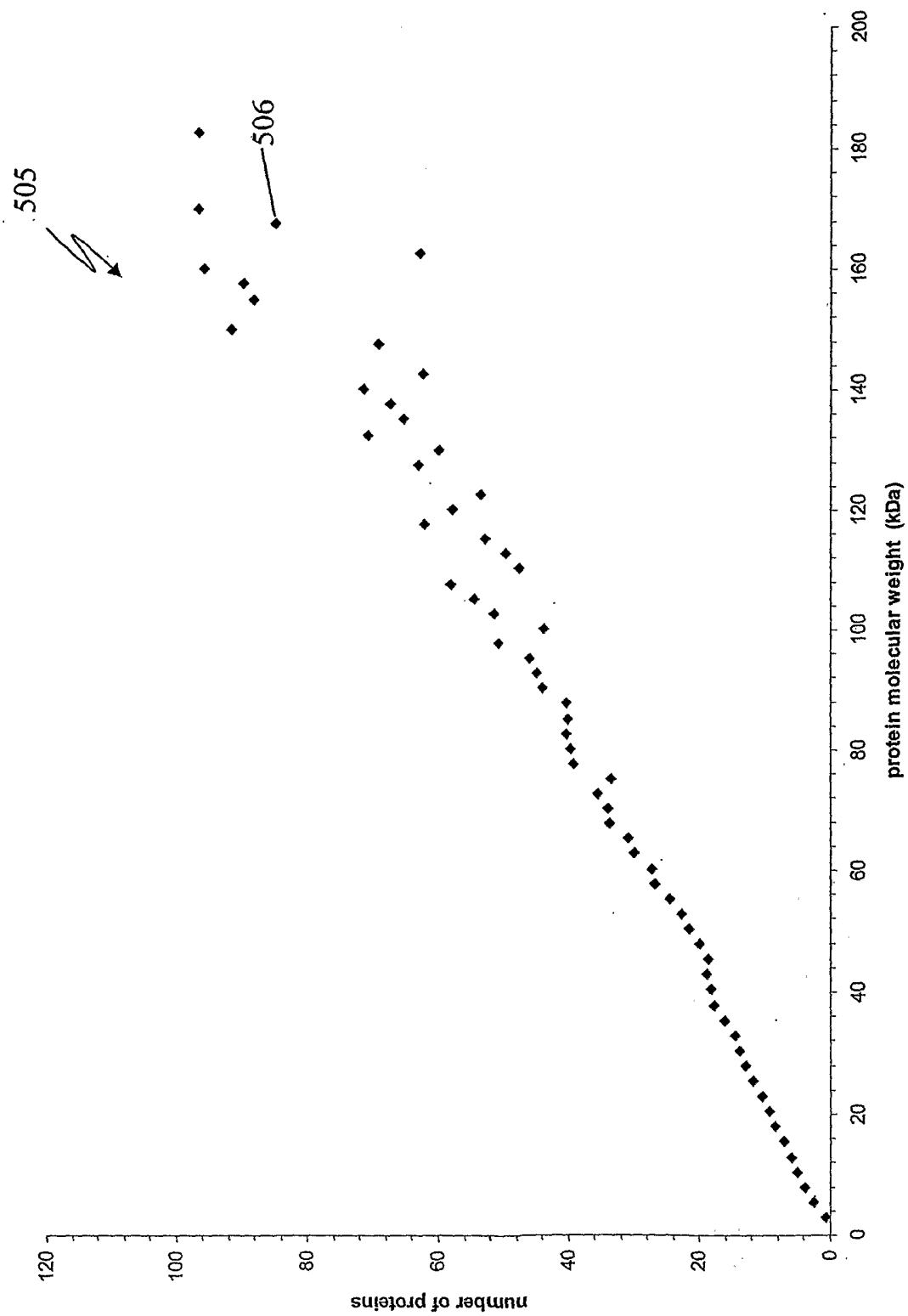


FIG. 5C

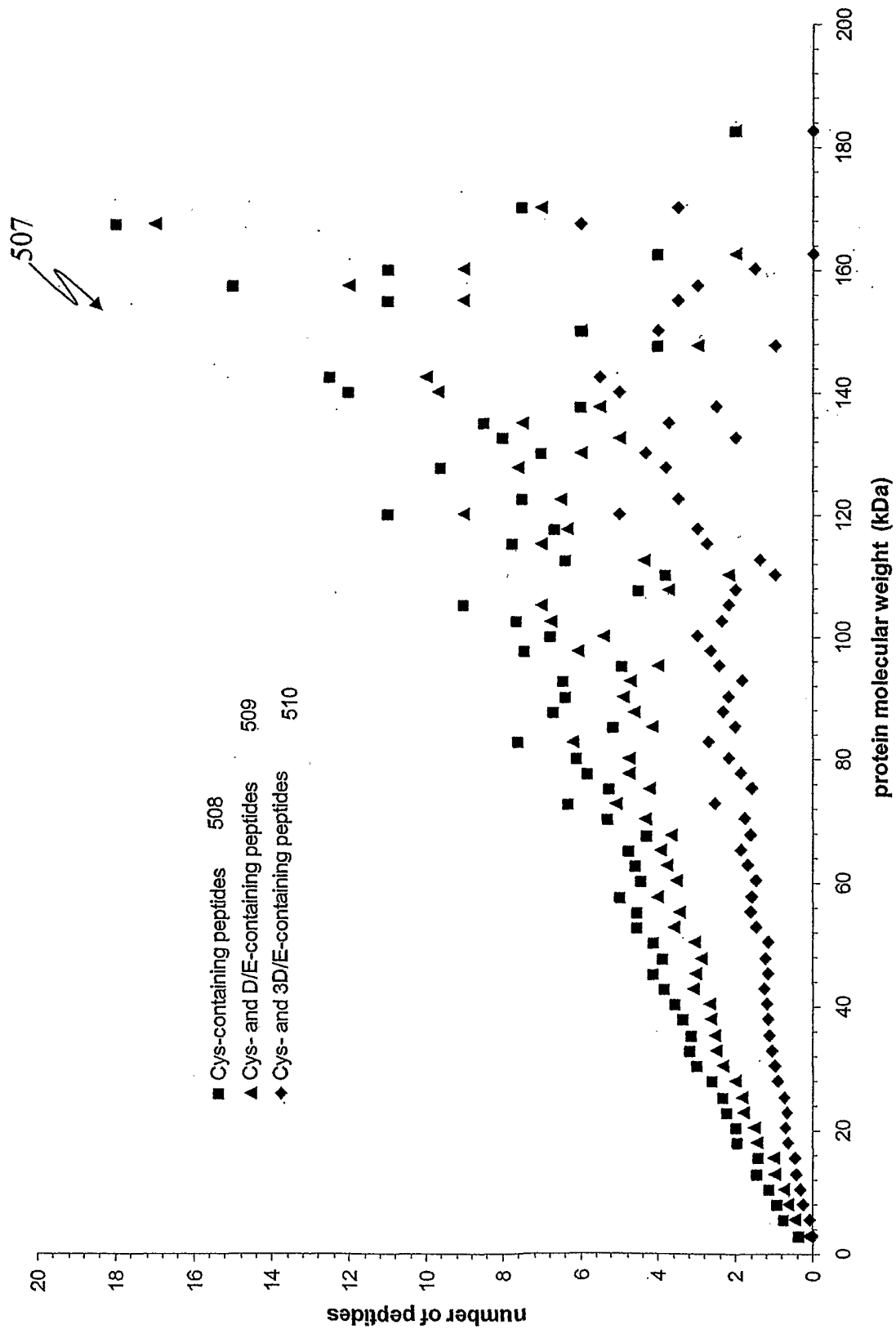


FIG. 5D

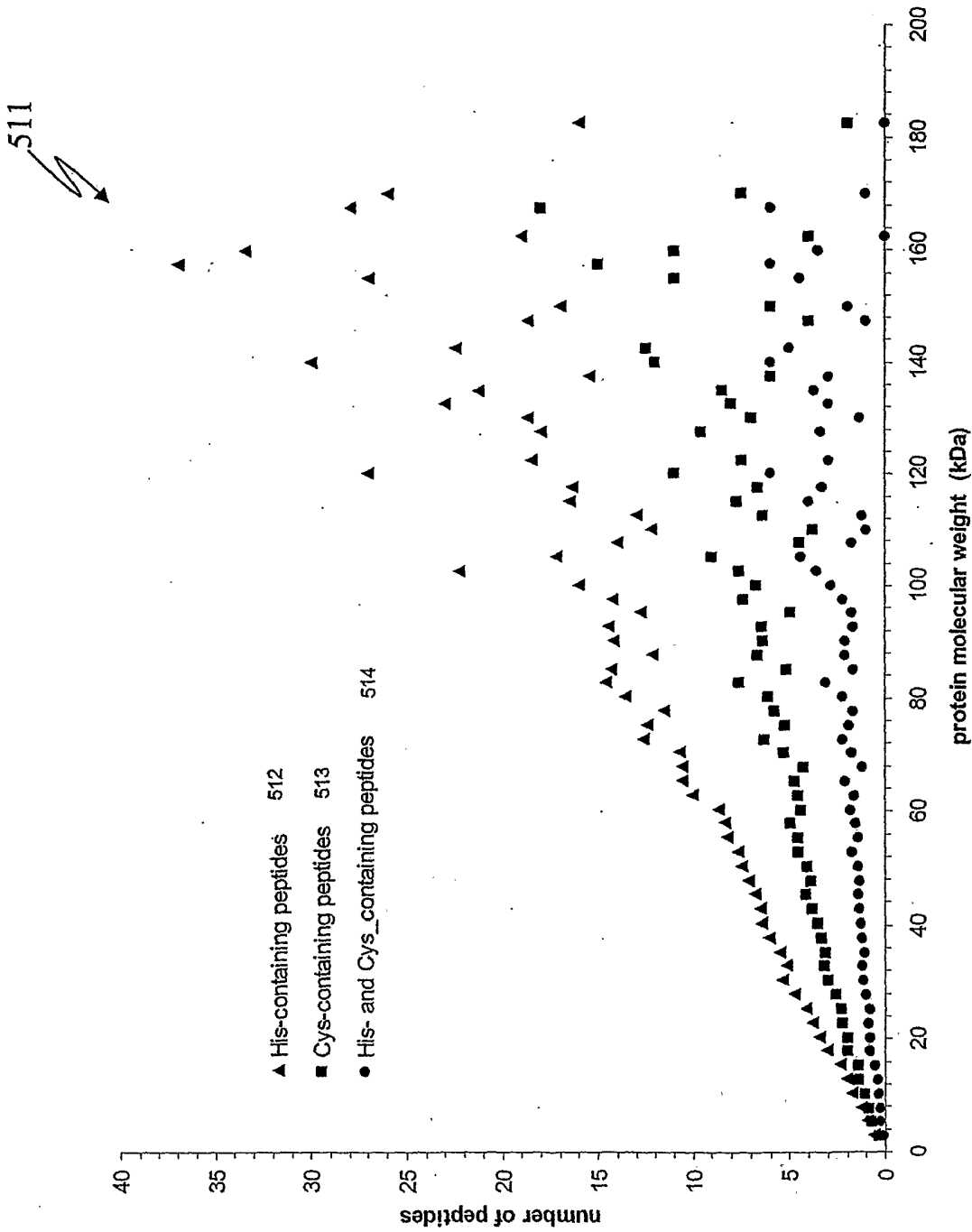


FIG. 5E

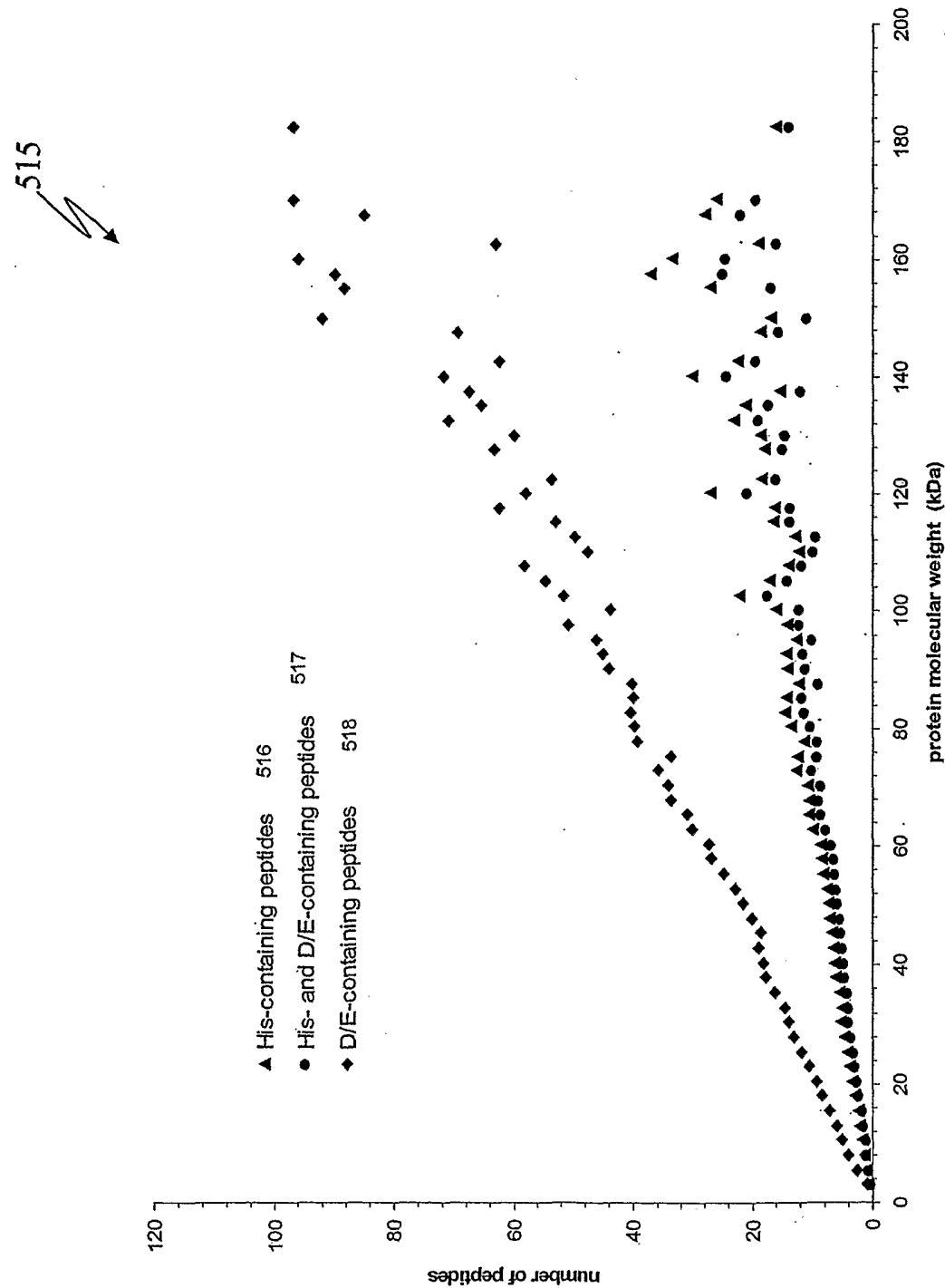


FIG. 5F

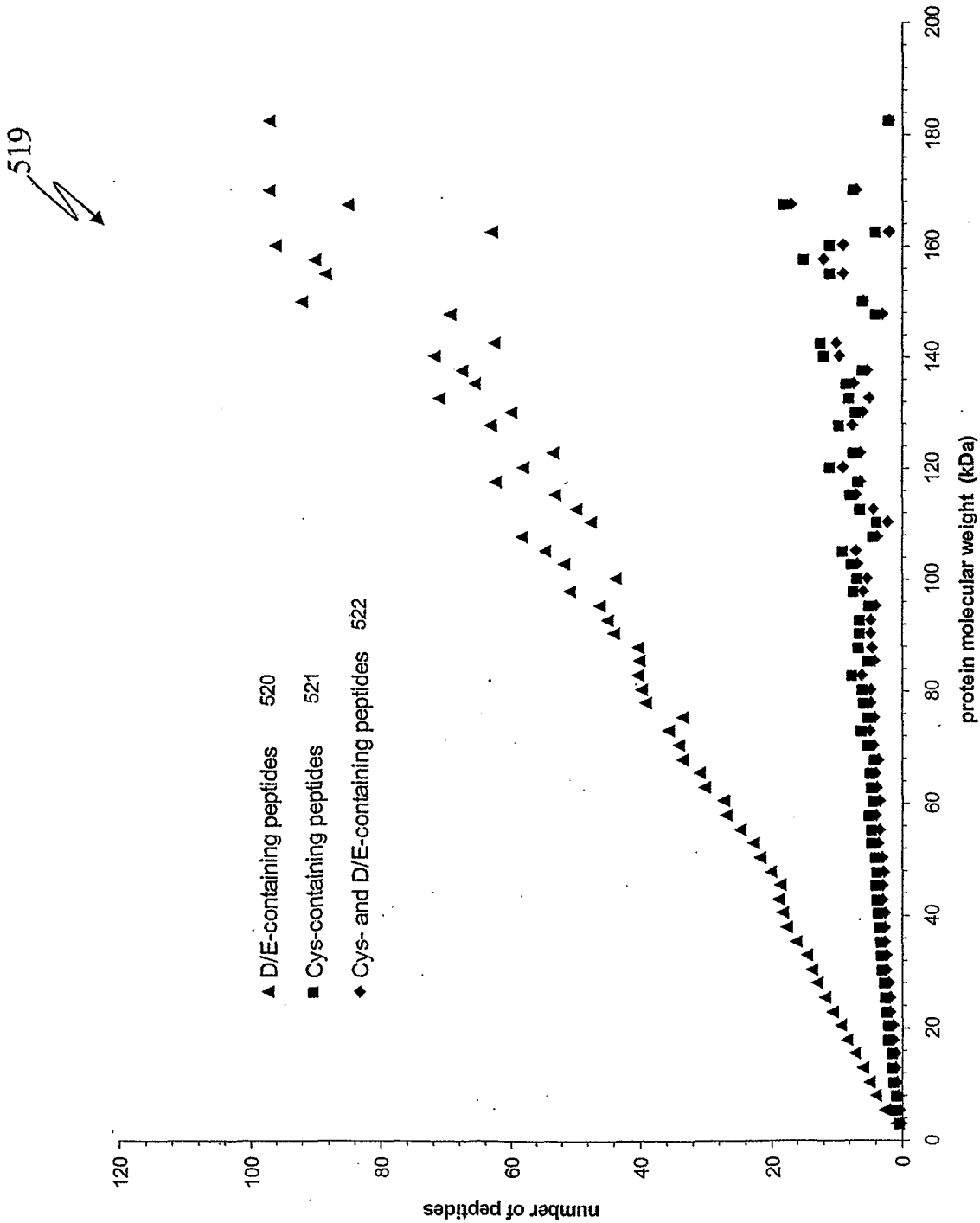


FIG. 5G

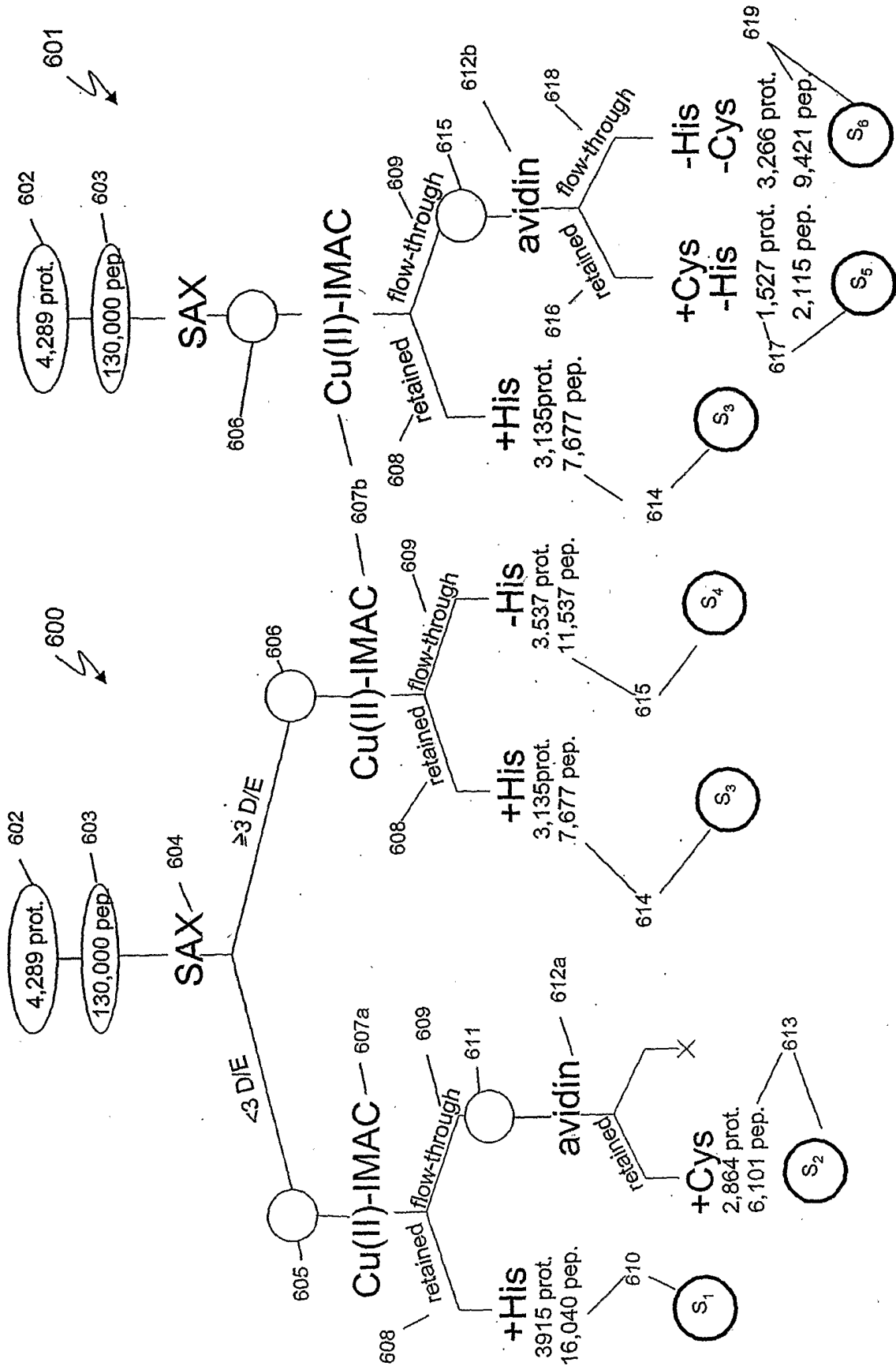


FIG. 6A

FIG. 6B

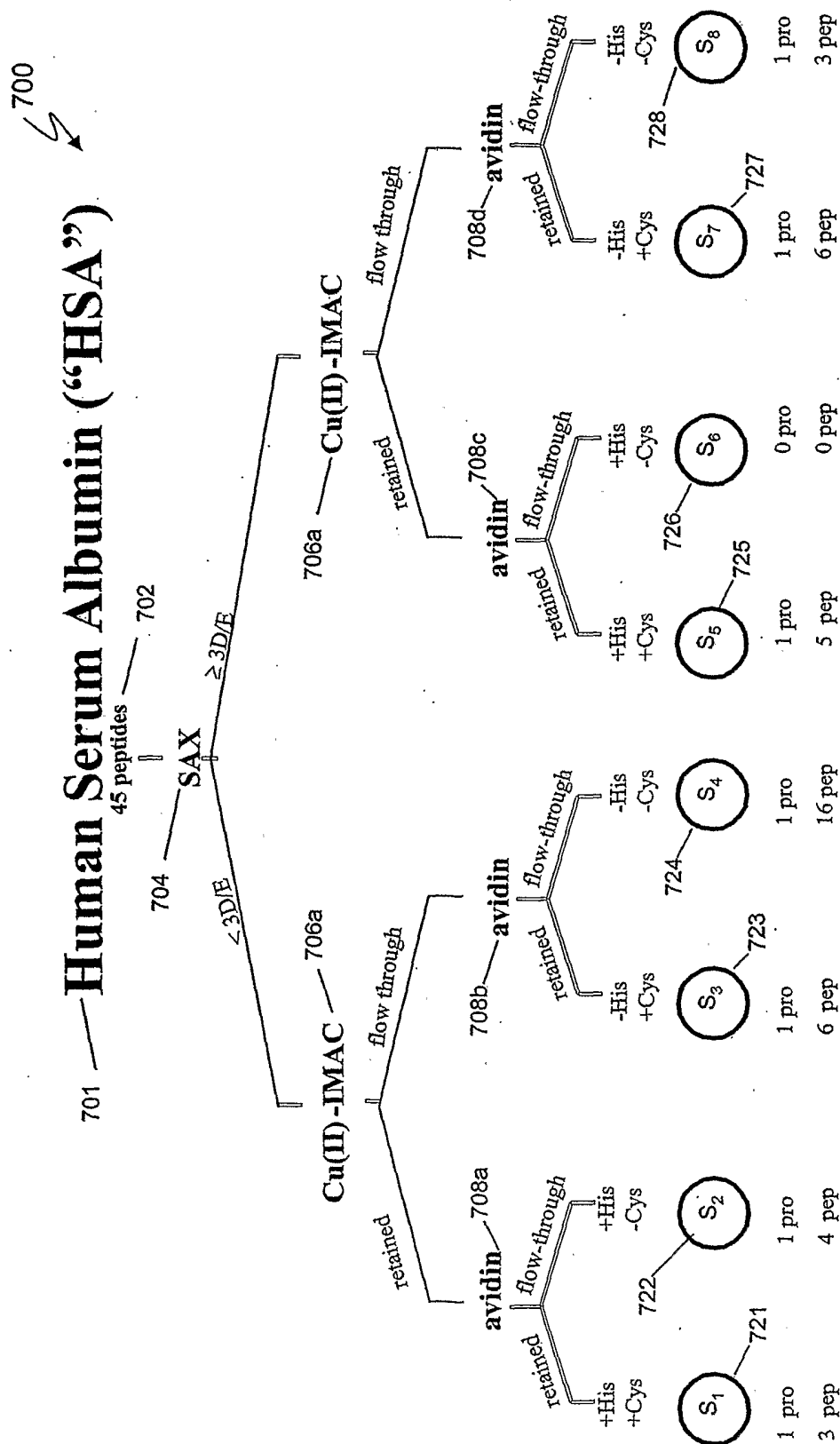


FIG. 7A

750 ↗

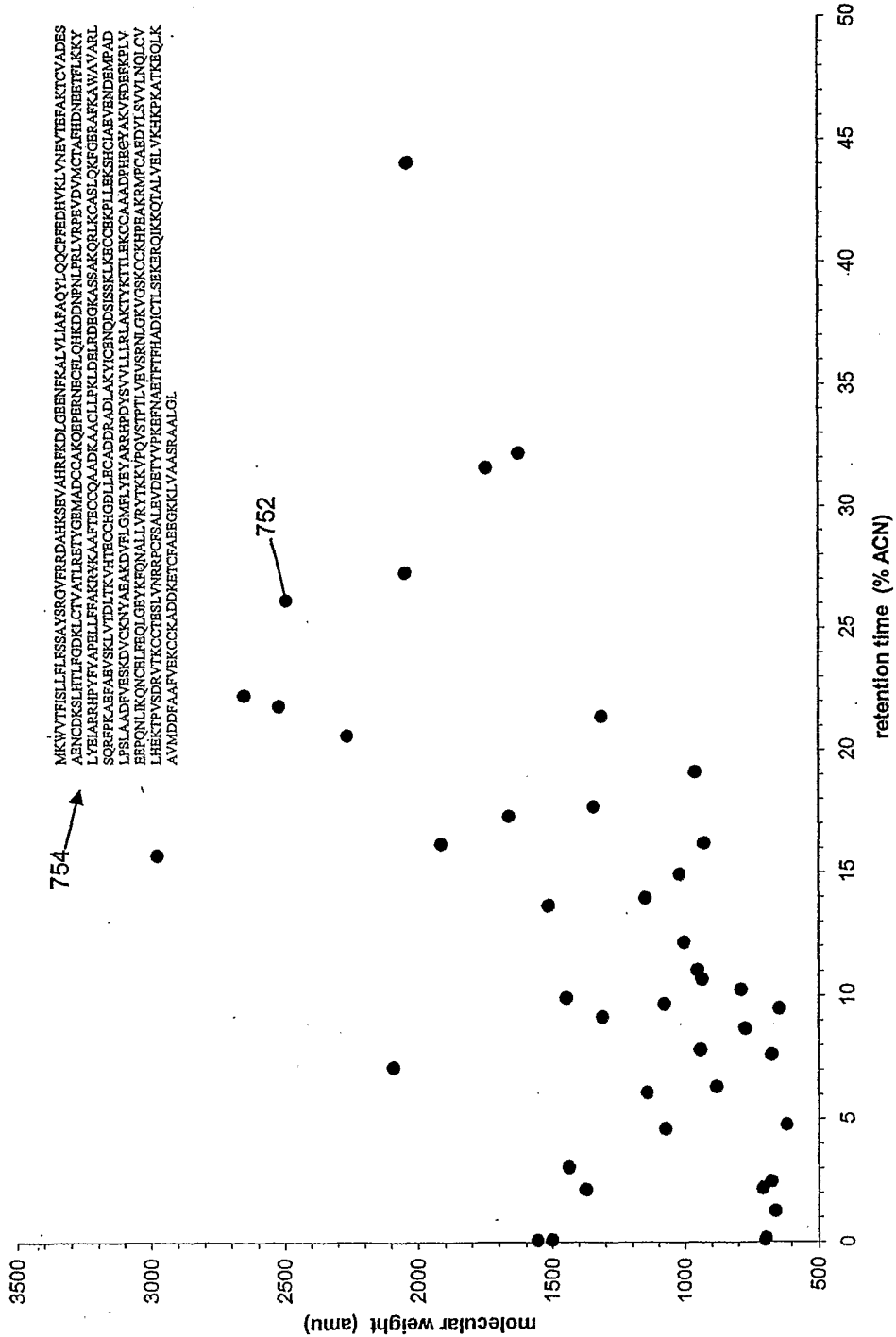
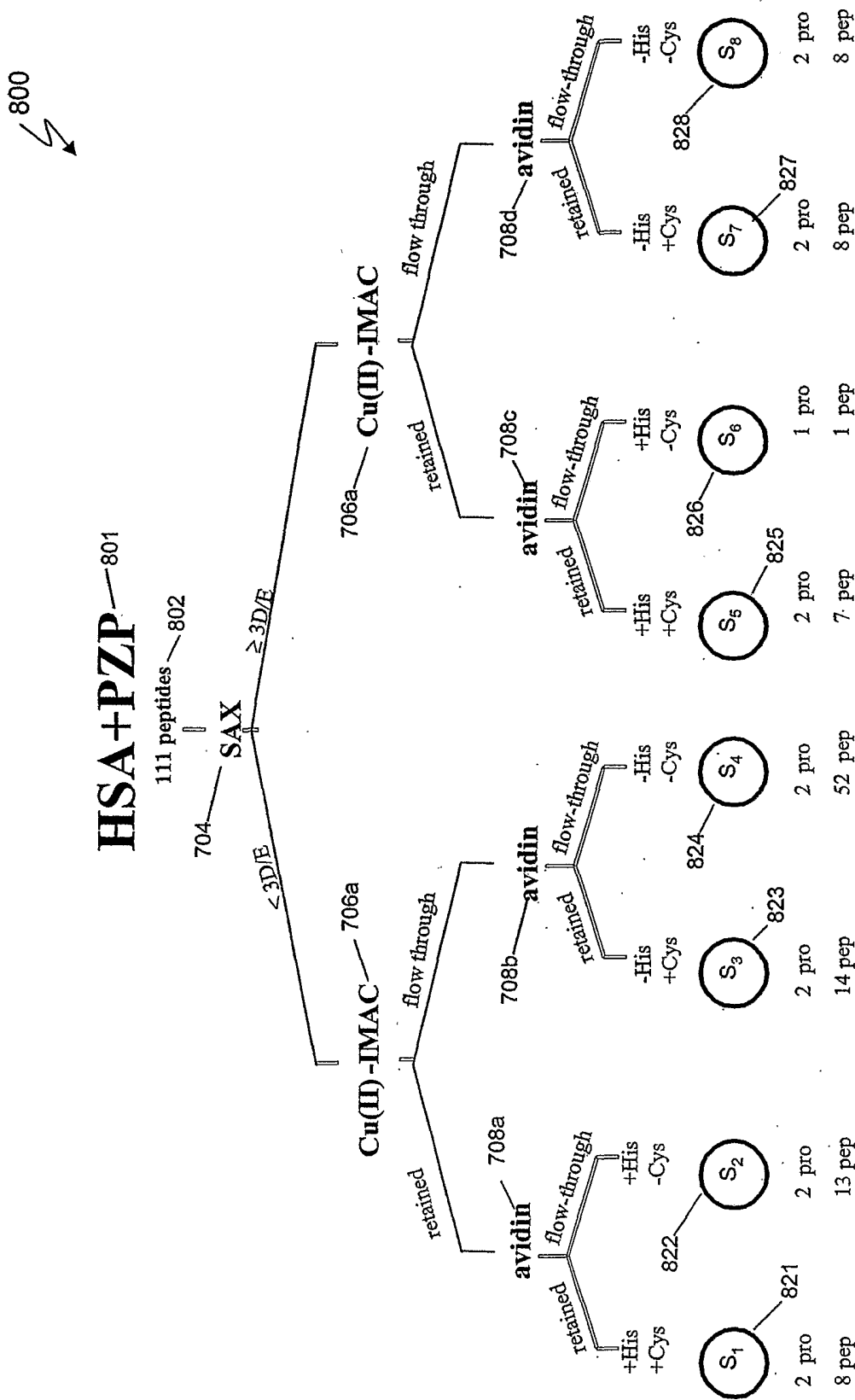
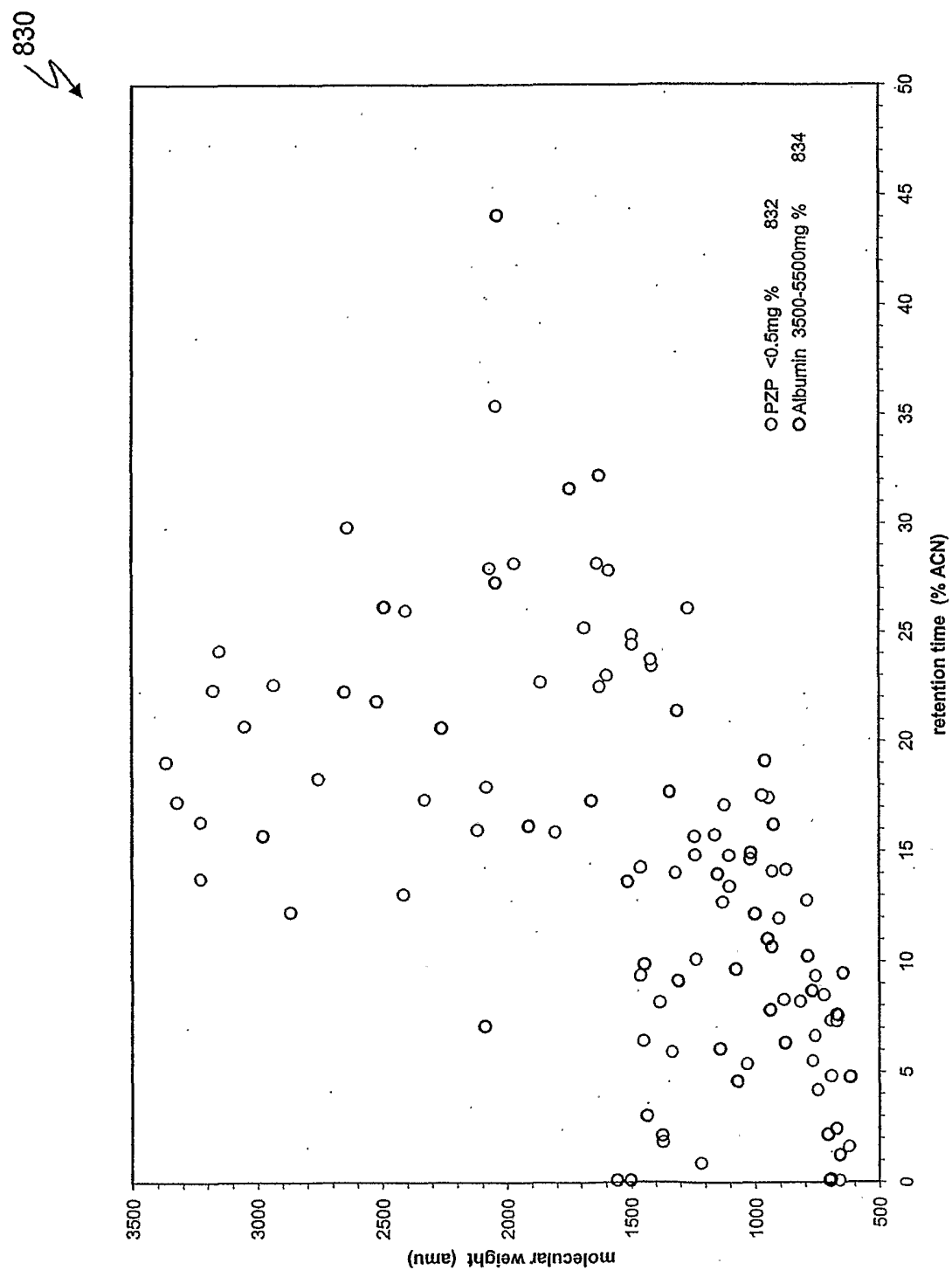


FIG. 7B





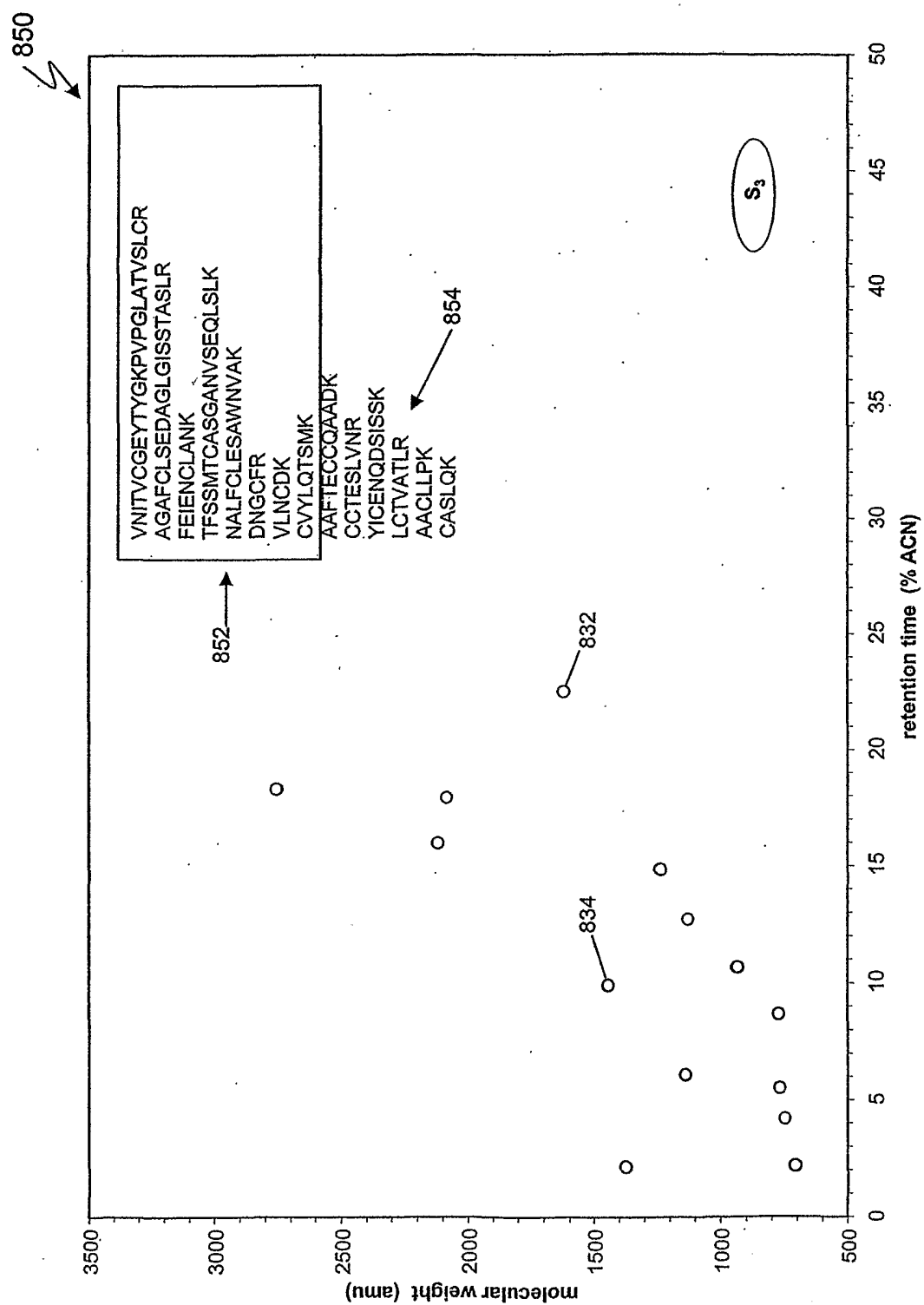


FIG. 8C

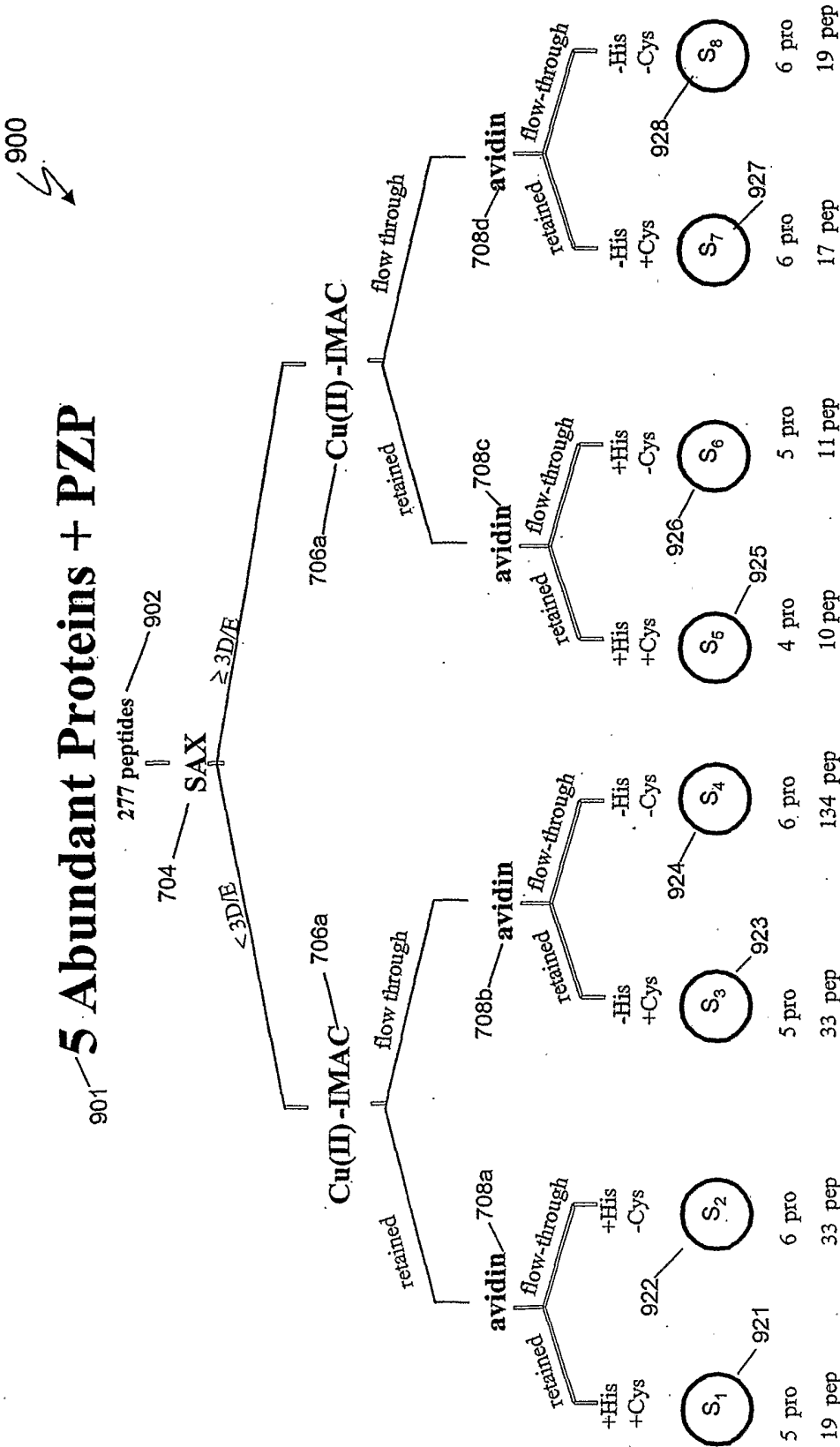


FIG. 9A

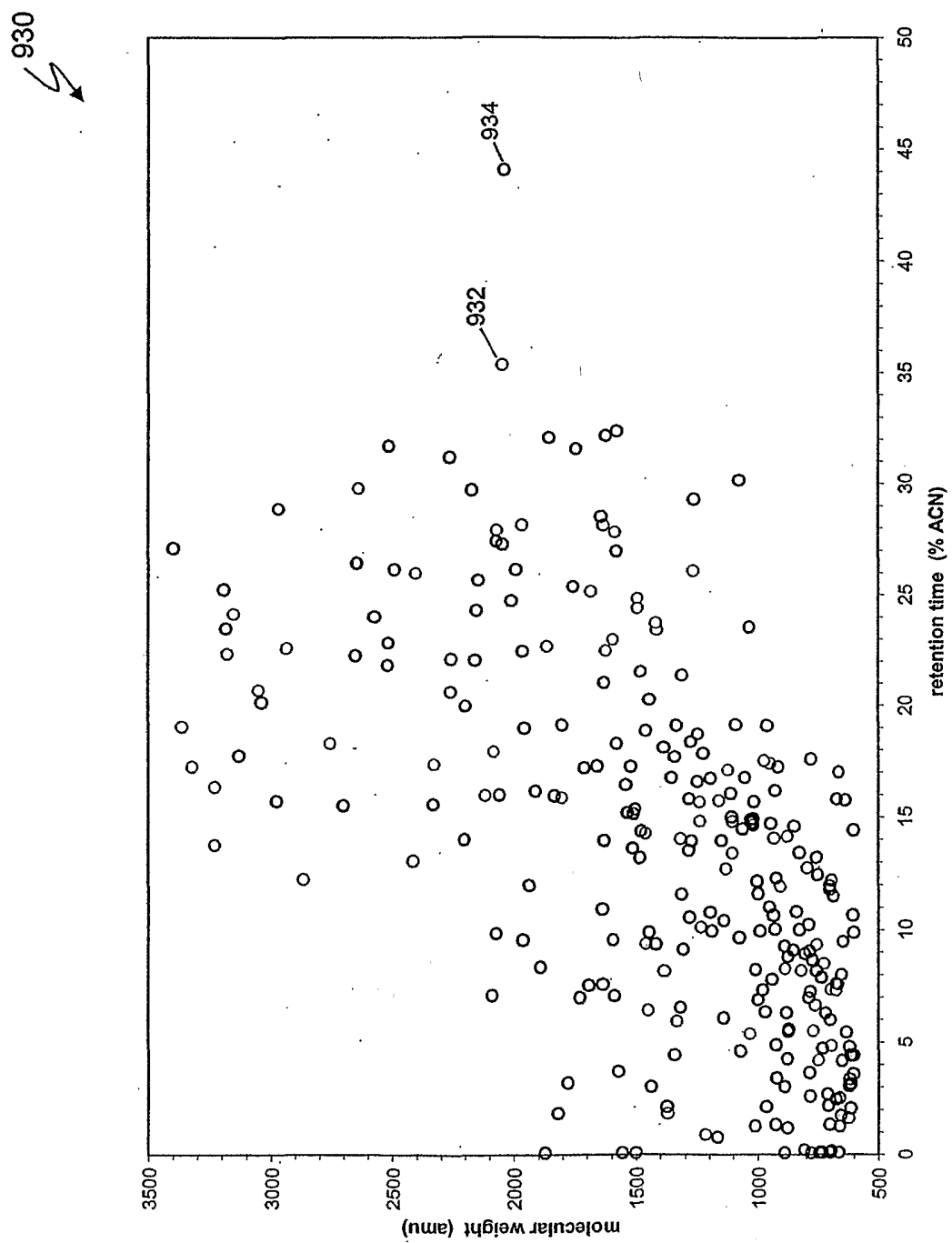


FIG. 9B

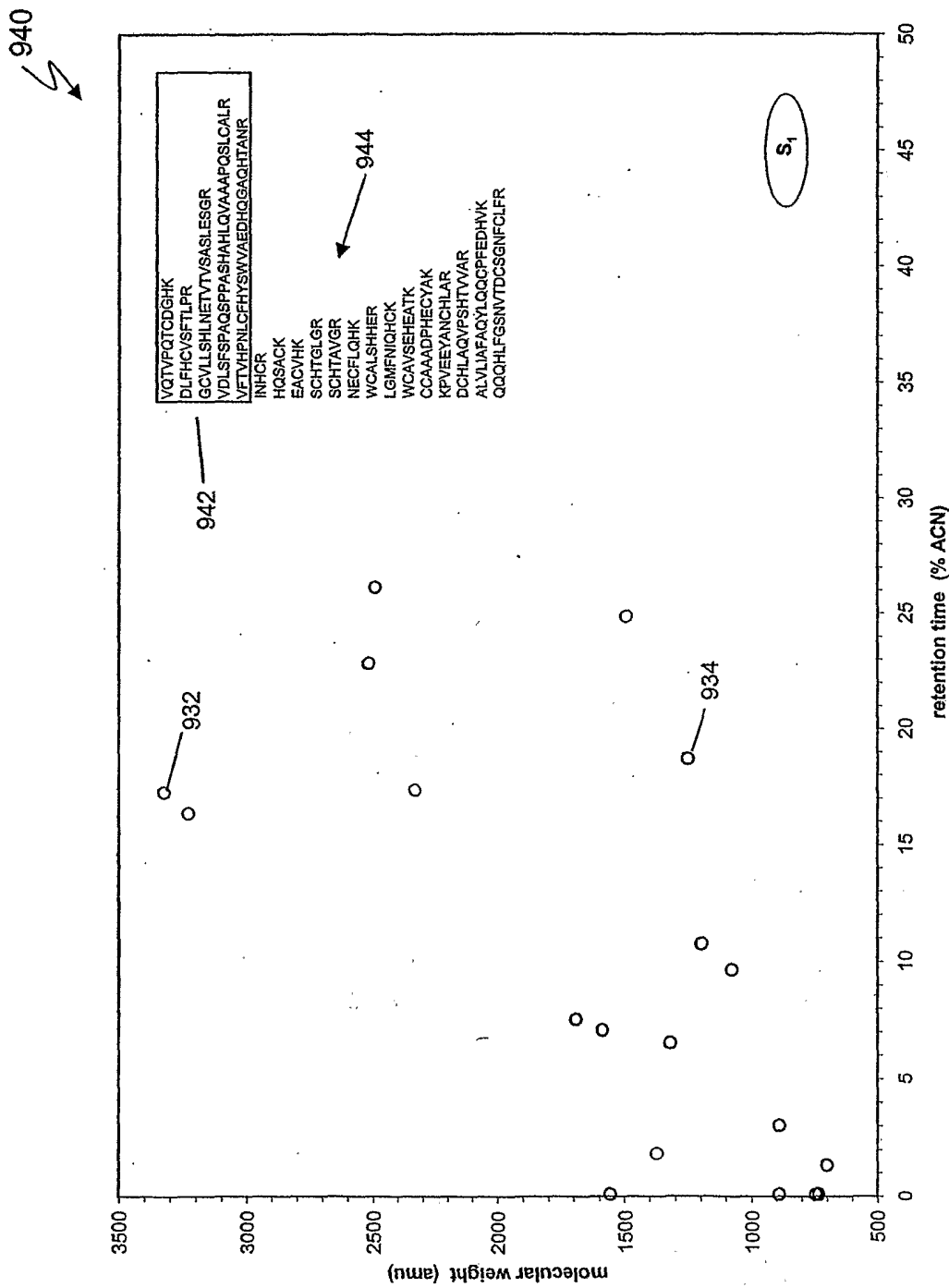


FIG. 9C

945 ↙

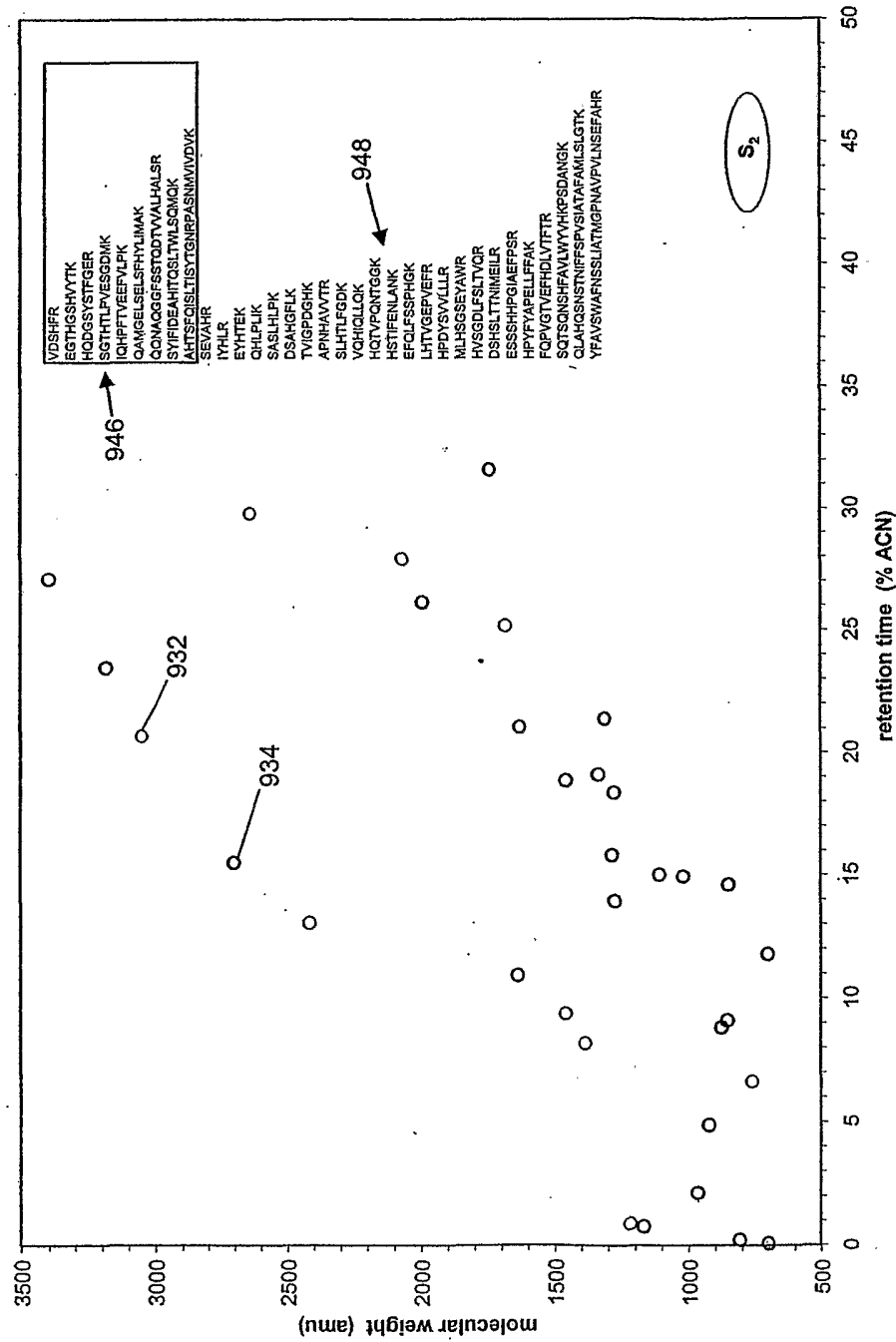


FIG. 9D

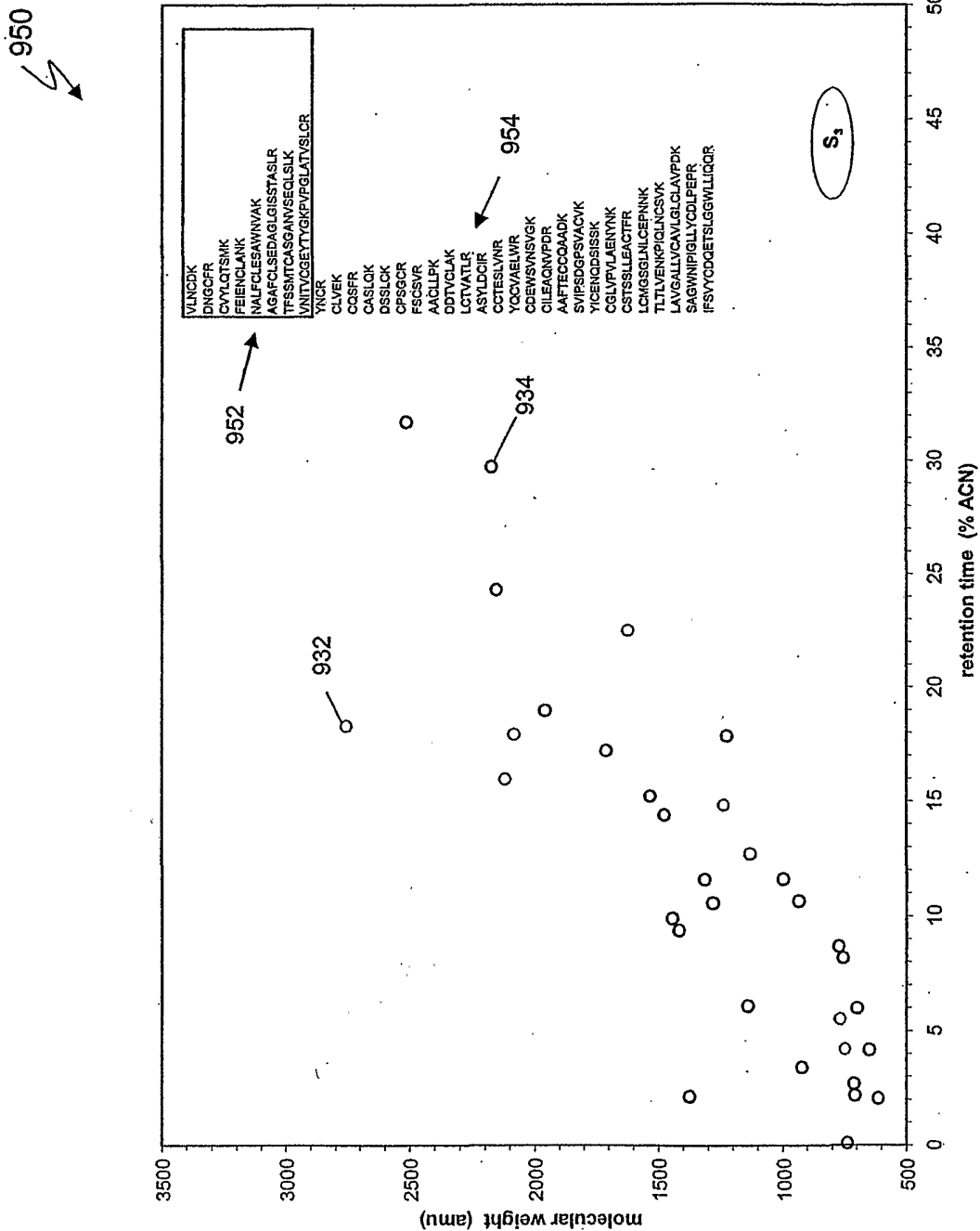


FIG. 9E

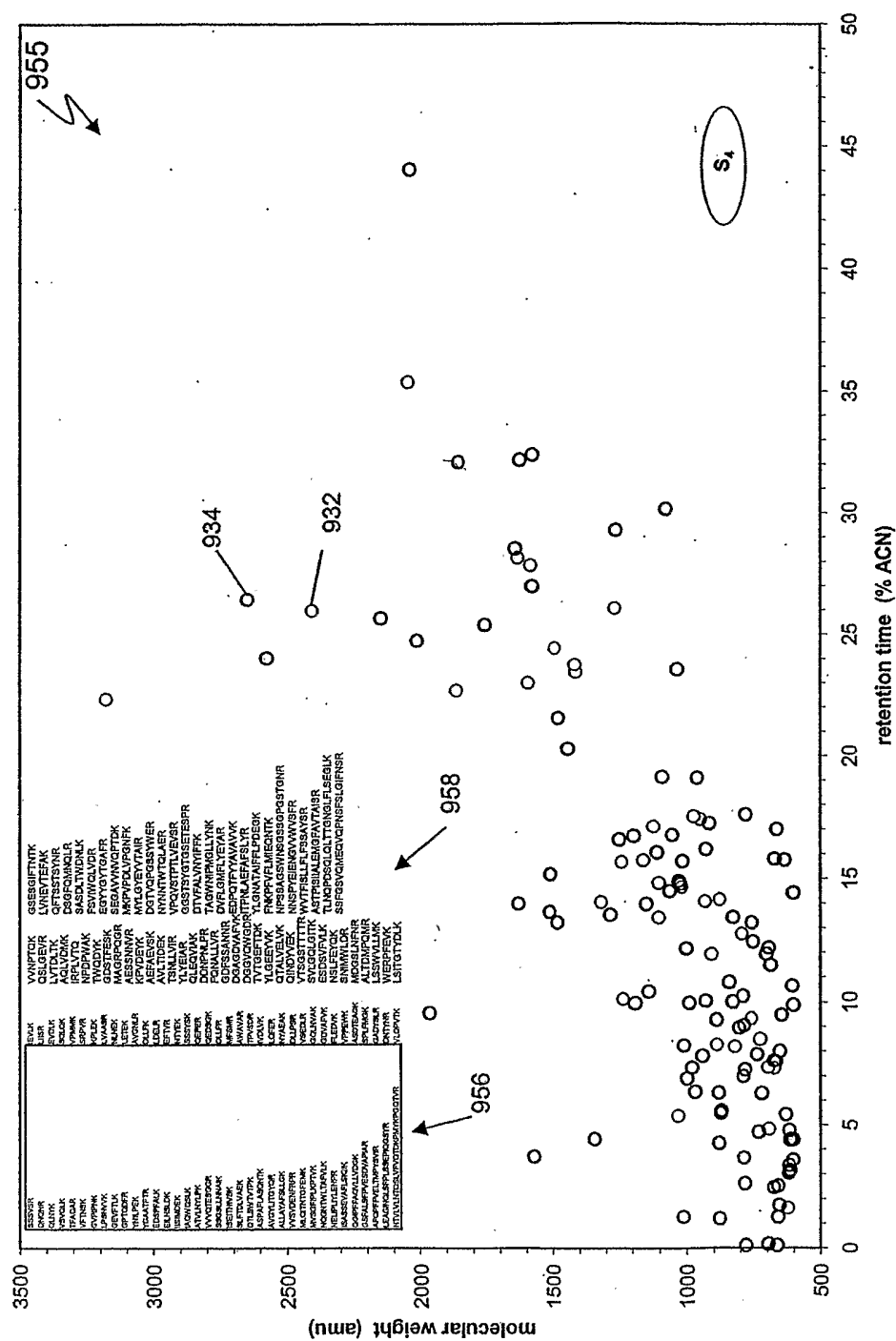


FIG. 9F

960

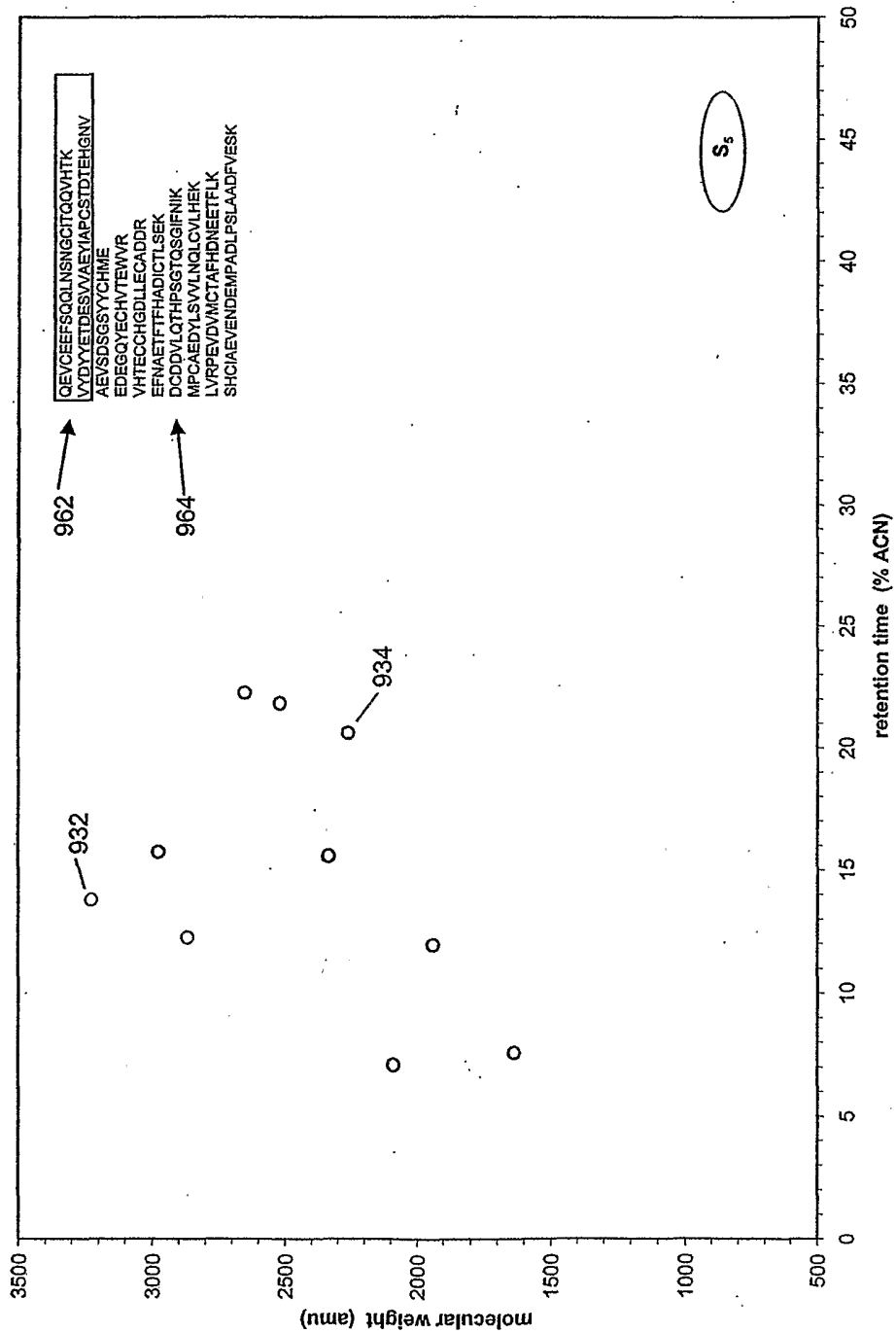


FIG. 9G

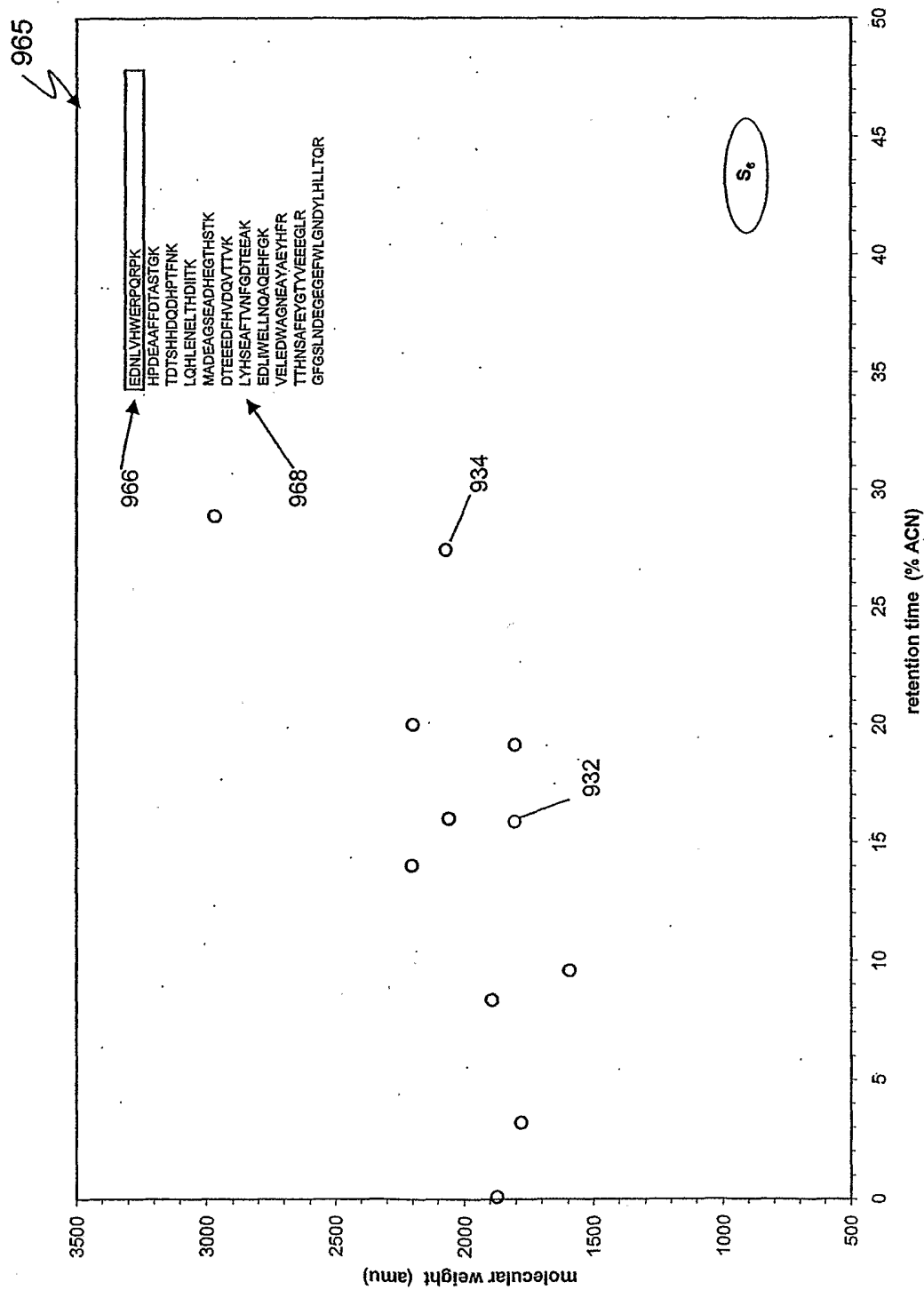


FIG. 9H

970
↙

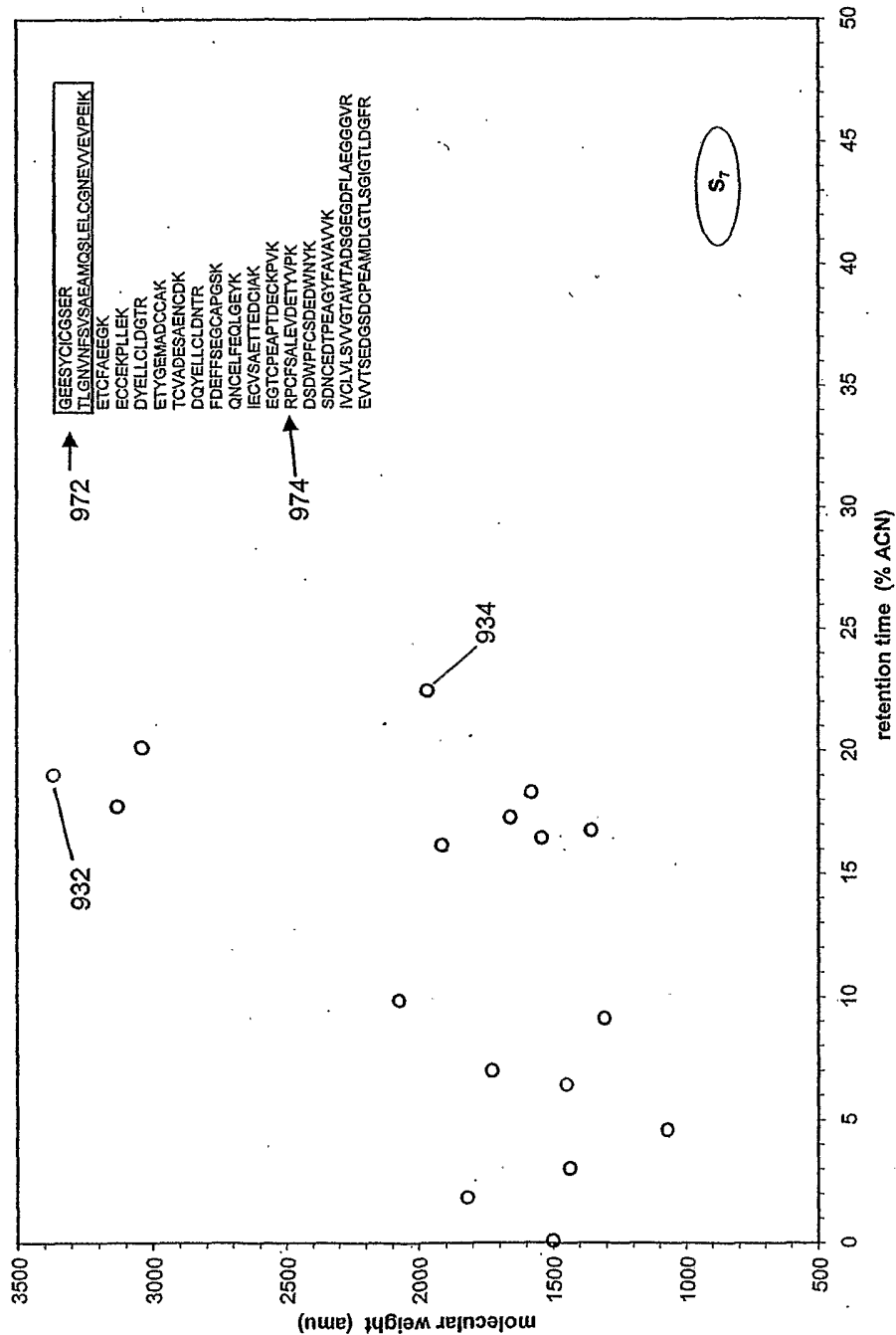


FIG. 9I

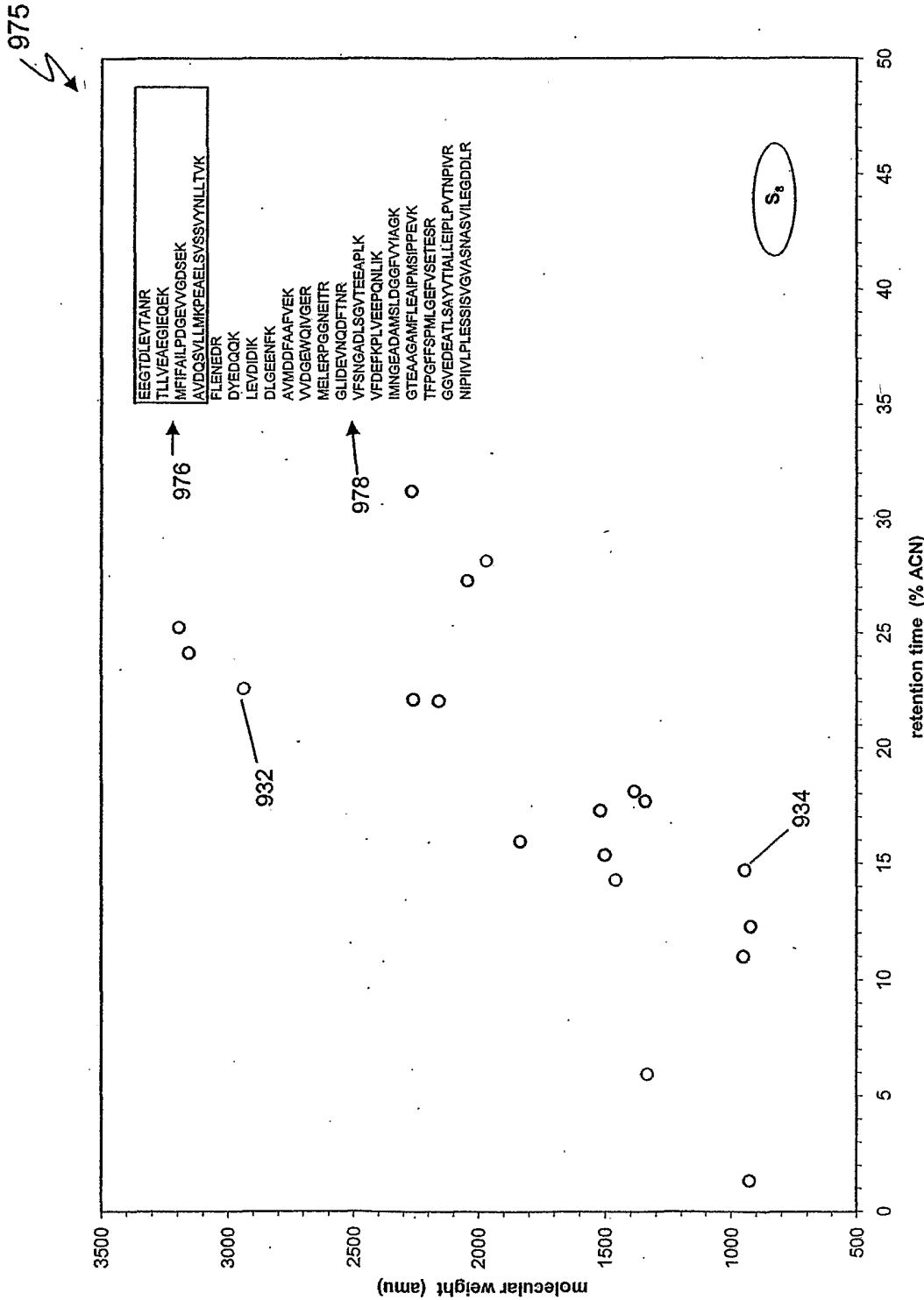


FIG. 9J

SEQUENCE LISTING

<110> Beyond Genomics, Inc. et al.

<120> Digital Chromatography

<130> BG9-012PC

<150> 60/490149

<151> 2003-07-25

<160> 277

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 609

<212> PRT

<213> Homo sapiens

<400> 1

```

Met Lys Trp Val Thr Phe Ile Ser Leu Leu Phe Leu Phe Ser Ser Ala
1      5      10      15
Tyr Ser Arg Gly Val Phe Arg Arg Asp Ala His Lys Ser Glu Val Ala
20     25     30
His Arg Phe Lys Asp Leu Gly Glu Glu Asn Phe Lys Ala Leu Val Leu
35     40     45
Ile Ala Phe Ala Gln Tyr Leu Gln Gln Cys Pro Phe Glu Asp His Val
50     55     60
Lys Leu Val Asn Glu Val Thr Glu Phe Ala Lys Thr Cys Val Ala Asp
65     70     75     80
Glu Ser Ala Glu Asn Cys Asp Lys Ser Leu His Thr Leu Phe Gly Asp
85     90     95
Lys Leu Cys Thr Val Ala Thr Leu Arg Glu Thr Tyr Gly Glu Met Ala
100    105    110
Asp Cys Cys Ala Lys Gln Glu Pro Glu Arg Asn Glu Cys Phe Leu Gln
115    120    125
His Lys Asp Asp Asn Pro Asn Leu Pro Arg Leu Val Arg Pro Glu Val
130    135    140
Asp Val Met Cys Thr Ala Phe His Asp Asn Glu Glu Thr Phe Leu Lys
145    150    155    160
Lys Tyr Leu Tyr Glu Ile Ala Arg Arg His Pro Tyr Phe Tyr Ala Pro
165    170    175
Glu Leu Leu Phe Phe Ala Lys Arg Tyr Lys Ala Ala Phe Thr Glu Cys
180    185    190
Cys Gln Ala Ala Asp Lys Ala Ala Cys Leu Leu Pro Lys Leu Asp Glu
195    200    205
Leu Arg Asp Glu Gly Lys Ala Ser Ser Ala Lys Gln Arg Leu Lys Cys
210    215    220
Ala Ser Leu Gln Lys Phe Gly Glu Arg Ala Phe Lys Ala Trp Ala Val
225    230    235    240
Ala Arg Leu Ser Gln Arg Phe Pro Lys Ala Glu Phe Ala Glu Val Ser
245    250    255
Lys Leu Val Thr Asp Leu Thr Lys Val His Thr Glu Cys Cys His Gly
260    265    270

Asp Leu Leu Glu Cys Ala Asp Asp Arg Ala Asp Leu Ala Lys Tyr Ile
275    280    285
Cys Glu Asn Gln Asp Ser Ile Ser Ser Lys Leu Lys Glu Cys Cys Glu
290    295    300

```

```

Lys Pro Leu Leu Glu Lys Ser His Cys Ile Ala Glu Val Glu Asn Asp
305          310          315          320
Glu Met Pro Ala Asp Leu Pro Ser Leu Ala Ala Asp Phe Val Glu Ser
          325          330          335
Lys Asp Val Cys Lys Asn Tyr Ala Glu Ala Lys Asp Val Phe Leu Gly
          340          345          350
Met Phe Leu Tyr Glu Tyr Ala Arg Arg His Pro Asp Tyr Ser Val Val
          355          360          365
Leu Leu Leu Arg Leu Ala Lys Thr Tyr Lys Thr Thr Leu Glu Lys Cys
370          375          380
Cys Ala Ala Ala Asp Pro His Glu Cys Tyr Ala Lys Val Phe Asp Glu
385          390          395          400
Phe Lys Pro Leu Val Glu Glu Pro Gln Asn Leu Ile Lys Gln Asn Cys
          405          410          415
Glu Leu Phe Glu Gln Leu Gly Glu Tyr Lys Phe Gln Asn Ala Leu Leu
          420          425          430
Val Arg Tyr Thr Lys Lys Val Pro Gln Val Ser Thr Pro Thr Leu Val
          435          440          445
Glu Val Ser Arg Asn Leu Gly Lys Val Gly Ser Lys Cys Cys Lys His
450          455          460

Pro Glu Ala Lys Arg Met Pro Cys Ala Glu Asp Tyr Leu Ser Val Val
465          470          475          480
Leu Asn Gln Leu Cys Val Leu His Glu Lys Thr Pro Val Ser Asp Arg
          485          490          495
Val Thr Lys Cys Cys Thr Glu Ser Leu Val Asn Arg Arg Pro Cys Phe
          500          505          510
Ser Ala Leu Glu Val Asp Glu Thr Tyr Val Pro Lys Glu Phe Asn Ala
          515          520          525
Glu Thr Phe Thr Phe His Ala Asp Ile Cys Thr Leu Ser Glu Lys Glu
          530          535          540
Arg Gln Ile Lys Lys Gln Thr Ala Leu Val Glu Leu Val Lys His Lys
545          550          555          560
Pro Lys Ala Thr Lys Glu Gln Leu Lys Ala Val Met Asp Asp Phe Ala
          565          570          575
Ala Phe Val Glu Lys Cys Cys Lys Ala Asp Asp Lys Glu Thr Cys Phe
          580          585          590
Ala Glu Glu Gly Lys Lys Leu Val Ala Ala Ser Arg Ala Ala Leu Gly
          595          600          605
Leu

```

<210> 2
 <211> 25
 <212> PRT
 <213> Homo sapiens

```

<400> 2
Val Asn Ile Thr Val Cys Gly Glu Tyr Thr Tyr Gly Lys Pro Val Pro
1          5          10          15
Gly Leu Ala Thr Val Ser Leu Cys Arg
          20          25

```

<210> 3
 <211> 21
 <212> PRT
 <213> Homo sapiens

<400> 3
Ala Gly Ala Phe Cys Leu Ser Glu Asp Ala Gly Leu Gly Ile Ser Ser
1 5 10 15
Thr Ala Ser Leu Arg
20

<210> 4
<211> 11
<212> PRT
<213> Homo sapiens

<400> 4
Phe Glu Ile Glu Phe Asn Cys Leu Ala Asn Lys
1 5 10

<210> 5
<211> 20
<212> PRT
<213> Homo sapiens

<400> 5
Thr Phe Ser Ser Met Thr Cys Ala Ser Gly Ala Asn Val Ser Glu Gln
1 5 10 15
Leu Ser Leu Lys
20

<210> 6
<211> 14
<212> PRT
<213> Homo sapiens

<400> 6
Asn Ala Leu Phe Cys Leu Glu Ser Ala Trp Asn Val Ala Lys
1 5 10

<210> 7
<211> 6
<212> PRT
<213> Homo sapiens

<400> 7
Asp Asn Gly Cys Phe Arg
1 5

<210> 8
<211> 6
<212> PRT
<213> Homo sapiens

<400> 8
Val Leu Asn Cys Asp Lys
1 5

<210> 9
<211> 9
<212> PRT
<213> Homo sapiens

<400> 9
Cys Val Tyr Leu Gln Thr Ser Met Lys
1 5

<210> 10
<211> 12
<212> PRT
<213> Homo sapiens

<400> 10
Ala Ala Phe Thr Glu Cys Cys Gln Ala Ala Asp Lys
1 5 10

<210> 11
<211> 9
<212> PRT
<213> Homo sapiens

<400> 11
Cys Cys Thr Glu Ser Leu Val Asn Arg
1 5

<210> 12
<211> 12
<212> PRT
<213> Homo sapiens

<400> 12
Tyr Ile Cys Glu Asn Gln Asp Ser Ile Ser Ser Lys
1 5 10

<210> 13
<211> 8
<212> PRT
<213> Homo sapiens

<400> 13
Leu Cys Thr Val Ala Thr Leu Arg
1 5

<210> 14
<211> 7
<212> PRT
<213> Homo sapiens

<400> 14
Ala Ala Cys Leu Leu Pro Lys
1 5

<210> 15
 <211> 6
 <212> PRT
 <213> Homo sapiens

<400> 15
 Cys Ala Ser Leu Gln Lys
 1 5

<210> 16
 <211> 12
 <212> PRT
 <213> Homo sapiens

<400> 16
 Val Gln Thr Val Pro Gln Thr Cys Asp Gly His Lys
 1 5 10

<210> 17
 <211> 12
 <212> PRT
 <213> Homo sapiens

<400> 17
 Asp Leu Phe His Cys Val Ser Phe Thr Leu Pro Arg
 1 5 10

<210> 18
 <211> 22
 <212> PRT
 <213> Homo sapiens

<400> 18
 Gly Cys Val Leu Leu Ser His Leu Asn Glu Thr Val Thr Val Ser Ala
 1 5 10 15
 Ser Leu Glu Ser Gly Arg
 20

<210> 19
 <211> 31
 <212> PRT
 <213> Homo sapiens

<400> 19
 Val Asp Leu Ser Phe Ser Pro Ala Gln Ser Pro Pro Ala Ser His Ala
 1 5 10 15
 His Leu Gln Val Ala Ala Ala Pro Gln Ser Leu Cys Ala Leu Arg
 20 25 30

<210> 20
 <211> 28
 <212> PRT
 <213> Homo sapiens

<400> 20

Val Phe Thr Val His Pro Asn Leu Cys Phe His Tyr Ser Trp Val Ala
1 5 10 15
Glu Asp His Gln Gly Ala Gln His Thr Ala Asn Arg
20 25

<210> 21

<211> 5

<212> PRT

<213> Homo sapiens

<400> 21

Ile Asn His Cys Arg
1 5

<210> 22

<211> 6

<212> PRT

<213> Homo sapiens

<400> 22

His Gln Ser Ala Cys Lys
1 5

<210> 23

<211> 6

<212> PRT

<213> Homo sapiens

<400> 23

Glu Ala Cys Val His Lys
1 5

<210> 24

<211> 8

<212> PRT

<213> Homo sapiens

<400> 24

Ser Cys His Thr Gly Leu Gly Arg
1 5

<210> 25

<211> 8

<212> PRT

<213> Homo sapiens

<400> 25

Ser Cys His Thr Ala Val Gly Arg
1 5

<210> 26

<211> 8

<212> PRT

<213> Homo sapiens

<400> 26

Asn Glu Cys Phe Leu Gln His Lys

1

5

<210> 27

<211> 9

<212> PRT

<213> Homo sapiens

<400> 27

Trp Cys Ala Leu Ser His His Glu Arg

1

5

<210> 28

<211> 10

<212> PRT

<213> Homo sapiens

<400> 28

Leu Gly Met Phe Asn Ile Gln His Cys Lys

1

5

10

<210> 29

<211> 11

<212> PRT

<213> Homo sapiens

<400> 29

Trp Cys Ala Val Ser Glu His Glu Ala Thr Lys

1

5

10

<210> 30

<211> 13

<212> PRT

<213> Homo sapiens

<400> 30

Cys Cys Ala Ala Ala Asp Pro His Glu Cys Tyr Ala Lys

1

5

10

<210> 31

<211> 13

<212> PRT

<213> Homo sapiens

<400> 31

Lys Pro Val Glu Glu Tyr Ala Asn Cys His Leu Ala Arg

1

5

10

<210> 32

<211> 15

<212> PRT

<213> Homo sapiens

<400> 32

Asp Cys His Leu Ala Gln Val Pro Ser His Thr Val Val Ala Arg
1 5 10 15

<210> 33

<211> 21

<212> PRT

<213> Homo sapiens

<400> 33

Ala Leu Val Leu Ile Ala Phe Ala Gln Tyr Leu Gln Gln Cys Pro Phe
1 5 10 15
Glu Asp His Val Lys
20

<210> 34

<211> 21

<212> PRT

<213> Homo sapiens

<400> 34

Gln Gln Gln His Leu Phe Gly Ser Asn Val Thr Asp Cys Ser Gly Asn
1 5 10 15
Phe Cys Leu Phe Arg
20

<210> 35

<211> 6

<212> PRT

<213> Homo sapiens

<400> 35

Val Asp Ser His Phe Arg
1 5

<210> 36

<211> 11

<212> PRT

<213> Homo sapiens

<400> 36

Glu Gly Thr His Gly Ser His Val Tyr Thr Lys
1 5 10

<210> 37

<211> 12

<212> PRT

<213> Homo sapiens

<400> 37

His Gln Asp Gly Ser Tyr Ser Thr Phe Gly Glu Arg
1 5 10

<210> 38
 <211> 14
 <212> PRT
 <213> Homo sapiens

<400> 38
 Ser Gly Thr His Thr Leu Pro Val Glu Ser Gly Asp Met Lys
 1 5 10

<210> 39
 <211> 14
 <212> PRT
 <213> Homo sapiens

<400> 39
 Ile Gln His Pro Phe Thr Val Glu Glu Phe Val Leu Pro Lys
 1 5 10

<210> 40
 <211> 18
 <212> PRT
 <213> Homo sapiens

<400> 40
 Gln Ala Met Gly Glu Leu Ser Glu Leu Ser Phe His Tyr Leu Ile Met
 1 5 10 15
 Ala Lys

<210> 41
 <211> 23
 <212> PRT
 <213> Homo sapiens

<400> 41
 Gln Gln Asn Ala Gln Gly Gly Phe Ser Ser Thr Gln Asp Thr Val Val
 1 5 10 15
 Ala Leu His Ala Leu Ser Arg
 20

<210> 42
 <211> 22
 <212> PRT
 <213> Homo sapiens

<400> 42
 Ser Tyr Ile Phe Ile Asp Glu Ala His Ile Thr Gln Ser Leu Thr Trp
 1 5 10 15
 Leu Ser Gln Met Gln Lys
 20

<210> 43
 <211> 28
 <212> PRT
 <213> Homo sapiens

<400> 43

Ala His Thr Ser Phe Gln Ile Ser Leu Thr Ile Ser Tyr Thr Gly Asn
1 5 10 15
Arg Pro Ala Ser Asn Met Val Ile Val Asp Val Lys
20 25

<210> 44

<211> 6

<212> PRT

<213> Homo sapiens

<400> 44

Ser Glu Val Ala His Arg
1 5

<210> 45

<211> 5

<212> PRT

<213> Homo sapiens

<400> 45

Ile Tyr His Leu Arg
1 5

<210> 46

<211> 6

<212> PRT

<213> Homo sapiens

<400> 46

Glu Tyr His Thr Glu Lys
1 5

<210> 47

<211> 7

<212> PRT

<213> Homo sapiens

<400> 47

Gln His Leu Pro Leu Ile Lys
1 5

<210> 48

<211> 8

<212> PRT

<213> Homo sapiens

<400> 48

Ser Ala Ser Leu His Leu Pro Lys
1 5

<210> 49

<211> 8

<212> PRT

<213> Homo sapiens

<400> 49

Asp Ser Ala His Gly Phe Leu Lys
1 5

<210> 50

<211> 9

<212> PRT

<213> Homo sapiens

<400> 50

Thr Val Ile Gly Pro Asp Gly His Lys
1 5

<210> 51

<211> 9

<212> PRT

<213> Homo sapiens

<400> 51

Ala Pro Asn His Ala Val Val Thr Arg
1 5

<210> 52

<211> 9

<212> PRT

<213> Homo sapiens

<400> 52

Ser Leu His Thr Leu Phe Gly Asp Lys
1 5

<210> 53

<211> 9

<212> PRT

<213> Homo sapiens

<400> 53

Val Gln His Ile Gln Leu Leu Gln Lys
1 5

<210> 54

<211> 11

<212> PRT

<213> Homo sapiens

<400> 54

His Gln Thr Val Pro Gln Asn Thr Gly Gly Lys
1 5 10

<210> 55

<211> 11

<212> PRT

<213> Homo sapiens

<400> 55
His Ser Thr Ile Phe Glu Asn Leu Ala Asn Lys
1 5 10

<210> 56
<211> 11
<212> PRT
<213> Homo sapiens

<400> 56
Glu Phe Gln Leu Phe Ser Ser Pro His Gly Lys
1 5 10

<210> 57
<211> 11
<212> PRT
<213> Homo sapiens

<400> 57
Leu His Thr Val Gly Glu Pro Val Glu Phe Arg
1 5 10

<210> 58
<211> 11
<212> PRT
<213> Homo sapiens

<400> 58
His Pro Asp Tyr Ser Val Val Leu Leu Leu Arg
1 5 10

<210> 59
<211> 11
<212> PRT
<213> Homo sapiens

<400> 59
Met Leu His Ser Gly Ser Glu Tyr Ala Trp Arg
1 5 10

<210> 60
<211> 13
<212> PRT
<213> Homo sapiens

<400> 60
His Val Ser Gly Asp Leu Phe Ser Leu Thr Val Gln Arg
1 5 10

<210> 61
<211> 13
<212> PRT
<213> Homo sapiens

<400> 61

Asp Ser His Ser Leu Thr Thr Asn Ile Met Glu Ile Leu
 1 5 10

<210> 62

<211> 15

<212> PRT

<213> Homo sapiens

<400> 62

Glu Ser Ser Ser His His Pro Gly Ile Ala Glu Phe Pro Ser Arg
 1 5 10 15

<210> 63

<211> 14

<212> PRT

<213> Homo sapiens

<400> 63

His Pro Tyr Phe Tyr Ala Pro Glu Leu Leu Phe Phe Ala Lys
 1 5 10

<210> 64

<211> 16

<212> PRT

<213> Homo sapiens

<400> 64

Phe Gln Pro Val Gly Thr Val Glu Phe His Asp Leu Val Thr Phe Thr
 1 5 10 15

<210> 65

<211> 24

<212> PRT

<213> Homo sapiens

<400> 65

Ser Gln Thr Ser Gln Asn Ser His Phe Ala Val Leu Trp Tyr Val His
 1 5 10 15
 Lys Pro Ser Asp Ala Asn Gly Lys
 20

<210> 66

<211> 29

<212> PRT

<213> Homo sapiens

<400> 66

Gln Leu Ala His Gln Ser Asn Ser Thr Asn Ile Phe Phe Ser Pro Val
 1 5 10 15
 Ser Ile Ala Thr Ala Phe Ala Met Leu Ser Leu Gly Thr
 20 25

<210> 67
<211> 31
<212> PRT
<213> Homo sapiens

<400> 67
Tyr Phe Ala Val Ser Trp Ala Phe Asn Ser Ser Leu Ile Ala Thr Met
1 5 10 15
Gly Pro Asn Ala Val Pro Val Leu Asn Ser Glu Phe Ala His Arg
20 25 30

<210> 68
<211> 4
<212> PRT
<213> Homo sapiens

<400> 68
Tyr Asn Cys Arg
1

<210> 69
<211> 5
<212> PRT
<213> Homo sapiens

<400> 69
Cys Leu Val Glu Lys
1 5

<210> 70
<211> 5
<212> PRT
<213> Homo sapiens

<400> 70
Cys Gln Ser Phe Arg
1 5

<210> 71
<211> 6
<212> PRT
<213> Homo sapiens

<400> 71
Asp Ser Ser Leu Cys Lys
1 5

<210> 72
<211> 6
<212> PRT
<213> Homo sapiens

<400> 72
Cys Pro Ser Gly Cys Arg
1 5

<210> 73
<211> 6
<212> PRT
<213> Homo sapiens

<400> 73
Phe Ser Cys Ser Val Arg
1 5

<210> 74
<211> 8
<212> PRT
<213> Homo sapiens

<400> 74
Asp Asp Thr Val Cys Leu Ala Lys
1 5

<210> 75
<211> 8
<212> PRT
<213> Homo sapiens

<400> 75
Ala Ser Tyr Leu Asp Cys Ile Arg
1 5

<210> 76
<211> 9
<212> PRT
<213> Homo sapiens

<400> 76
Tyr Gln Cys Val Ala Glu Leu Trp Arg
1 5

<210> 77
<211> 11
<212> PRT
<213> Homo sapiens

<400> 77
Cys Asp Glu Trp Ser Val Asn Ser Val Gly Lys
1 5 10

<210> 78
<211> 11
<212> PRT
<213> Homo sapiens

<400> 78
Cys Ile Leu Glu Ala Gln Asn Val Pro Asp Arg
1 5 10

<210> 79
<211> 14
<212> PRT
<213> Homo sapiens

<400> 79
Ser Val Ile Pro Ser Asp Gly Pro Ser Val Ala Cys Val Lys
1 5 10

<210> 80
<211> 13
<212> PRT
<213> Homo sapiens

<400> 80
Cys Gly Leu Val Pro Val Leu Ala Glu Asn Tyr Asn Lys
1 5 10

<210> 81
<211> 13
<212> PRT
<213> Homo sapiens

<400> 81
Cys Ser Thr Ser Ser Leu Leu Glu Ala Cys Thr Phe Arg
1 5 10

<210> 82
<211> 15
<212> PRT
<213> Homo sapiens

<400> 82
Leu Cys Met Gly Ser Gly Leu Asn Leu Cys Glu Pro Asn Asn Lys
1 5 10 15

<210> 83
<211> 17
<212> PRT
<213> Homo sapiens

<400> 83
Thr Leu Thr Leu Val Glu Asn Lys Pro Ile Gln Leu Asn Cys Ser Val
1 5 10 15
Lys

<210> 84
<211> 21
<212> PRT
<213> Homo sapiens

<400> 84
Leu Ala Val Gly Ala Leu Leu Val Cys Ala Val Leu Gly Leu Cys Leu
1 5 10 15

Ala Val Pro Asp Lys
20

<210> 85
<211> 19
<212> PRT
<213> Homo sapiens

<400> 85
Ser Ala Gly Trp Asn Ile Pro Ile Gly Leu Leu Tyr Cys Asp Leu Pro
1 5 10 15
Glu Pro Arg

<210> 86
<211> 21
<212> PRT
<213> Homo sapiens

<400> 86
Ile Phe Ser Val Tyr Cys Asp Gln Glu Thr Ser Leu Gly Gly Trp Leu
1 5 10 15
Leu Ile Gln Gln Arg
20

<210> 87
<211> 6
<212> PRT
<213> Homo sapiens

<400> 87
Ser Ser Ser Val Ser Arg
1 5

<210> 88
<211> 5
<212> PRT
<213> Homo sapiens

<400> 88
Gln Asn Gln Asn Arg
1 5

<210> 89
<211> 5
<212> PRT
<213> Homo sapiens

<400> 89
Gln Leu Asn Tyr Lys
1 5

<210> 90
<211> 6
<212> PRT

<213> Homo sapiens

<400> 90

Val Ser Val Gln Leu Lys
1 5

<210> 91

<211> 6

<212> PRT

<213> Homo sapiens

<400> 91

Thr Phe Ala Gln Ala Arg
1 5

<210> 92

<211> 6

<212> PRT

<213> Homo sapiens

<400> 92

Val Phe Thr Asn Ser Lys
1 5

<210> 93

<211> 7

<212> PRT

<213> Homo sapiens

<400> 93

Gly Val Pro Ile Pro Asn Lys
1 5

<210> 94

<211> 7

<212> PRT

<213> Homo sapiens

<400> 94

Leu Pro Ser Asn Val Val Lys
1 5

<210> 95

<211> 7

<212> PRT

<213> Homo sapiens

<400> 95

Gly Glu Val Phe Thr Leu Lys
1 5

<210> 96

<211> 7

<212> PRT

<213> Homo sapiens

<400> 96

Gly Pro Thr Gln Asp Phe Arg
1 5

<210> 97

<211> 7

<212> PRT

<213> Homo sapiens

<400> 97

Tyr Asn Ile Leu Pro Glu Lys
1 5

<210> 98

<211> 8

<212> PRT

<213> Homo sapiens

<400> 98

Tyr Gly Ala Ala Thr Phe Thr Arg
1 5

<210> 99

<211> 8

<212> PRT

<213> Homo sapiens

<400> 99

Glu Asp Ser Pro Phe Ala Leu Lys
1 5

<210> 100

<211> 8

<212> PRT

<213> Homo sapiens

<400> 100

Glu Ile Leu Asn Ser Leu Asp Lys
1 5

<210> 101

<211> 8

<212> PRT

<213> Homo sapiens

<400> 101

Ile Ile Ser Ile Met Asp Glu Lys
1 5

<210> 102

<211> 8

<212> PRT

<213> Homo sapiens

<400> 102

Ile Ala Gln Trp Gln Ser Leu Lys

1 5

<210> 103

<211> 8

<212> PRT

<213> Homo sapiens

<400> 103

Ala Thr Val Leu Asn Tyr Leu Pro

1 5

<210> 104

<211> 10

<212> PRT

<213> Homo sapiens

<400> 104

Val Val Val Gln Thr Glu Ser Gly Gly Arg

1 5 10

<210> 105

<211> 11

<212> PRT

<213> Homo sapiens

<400> 105

Ser Ser Gly Ser Leu Leu Asn Asn Ala Ile Lys

1 5 10

<210> 106

<211> 10

<212> PRT

<213> Homo sapiens

<400> 106

Ile Ser Glu Ile Thr Asn Ile Val Ser Lys

1 5 10

<210> 107

<211> 10

<212> PRT

<213> Homo sapiens

<400> 107

Ser Leu Phe Thr Asp Leu Val Ala Glu Lys

1 5 10

<210> 108

<211> 10

<212> PRT

<213> Homo sapiens

<400> 108

Gln Thr Leu Ser Trp Thr Val Thr Pro Lys
1 5 10

<210> 109

<211> 12

<212> PRT

<213> Homo sapiens

<400> 109

Ala Ser Pro Ala Phe Leu Ala Ser Gln Asn Thr Lys
1 5 10

<210> 110

<211> 11

<212> PRT

<213> Homo sapiens

<400> 110

Ala Val Gly Tyr Leu Ile Thr Gly Tyr Gln Arg
1 5 10

<210> 111

<211> 12

<212> PRT

<213> Homo sapiens

<400> 111

Ala Leu Leu Ala Tyr Ala Phe Ser Leu Leu Gly Lys
1 5 10

<210> 112

<211> 11

<212> PRT

<213> Homo sapiens

<400> 112

Val Val Ser Val Asp Glu Asn Phe Arg Pro Arg
1 5 10

<210> 113

<211> 13

<212> PRT

<213> Homo sapiens

<400> 113

Met Leu Gln Ile Thr Asn Thr Gly Phe Glu Met Ile Lys
1 5 10

<210> 114

<211> 13

<212> PRT

<213> Homo sapiens

<400> 114

Met Val Ser Gly Phe Ile Pro Leu Lys Pro Thr Val Lys
1 5 10

<210> 115

<211> 13

<212> PRT

<213> Homo sapiens

<400> 115

Asn Gln Gly Asn Thr Trp Leu Thr Ala Phe Val Leu Lys
1 5 10

<210> 116

<211> 13

<212> PRT

<213> Homo sapiens

<400> 116

Asn Glu Leu Ile Pro Leu Ile Tyr Leu Glu Asn Pro Arg
1 5 10

<210> 117

<211> 15

<212> PRT

<213> Homo sapiens

<400> 117

Ile Ser Ala Ser Ser Glu Val Ala Phe Leu Ser Ile Gln Ile Lys
1 5 10 15

<210> 118

<211> 15

<212> PRT

<213> Homo sapiens

<400> 118

Gln Gly Ile Pro Phe Phe Ala Gln Val Leu Leu Val Asp Gly Lys
1 5 10 15

<210> 119

<211> 18

<212> PRT

<213> Homo sapiens

<400> 119

Gly Ser Phe Ala Leu Ser Phe Pro Val Glu Ser Asp Val Ala Pro Ile
1 5 10 15
Ala Arg

<210> 120

<211> 17

<212> PRT

<213> Homo sapiens

<400> 120

Ala Phe Gln Pro Phe Phe Val Glu Leu Thr Met Pro Tyr Ser Val Ile
1 5 10 15
Arg

<210> 121

<211> 22

<212> PRT

<213> Homo sapiens

<400> 121

Leu Glu Ala Gly Ile Asn Gln Leu Ser Phe Pro Leu Ser Ser Glu Pro
1 5 10 15
Ile Gln Gly Ser Tyr Arg
20

<210> 122

<211> 28

<212> PRT

<213> Homo sapiens

<400> 122

Asn Thr Val Leu Val Leu Asn Thr Gln Ser Leu Val Phe Val Gln Thr
1 5 10 15
Asp Lys Pro Met Tyr Lys Pro Gly Gln Thr Val Arg
20 25

<210> 123

<211> 5

<212> PRT

<213> Homo sapiens

<400> 123

Ile Glu Val Leu Lys
1 5

<210> 124

<211> 5

<212> PRT

<213> Homo sapiens

<400> 124

Leu Ile Ile Ser Arg
1 5

<210> 125

<211> 5

<212> PRT

<213> Homo sapiens

<400> 125

Glu Val Asp Leu Arg
1 5

<210> 126

<211> 5

<212> PRT

<213> Homo sapiens

<400> 126

Ser Gln Leu Gln Lys
1 5

<210> 127

<211> 5

<212> PRT

<213> Homo sapiens

<400> 127

Val Pro Met Met Lys
1 5

<210> 128

<211> 5

<212> PRT

<213> Homo sapiens

<400> 128

Ser Arg Pro Val Arg
1 5

<210> 129

<211> 5

<212> PRT

<213> Homo sapiens

<400> 129

Lys Pro Leu Glu Lys
1 5

<210> 130

<211> 5

<212> PRT

<213> Homo sapiens

<400> 130

Leu Val Ala Ala Ser
1 5

<210> 131

<211> 5

<212> PRT

<213> Homo sapiens

<400> 131

Asn Leu Asn Glu Lys
1 5

<210> 132
<211> 5
<212> PRT
<213> Homo sapiens

<400> 132
Leu Glu Thr Glu Lys
1 5

<210> 133
<211> 6
<212> PRT
<213> Homo sapiens

<400> 133
Ala Val Gly Asn Leu Arg
1 5

<210> 134
<211> 5
<212> PRT
<213> Homo sapiens

<400> 134
Asp Leu Leu Phe Arg
1 5

<210> 135
<211> 5
<212> PRT
<213> Homo sapiens

<400> 135
Leu Asp Glu Leu Arg
1 5

<210> 136
<211> 5
<212> PRT
<213> Homo sapiens

<400> 136
Glu Phe Thr Val Arg
1 5

<210> 137
<211> 5
<212> PRT
<213> Homo sapiens

<400> 137
Asn Thr Tyr Glu Lys
1 5

<210> 138
<211> 6
<212> PRT
<213> Homo sapiens

<400> 138
Ser Ser Ser Tyr Ser Lys
1 5

<210> 139
<211> 5
<212> PRT
<213> Homo sapiens

<400> 139
Gln Glu Pro Glu Arg
1 5

<210> 140
<211> 6
<212> PRT
<213> Homo sapiens

<400> 140
Gln Glu Asp Ser Gly Lys
1 5

<210> 141
<211> 5
<212> PRT
<213> Homo sapiens

<400> 141
Asp Leu Leu Phe Arg
1 5

<210> 142
<211> 5
<212> PRT
<213> Homo sapiens

<400> 142
Met Phe Ser Met Arg
1 5

<210> 143
<211> 6
<212> PRT
<213> Homo sapiens

<400> 143
Ala Trp Ala Val Ala Arg
1 5

<210> 144
<211> 6
<212> PRT
<213> Homo sapiens

<400> 144
Thr Pro Val Ser Asp Arg
1 5

<210> 145
<211> 6
<212> PRT
<213> Homo sapiens

<400> 145
Ile Val Asp Leu Val Lys
1 5

<210> 146
<211> 5
<212> PRT
<213> Homo sapiens

<400> 146
Leu Gln Phe Glu Arg
1 5

<210> 147
<211> 6
<212> PRT
<213> Homo sapiens

<400> 147
Asn Tyr Ala Glu Ala Lys
1 5

<210> 148
<211> 6
<212> PRT
<213> Homo sapiens

<400> 148
Asp Leu Leu Pro Ser Arg
1 5

<210> 149
<211> 6
<212> PRT
<213> Homo sapiens

<400> 149
Val Ser Glu Asp Leu Arg
1 5

<210> 150
<211> 7
<212> PRT
<213> Homo sapiens

<400> 150
Gly Gln Leu Asn Val Ala Lys
1 5

<210> 151
<211> 7
<212> PRT
<213> Homo sapiens

<400> 151
Gly Asp Val Ala Phe Val Lys
1 5

<210> 152
<211> 6
<212> PRT
<213> Homo sapiens

<400> 152
Phe Leu Glu Asp Val Lys
1 5

<210> 153
<211> 6
<212> PRT
<213> Homo sapiens

<400> 153
Val Pro Pro Glu Trp Lys
1 5

<210> 154
<211> 8
<212> PRT
<213> Homo sapiens

<400> 154
Ala Ser Asp Thr Glu Ala Gly Lys
1 5

<210> 155
<211> 7
<212> PRT
<213> Homo sapiens

<400> 155
Ser Pro Leu Phe Met Gly Lys
1 5

<210> 156
<211> 7
<212> PRT
<213> Homo sapiens

<400> 156
Gly Ala Asp Tyr Ser Leu Arg
1 5

<210> 157
<211> 6
<212> PRT
<213> Homo sapiens

<400> 157
Asp Asn Thr Tyr Asn Arg
1 5

<210> 158
<211> 7
<212> PRT
<213> Homo sapiens

<400> 158
Val Leu Gln Pro Val Thr Lys
1 5

<210> 159
<211> 7
<212> PRT
<213> Homo sapiens

<400> 159
Val Val Asn Pro Thr Gln Lys
1 5

<210> 160
<211> 7
<212> PRT
<213> Homo sapiens

<400> 160
Gln Ser Leu Gly Glu Val Arg
1 5

<210> 161
<211> 7
<212> PRT
<213> Homo sapiens

<400> 161
Leu Val Thr Asp Leu Thr Lys
1 5

<210> 162
<211> 7
<212> PRT
<213> Homo sapiens

<400> 162
Ala Gln Leu Val Asp Met Lys
1 5

<210> 163
<211> 7
<212> PRT
<213> Homo sapiens

<400> 163
Ile Arg Pro Leu Val Thr Gln
1 5

<210> 164
<211> 7
<212> PRT
<213> Homo sapiens

<400> 164
Asn Pro Asp Pro Trp Ala Lys
1 5

<210> 165
<211> 6
<212> PRT
<213> Homo sapiens

<400> 165
Thr Trp Gln Asp Tyr Lys
1 5

<210> 166
<211> 8
<212> PRT
<213> Homo sapiens

<400> 166
Gly Asp Ser Thr Phe Glu Ser Lys
1 5

<210> 167
<211> 8
<212> PRT
<213> Homo sapiens

<400> 167
Met Ala Gly Arg Pro Gln Gly Arg
1 5

<210> 168
<211> 8
<212> PRT
<213> Homo sapiens

<400> 168
Ala Glu Ser Ser Asn Asn Val Arg
1 5

<210> 169
<211> 7
<212> PRT
<213> Homo sapiens

<400> 169
Lys Pro Val Asp Glu Tyr Lys
1 5

<210> 170
<211> 8
<212> PRT
<213> Homo sapiens

<400> 170
Ala Glu Phe Ala Glu Val Ser Lys
1 5

<210> 171
<211> 8
<212> PRT
<213> Homo sapiens

<400> 171
Ala Val Leu Thr Ile Asp Glu Lys
1 5

<210> 172
<211> 8
<212> PRT
<213> Homo sapiens

<400> 172
Thr Ser Asn Leu Leu Val Ile Arg
1 5

<210> 173
<211> 7
<212> PRT
<213> Homo sapiens

<400> 173
Tyr Leu Tyr Glu Ile Ala Arg
1 5

<210> 174
<211> 8
<212> PRT
<213> Homo sapiens

<400> 174
Gln Leu Glu Gln Val Ile Ala Lys
1 5

<210> 175
<211> 8
<212> PRT
<213> Homo sapiens

<400> 175
Asp Asp Asn Pro Asn Leu Pro Arg
1 5

<210> 176
<211> 8
<212> PRT
<213> Homo sapiens

<400> 176
Phe Gln Asn Ala Leu Leu Val Arg
1 5

<210> 177
<211> 9
<212> PRT
<213> Homo sapiens

<400> 177
Gly Asp Phe Ser Ser Ala Asn Asn Arg
1 5

<210> 178
<211> 10
<212> PRT
<213> Homo sapiens

<400> 178
Asp Gly Ala Gly Asp Val Ala Phe Val Lys
1 5 10

<210> 179
<211> 9
<212> PRT
<213> Homo sapiens

<400> 179
Asp Gly Gly Val Gln Trp Gly Asp Arg
1 5

<210> 180
<211> 9
<212> PRT
<213> Homo sapiens

<400> 180
Thr Val Thr Gly Glu Phe Thr Asp Lys
1 5

<210> 181
<211> 8
<212> PRT
<213> Homo sapiens

<400> 181
Tyr Leu Gly Glu Glu Tyr Val Lys
1 5

<210> 182
<211> 9
<212> PRT
<213> Homo sapiens

<400> 182
Gln Thr Ala Leu Val Glu Leu Val Lys
1 5

<210> 183
<211> 8
<212> PRT
<213> Homo sapiens

<400> 183
Gln Ile Asn Asp Tyr Val Glu Lys
1 5

<210> 184
<211> 10
<212> PRT
<213> Homo sapiens

<400> 184
Val Thr Ser Gly Ser Thr Thr Thr Thr Arg
1 5 10

<210> 185
<211> 10
<212> PRT
<213> Homo sapiens

<400> 185
Ser Val Leu Gly Gln Leu Gly Ile Thr Lys
1 5 10

<210> 186
<211> 9
<212> PRT
<213> Homo sapiens

<400> 186
Glu Ser Asp Ser Val Phe Val Leu Lys
1 5

<210> 187
<211> 8
<212> PRT
<213> Homo sapiens

<400> 187
Asn Ser Leu Phe Glu Tyr Gln Lys
1 5

<210> 188
<211> 8
<212> PRT
<213> Homo sapiens

<400> 188
Ser Asn Ile Met Trp Leu Asp Arg
1 5

<210> 189
<211> 9
<212> PRT
<213> Homo sapiens

<400> 189
Met Asp Gly Ser Leu Asn Phe Asn Arg
1 5

<210> 190
<211> 9
<212> PRT
<213> Homo sapiens

<400> 190
Ala Leu Thr Asp Met Pro Gln Met Arg
1 5

<210> 191
<211> 9
<212> PRT
<213> Homo sapiens

<400> 191
Leu Ser Ser Trp Val Leu Leu Met Arg
1 5

<210> 192
<211> 8
<212> PRT
<213> Homo sapiens

<400> 192
Trp Glu Arg Pro Phe Glu Val Lys
1 5

<210> 193
<211> 10
<212> PRT
<213> Homo sapiens

<400> 193
Leu Ser Ile Thr Gly Thr Tyr Asp Leu Lys
1 5 10

<210> 194
<211> 11
<212> PRT
<213> Homo sapiens

<400> 194
Gly Ser Glu Ser Gly Ile Phe Thr Asn Thr Lys
1 5 10

<210> 195
<211> 10
<212> PRT
<213> Homo sapiens

<400> 195
Leu Val Asn Glu Val Thr Glu Phe Ala Lys
1 5 10

<210> 196
<211> 10
<212> PRT
<213> Homo sapiens

<400> 196
Gln Phe Thr Ser Ser Thr Ser Tyr Asn Arg
1 5 10

<210> 197
<211> 10
<212> PRT
<213> Homo sapiens

<400> 197
Asp Ser Gly Phe Gln Met Asn Gln Leu Arg
1 5 10

<210> 198
<211> 11
<212> PRT
<213> Homo sapiens

<400> 198
Ser Ala Ser Asp Leu Thr Trp Asp Asn Leu Lys
1 5 10

<210> 199
<211> 10
<212> PRT
<213> Homo sapiens

<400> 199
Phe Ser Val Ile Trp Gln Leu Val Asp Arg
1 5 10

<210> 200
<211> 11
<212> PRT
<213> Homo sapiens

<400> 200
Glu Gly Tyr Tyr Gly Tyr Thr Gly Ala Phe Arg
1 5 10

<210> 201
<211> 13
<212> PRT
<213> Homo sapiens

<400> 201
Ser Glu Gly Ala Val Val Asn Val Gln Pro Thr Asp Lys
1 5 10

<210> 202
<211> 13
<212> PRT
<213> Homo sapiens

<400> 202
Met Lys Pro Val Pro Asp Leu Val Pro Gly Asn Phe Lys
1 5 10

<210> 203
<211> 12
<212> PRT
<213> Homo sapiens

<400> 203
Met Tyr Leu Gly Tyr Glu Tyr Val Thr Ala Ile Arg
1 5 10

<210> 204
<211> 13
<212> PRT
<213> Homo sapiens

<400> 204
Asp Gly Thr Val Gln Pro Gly Ser Ser Tyr Trp Glu Arg
1 5 10

<210> 205
<211> 12
<212> PRT
<213> Homo sapiens

<400> 205
Asn Tyr Asn Asn Thr Trp Thr Gln Leu Ala Glu Arg
1 5 10

<210> 206
<211> 14
<212> PRT
<213> Homo sapiens

<400> 206
Val Pro Gln Val Ser Thr Pro Thr Leu Val Glu Val Ser Arg
1 5 10

<210> 207
<211> 16
<212> PRT
<213> Homo sapiens

<400> 207
Gly Gly Ser Thr Ser Tyr Gly Thr Gly Ser Glu Thr Glu Ser Pro Arg
1 5 10 15

<210> 208
<211> 13
<212> PRT
<213> Homo sapiens

<400> 208
Asp Thr Val Phe Ala Leu Val Asn Tyr Ile Phe Phe Lys
1 5 10

<210> 209
<211> 14
<212> PRT
<213> Homo sapiens

<400> 209
Thr Ala Gly Trp Asn Ile Pro Met Gly Leu Leu Tyr Asn Lys
1 5 10

<210> 210
<211> 13
<212> PRT
<213> Homo sapiens

<400> 210
Asp Val Phe Leu Gly Met Phe Leu Tyr Glu Tyr Ala Arg
1 5 10

<210> 211
<211> 14
<212> PRT
<213> Homo sapiens

<400> 211
Glu Asp Pro Gln Thr Phe Tyr Tyr Ala Val Ala Val Val Lys
1 5 10

<210> 212
<211> 14
<212> PRT
<213> Homo sapiens

<400> 212
Ile Thr Pro Asn Leu Ala Glu Phe Ala Phe Ser Leu Tyr Arg
1 5 10

<210> 213
<211> 16
<212> PRT
<213> Homo sapiens

<400> 213
Tyr Leu Gly Asn Ala Thr Ala Ile Phe Phe Leu Pro Asp Glu Gly Lys
1 5 10 15

<210> 214
<211> 15
<212> PRT
<213> Homo sapiens

<400> 214
Phe Asn Lys Pro Phe Val Phe Leu Met Ile Glu Gln Asn Thr Lys
1 5 10 15

<210> 215
<211> 21
<212> PRT
<213> Homo sapiens

<400> 215
Asn Pro Ser Ser Ala Gly Ser Trp Asn Ser Gly Ser Ser Gly Pro Gly
1 5 10 15
Ser Thr Gly Asn Arg
20

<210> 216
 <211> 17
 <212> PRT
 <213> Homo sapiens

<400> 216
 Asn Asn Ser Pro Tyr Glu Ile Glu Asn Gly Val Val Trp Val Ser Phe
 1 5 10 15
 Arg

<210> 217
 <211> 17
 <212> PRT
 <213> Homo sapiens

<400> 217
 Trp Val Thr Phe Ile Ser Leu Leu Phe Leu Phe Ser Ser Ala Tyr Ser
 1 5 10 15
 Arg

<210> 218
 <211> 21
 <212> PRT
 <213> Homo sapiens

<400> 218
 Ala Ser Thr Pro Ile Ser Ile Ile Ala Leu Glu Met Gly Phe Ala Val
 1 5 10 15
 Thr Ala Ile Ser Arg
 20

<210> 219
 <211> 24
 <212> PRT
 <213> Homo sapiens

<400> 219
 Thr Leu Asn Gln Pro Asp Ser Gln Leu Gln Leu Thr Thr Gly Asn Gly
 1 5 10 15
 Leu Phe Leu Ser Glu Gly Leu Lys
 20

<210> 220
 <211> 24
 <212> PRT
 <213> Homo sapiens

<400> 220
 Ser Ser Phe Gly Ser Val Gln Met Glu Gln Val Gln Pro Asn Ser Phe
 1 5 10 15
 Ser Leu Gly Ile Phe Asn Ser Arg
 20

<210> 221
 <211> 24
 <212> PRT
 <213> Homo sapiens

<400> 221
 Gln Glu Val Cys Glu Glu Phe Ser Gln Gln Leu Asn Ser Asn Gly Cys
 1 5 10 15
 Ile Thr Gln Gln Val His Thr Lys
 20

<210> 222
 <211> 28
 <212> PRT
 <213> Homo sapiens

<400> 222
 Val Tyr Asp Tyr Tyr Glu Thr Asp Glu Ser Val Val Ala Glu Tyr Ile
 1 5 10 15
 Ala Pro Cys Ser Thr Asp Thr Glu His Gly Asn Val
 20 25

<210> 223
 <211> 14
 <212> PRT
 <213> Homo sapiens

<400> 223
 Ala Glu Val Ser Asp Ser Gly Ser Tyr Tyr Cys His Met Glu
 1 5 10

<210> 224
 <211> 15
 <212> PRT
 <213> Homo sapiens

<400> 224
 Glu Asp Glu Gly Gln Tyr Glu Cys His Val Thr Glu Trp Val Arg
 1 5 10 15

<210> 225
 <211> 17
 <212> PRT
 <213> Homo sapiens

<400> 225
 Val His Thr Glu Cys Cys His Gly Asp Leu Leu Glu Cys Ala Asp Asp
 1 5 10 15
 Arg

<210> 226
 <211> 19
 <212> PRT
 <213> Homo sapiens

<400> 226

Glu Phe Asn Ala Glu Thr Phe Thr Phe His Ala Asp Ile Cys Thr Leu
 1 5 10 15
 Ser Glu Lys

<210> 227

<211> 21

<212> PRT

<213> Homo sapiens

<400> 227

Asp Cys Asp Asp Val Leu Gln Thr His Pro Ser Gly Thr Gln Ser Gly
 1 5 10 15
 Ile Phe Asn Ile Lys
 20

<210> 228

<211> 21

<212> PRT

<213> Homo sapiens

<400> 228

Met Pro Cys Ala Glu Asp Tyr Leu Ser Val Val Leu Asn Gln Leu Cys
 1 5 10 15
 Val Leu His Glu Lys
 20

<210> 229

<211> 22

<212> PRT

<213> Homo sapiens

<400> 229

Leu Val Arg Pro Glu Val Asp Val Met Cys Thr Ala Phe His Asp Asn
 1 5 10 15
 Glu Glu Thr Phe Leu Lys
 20

<210> 230

<211> 27

<212> PRT

<213> Homo sapiens

<400> 230

Ser His Cys Ile Ala Glu Val Glu Asn Asp Glu Met Pro Ala Asp Leu
 1 5 10 15
 Pro Ser Leu Ala Ala Asp Phe Val Glu Ser Lys
 20 25

<210> 231

<211> 14

<212> PRT

<213> Homo sapiens

<400> 231

Glu Asp Asn Leu Val His Trp Glu Arg Pro Gln Arg Pro Lys
1 5 10

<210> 232

<211> 14

<212> PRT

<213> Homo sapiens

<400> 232

His Pro Asp Glu Ala Ala Phe Phe Asp Thr Ala Ser Thr Gly
1 5 10

<210> 233

<211> 15

<212> PRT

<213> Homo sapiens

<400> 233

Thr Asp Thr Ser His His Asp Gln Asp His Pro Thr Phe Asn Lys
1 5 10 15

<210> 234

<211> 15

<212> PRT

<213> Homo sapiens

<400> 234

Leu Gln His Leu Glu Asn Glu Leu Thr His Asp Ile Ile Thr Lys
1 5 10 15

<210> 235

<211> 17

<212> PRT

<213> Homo sapiens

<400> 235

Met Ala Asp Glu Ala Gly Ser Glu Ala Asp His Glu Gly Thr His Ser
1 5 10 15
Thr

<210> 236

<211> 16

<212> PRT

<213> Homo sapiens

<400> 236

Asp Thr Glu Glu Asp Phe His Val Asp Gln Val Thr Thr Val Lys
1 5 10 15

<210> 237
 <211> 18
 <212> PRT
 <213> Homo sapiens

<400> 237
 Leu Tyr His Ser Glu Ala Phe Thr Val Asn Phe Gly Asp Thr Glu Glu
 1 5 10 15
 Ala Lys

<210> 238
 <211> 17
 <212> PRT
 <213> Homo sapiens

<400> 238
 Glu Asp Leu Ile Trp Glu Leu Leu Asn Gln Ala Gln Glu His Phe Gly
 1 5 10 15
 Lys

<210> 239
 <211> 18
 <212> PRT
 <213> Homo sapiens

<400> 239
 Val Glu Leu Glu Asp Trp Ala Gly Asn Glu Ala Tyr Ala Glu Tyr His
 1 5 10 15
 Phe Arg

<210> 240
 <211> 19
 <212> PRT
 <213> Homo sapiens

<400> 240
 Thr Thr His Asn Ser Ala Phe Glu Tyr Gly Thr Tyr Val Glu Glu Glu
 1 5 10 15
 Gly Leu Arg

<210> 241
 <211> 26
 <212> PRT
 <213> Homo sapiens

<400> 241
 Gly Phe Gly Ser Leu Asn Asp Glu Gly Glu Gly Glu Phe Trp Leu Gly
 1 5 10 15
 Asn Asp Tyr Leu His Leu Leu Thr Gln Arg
 20 25

<210> 242
<211> 12
<212> PRT
<213> Homo sapiens

<400> 242
Gly Glu Glu Ser Tyr Cys Ile Cys Gly Ser Glu Arg
1 5 10

<210> 243
<211> 31
<212> PRT
<213> Homo sapiens

<400> 243
Thr Leu Gly Asn Val Asn Phe Ser Val Ser Ala Glu Ala Met Gln Ser
1 5 10 15
Leu Glu Leu Cys Gly Asn Glu Val Val Glu Val Pro Glu Ile Lys
20 25 30

<210> 244
<211> 9
<212> PRT
<213> Homo sapiens

<400> 244
Glu Thr Cys Phe Ala Glu Glu Gly Lys
1 5

<210> 245
<211> 10
<212> PRT
<213> Homo sapiens

<400> 245
Glu Cys Cys Glu Lys Pro Leu Leu Glu Lys
1 5 10

<210> 246
<211> 11
<212> PRT
<213> Homo sapiens

<400> 246
Asp Tyr Glu Leu Leu Cys Leu Asp Gly Thr Arg
1 5 10

<210> 247
<211> 12
<212> PRT
<213> Homo sapiens

<400> 247
Glu Thr Tyr Gly Glu Met Ala Asp Cys Cys Ala Lys
1 5 10

<210> 248
<211> 13
<212> PRT
<213> Homo sapiens

<400> 248
Thr Cys Val Ala Asp Glu Ser Ala Glu Asn Cys Asp Lys
1 5 10

<210> 249
<211> 12
<212> PRT
<213> Homo sapiens

<400> 249
Asp Gln Tyr Glu Leu Leu Cys Leu Asp Asn Thr Arg
1 5 10

<210> 250
<211> 14
<212> PRT
<213> Homo sapiens

<400> 250
Phe Asp Glu Phe Phe Ser Glu Gly Cys Ala Pro Gly Ser Lys
1 5 10

<210> 251
<211> 13
<212> PRT
<213> Homo sapiens

<400> 251
Gln Asn Cys Glu Leu Phe Glu Gln Leu Gly Glu Tyr Lys
1 5 10

<210> 252
<211> 15
<212> PRT
<213> Homo sapiens

<400> 252
Ile Glu Cys Val Ser Ala Glu Thr Thr Glu Asp Cys Ile Ala Lys
1 5 10 15

<210> 253
<211> 16
<212> PRT
<213> Homo sapiens

<400> 253
Glu Gly Thr Cys Pro Glu Ala Pro Thr Asp Glu Cys Lys Pro Val Lys
1 5 10 15

<210> 254
 <211> 16
 <212> PRT
 <213> Homo sapiens

<400> 254
 Arg Pro Cys Phe Ser Ala Leu Glu Val Asp Glu Thr Tyr Val Pro Lys
 1 5 10 15

<210> 255
 <211> 15
 <212> PRT
 <213> Homo sapiens

<400> 255
 Asp Ser Asp Trp Pro Phe Cys Ser Asp Glu Asp Trp Asn Tyr Lys
 1 5 10 15

<210> 256
 <211> 19
 <212> PRT
 <213> Homo sapiens

<400> 256
 Ser Asp Asn Cys Glu Asp Thr Pro Glu Ala Gly Tyr Phe Ala Val Ala
 1 5 10 15
 Val Val Lys

<210> 257
 <211> 30
 <212> PRT
 <213> Homo sapiens

<400> 257
 Ile Val Cys Leu Val Leu Ser Val Val Gly Thr Ala Trp Thr Ala Asp
 1 5 10 15
 Ser Gly Glu Gly Asp Phe Leu Ala Glu Gly Gly Gly Val Arg
 20 25 30

<210> 258
 <211> 30
 <212> PRT
 <213> Homo sapiens

<400> 258
 Glu Val Val Thr Ser Glu Asp Gly Ser Asp Cys Pro Glu Ala Met Asp
 1 5 10 15
 Leu Gly Thr Leu Ser Gly Ile Gly Thr Leu Asp Gly Phe Arg
 20 25 30

<210> 259
 <211> 12
 <212> PRT
 <213> Homo sapiens

<400> 259

Glu Glu Gly Thr Asp Leu Glu Val Thr Ala Asn Arg
1 5 10

<210> 260

<211> 13

<212> PRT

<213> Homo sapiens

<400> 260

Thr Leu Leu Val Glu Ala Glu Gly Ile Glu Gln Glu Lys
1 5 10

<210> 261

<211> 18

<212> PRT

<213> Homo sapiens

<400> 261

Met Phe Ile Phe Ala Ile Leu Pro Asp Gly Glu Val Val Gly Asp Ser
1 5 10 15
Glu Lys

<210> 262

<211> 27

<212> PRT

<213> Homo sapiens

<400> 262

Ala Val Asp Gln Ser Val Leu Leu Met Lys Pro Glu Ala Glu Leu Ser
1 5 10 15
Val Ser Ser Val Tyr Asn Leu Leu Thr Val Lys
20 25

<210> 263

<211> 7

<212> PRT

<213> Homo sapiens

<400> 263

Phe Leu Glu Asn Glu Asp Arg
1 5

<210> 264

<211> 7

<212> PRT

<213> Homo sapiens

<400> 264

Asp Tyr Glu Asp Gln Gln Lys
1 5

<210> 265
<211> 8
<212> PRT
<213> Homo sapiens

<400> 265
Leu Glu Val Asp Ile Asp Ile Lys
1 5

<210> 266
<211> 8
<212> PRT
<213> Homo sapiens

<400> 266
Asp Leu Gly Glu Glu Asn Phe Lys
1 5

<210> 267
<211> 12
<212> PRT
<213> Homo sapiens

<400> 267
Ala Val Met Asp Asp Phe Ala Ala Phe Val Glu Lys
1 5 10

<210> 268
<211> 12
<212> PRT
<213> Homo sapiens

<400> 268
Val Val Asp Gly Glu Trp Gln Ile Val Gly Glu Arg
1 5 10

<210> 269
<211> 13
<212> PRT
<213> Homo sapiens

<400> 269
Met Glu Leu Glu Arg Pro Gly Gly Asn Glu Ile Thr Arg
1 5 10

<210> 270
<211> 13
<212> PRT
<213> Homo sapiens

<400> 270
Gly Leu Ile Asp Glu Val Asn Gln Asp Phe Thr Asn Arg
1 5 10

<210> 271
<211> 18
<212> PRT
<213> Homo sapiens

<400> 271
Val Phe Ser Asn Gly Ala Asp Leu Ser Gly Val Thr Glu Glu Ala Pro
1 5 10 15
Leu Lys

<210> 272
<211> 17
<212> PRT
<213> Homo sapiens

<400> 272
Val Phe Asp Glu Phe Lys Pro Leu Val Glu Glu Pro Gln Asn Leu Ile
1 5 10 15
Lys

<210> 273
<211> 21
<212> PRT
<213> Homo sapiens

<400> 273
Ile Met Asn Gly Glu Ala Asp Ala Met Ser Leu Asp Gly Gly Phe Val
1 5 10 15
Tyr Ile Ala Gly Lys
20

<210> 274
<211> 22
<212> PRT
<213> Homo sapiens

<400> 274
Gly Thr Glu Ala Ala Gly Ala Met Phe Leu Glu Ala Ile Pro Met Ser
1 5 10 15
Ile Pro Pro Glu Val Lys
20

<210> 275
<211> 20
<212> PRT
<213> Homo sapiens

<400> 275
Thr Phe Pro Gly Phe Phe Ser Pro Met Leu Gly Glu Phe Val Ser Glu
1 5 10 15
Thr Glu Ser Arg
20

<210> 276
<211> 30
<212> PRT
<213> Homo sapiens

<400> 276
Gly Gly Val Glu Asp Glu Ala Thr Leu Ser Ala Tyr Val Thr Ile Ala
1 5 10 15
Leu Leu Glu Ile Pro Leu Pro Val Thr Asn Pro Ile Val Arg
20 25 30

<210> 277
<211> 31
<212> PRT
<213> Homo sapiens

<400> 277
Asn Ile Pro Ile Ile Val Leu Pro Leu Glu Ser Ser Ile Ser Val Gly
1 5 10 15
Val Ala Ser Asn Ala Ser Val Ile Leu Glu Gly Asp Asp Leu Arg
20 25 30