

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
17 February 2005 (17.02.2005)

PCT

(10) International Publication Number
WO 2005/015239 A2

(51) International Patent Classification⁷: **G01N 33/68**

[US/US]; 1161 York Avenue, Apartment 8F, New York, New York 10021 (US).

(21) International Application Number:

PCT/US2004/025464

(74) Agents: STEFFEY, Charles, E. et al.; Schwegman, Lundberg, Woessner & Kluth, P.A., P.O. Box 2938, Minneapolis, MN 55402 (US).

(22) International Filing Date: 6 August 2004 (06.08.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/493,159

7 August 2003 (07.08.2003) US

(71) Applicant (for all designated States except US): **CORNELL RESEARCH FOUNDATION, INC.** [US/US]; 20 Thornwood Drive, Suite 105, Ithaca, New York 14850 (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.

(72) Inventor; and

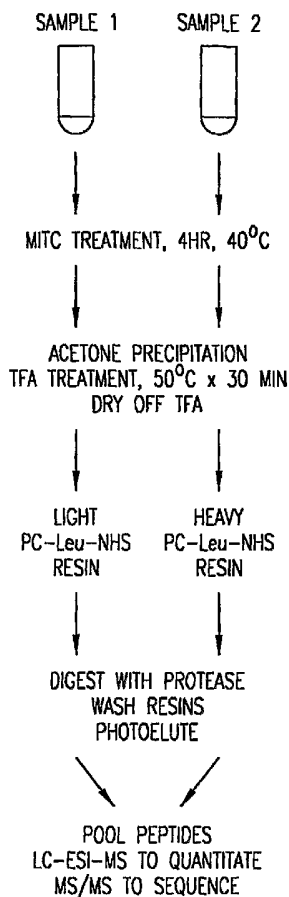
(75) Inventor/Applicant (for US only): **JAFFREY, Samie, R.**

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH,

[Continued on next page]

(54) Title: METHOD FOR N-TERMINAL LABELING OF PROTEINS

(57) Abstract: A method to prepare N-terminal modified peptides is provided.



WO 2005/015239 A2



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.

METHOD FOR N-TERMINAL LABELING OF PROTEINS

Cross-Reference to Related Applications

This application claims the benefit of the filing date of U.S. application
5 Serial No. 60/493,159, filed August 7, 2003, the disclosure of which is
incorporated herein by reference.

Background of the Invention

Over the past twenty years there have been tremendous advances in the
10 understanding of the molecular basis of diseases such as cancer, neurological
disorders, rheumatologic disorders, mental illness, and cardiovascular diseases.
However, substantial gaps both in the understanding of the pathogenesis as well
as the development of strategies for the early detection of disease remain. The
goal of proteomics is the identification and quantitation of proteins, post-
15 translational modifications, or functional properties of proteins in a biological
sample. Proteomics can provide insight into disease in several ways. For
instance, proteomics can identify proteins whose expression or levels of
expression are altered in breast cancer, or which proteins may be required for the
development or progression of breast cancer. Moreover, some proteins with
20 altered expression patterns may be useful to classify a disease into subtypes. This
sort of classification system may improve prognosis or choices of therapeutic
regimens. Lastly, expression patterns may help to identify proteins that could
serve as biomarkers for the early diagnosis of diseases. Markers that identify
diseases at a stage where medical intervention is possible would significantly
25 reduce the morbidity or mortality of various diseases.

Proteases are proteins that hydrolyze proteins at specific amino acid
residues. They can digest proteins at numerous sites, and thus have a role in
general protein degradation, or they can cleave proteins at highly specific sites,
often modulating the activity, function, or structure of their substrates. Protease
30 activity is often modulated in disease, thus proteases are often considered as
excellent targets for the chemotherapy of disease. However, determining
whether a protease is active in a tissue, e.g., by measuring protease activity, or
determining the substrate(s) of a protease, can be difficult.

2D-electrophoresis has historically been the primary tool of proteomics, although mass spectrometry (MS) is being recognized as having the potential to provide more information in a much higher throughput fashion. MS provides molecular weight data for peptides, e.g., peptides between 5 and 25 amino acids in length. An additional power of MS is called MS/MS, in which peptides are chosen in a first round of MS and then fragmented, and the mass of the fragments determined in a second round of MS. The fragmentation spectra can then be interpreted by a computer that compares it to virtual fragmentation spectra of peptides predicted by theoretical digests of protein sequences in databases. The computer makes predictions about the likely amino acid sequence, thereby identifying the protein from which the peptide originated.

The simplest MS-based proteomic experiment involves trypsinizing a cellular lysate, loading the peptides onto a MS instrument, and sequencing each peptide by MS/MS. The limitation of this approach is that only 200 peptides can be analyzed at a time. This is clearly not feasible at a cellular level. For example, digestion of the yeast proteome, which has about 6,000 proteins, results in 344,000 peptides (Gygi et al., 1999). Although liquid chromatography (LC) prior to MS may reduce the complexity of the digested mixture by two orders of magnitude, a mixture representing a proteome would still be far too complex to be analyzed by LC/MS. Peptide complexity hampers MS for specific reasons: (1) in highly complex mixtures, individual peaks seen on MS may represent combinations of several different peptides with a similar molecular weight, making subsequent fragmentation spectra uninterpretable; (2) because usually only about 1 µg of a peptide mixture is analyzed in a typical MS experiment, low abundance peptides may not achieve the requisite threshold concentration needed for detection on MS; and (3) during the ionization process in MS, a phenomenon called suppression occurs (Gustavsson et al., 2001), in which abundant peptides are ionized much more efficiently relative to the ionization of low-abundance peptides, and nonionized peptides cannot be detected. Thus, a major effort of proteomics research is to develop methods to simplify complex mixtures so that fewer peptides are generated per protein present in the starting material.

A major development to overcome the complexity problem is the isotope-coded affinity tag (ICAT) method (Gygi et al., 1999). The method

utilizes proteins in two samples and specific ICAT reagents which are composed of three components: a chemically-reactive iodoacetamidyl group that reacts with cysteines; a linker which can be isotopically labeled; and a biotin tag (Gygi et al., 1999). One sample is treated with the ICAT reagent, resulting in labeling of cysteine residues. The other sample is treated with a similar ICAT reagent except that eight of the hydrogens in the linker are replaced with ^2H atoms. This reagent is referred to as the "heavy" ICAT reagent. Next, proteins from the two samples are mixed, digested with trypsin, and cysteine-containing peptides are affinity isolated using avidin-agarose. Most peptides do not bind to avidin and are washed away, resulting in a substantial reduction in the complexity of the peptide mixture. Most proteins have at least one cysteine, so that the final peptide mixture contains representatives from nearly every protein. Because both heavy and light ICAT reagents are used, when MS is performed, every peptide appears as a doublet. The lower molecular weight peak in the doublet represents a peptide from a protein treated with the light reagent, and the higher molecular weight peak, separated by a mass difference of 8, represents the same peptide derived from a protein treated with the heavy reagent. The relative ratio of their signal intensities allows for quantification of the tagged peptides, and MS/MS is used to determine their amino acid sequence.

The ICAT protocol has been used mostly with yeast or bacteria, but not extensively with mammalian cells, for several reasons. One reason is that the complexity of the eukaryotic proteome is so much more vast than yeast that the reduction in peptide complexity achieved by ICAT is still not sufficient for analyzing the eukaryotic proteome at a sufficient depth. For example, when applied to yeast which have 6,000 proteins, the ICAT procedure reduces the number of peptides from 344,000 to approximately 30,000 and so can detect approximately 15% of this proteome (Gygi et al., 1999). Eukaryotic organisms possess at least 40,000 proteins and the complexity reduction that is achieved is not sufficient to detect moderate or low-abundance proteins. Other technical problems have been noted with the ICAT protocol that limit its effectiveness include nonspecific peptide elution (i.e., elution of peptides that do not contain biotin) due to technical issues with the ICAT protocol, complex MS/MS fragmentation spectra because of unusual fragmentation patterns of the ICAT reagent itself, non-coelution of heavy and light ICAT-labeled peptides on RP-LC

(Zhang et al., 2001), and ICAT reagents have some, albeit small, reactivity with residues other than cysteines (Boja et al., 2001).

Thus, there is a need for a method to simplify complex mixtures so that proteins in complex mixtures can be readily identified.

5

Summary of the Invention

The invention provides a method to label the N-terminus of proteins in a reliable, reproducible manner. For example, the method includes providing a hydrosylate comprising free (unlabeled) peptides and N-terminal transferring moiety-modified peptides. The hydrolysate is prepared by hydrolyzing a plurality of N-terminal transferring moiety-modified proteins obtained by contacting an N-terminal transferring moiety with a sample comprising a plurality of proteins. The plurality of proteins comprise proteins having side chain-modified amines and a free alpha amine group on the amino terminus. In one embodiment, the N-terminal transferring moiety comprises a tag, an optional linker, and a molecule with an amine reactive group, e.g., the N-terminal transferring moiety is any molecule that can be detected or isolated, for instance, due to the tag, and can react with an amine. In one embodiment, the molecule with the amine reactive group is covalently linked to the tag, such as an insoluble support, e.g., a bead or resin, via a linker. In one embodiment, the tag is biotin, iminobiotin, or a biotin-related molecule that can bind to avidin or streptavidin, or a solid support having those molecules. In one embodiment, the tag is dinitrophenol which can bind to anti-dinitrophenol antibodies, e.g., antibodies attached to a solid support. In one embodiment, the linker is an acid-labile or base-labile linker thus permitting the removal of the tag with acid or base, respectively, for example, after isolation of the N-terminal transferring moiety-modified proteins. In one embodiment, the linker comprises a photocleavable molecule. Thus, a complex protein mixture is treated with a series of chemical steps that ultimately results in the incorporation of a tag, optionally a linker, and a portion of the molecule with amine reactive group on the N-terminus of proteins in the mixture. In one embodiment, a hydrolysate is obtained by treatment with trypsin or another protease, or a chemical, e.g., cyanogen bromide or 2-(2-nitrophenylsulfenyl)-3-methylindole (BNPS-skatole), that results in a reproducible chemical cleavage of proteins. Peptides in the hydrosylate that are

derived from the N-terminus of proteins are recovered by virtue of their having the tag. In one embodiment, N-terminal transferring moiety-modified peptides in which the tag is linked to a bead are recovered by centrifugation or filtration. In another embodiment, N-terminal transferring moiety-modified peptides
5 having a tag which is an affinity marker such as biotin are recovered with a ligand for the affinity marker, for instance, an avidin-containing solid support. The isolation of the N-terminal modified peptides, and their subsequent identification, allows the characterization of proteins present in samples from humans and other mammalian sources. The ability to selectively isolate the
10 amino terminal peptides from a complex mixture of peptides provides a straightforward way to perform proteome analysis, and so the method of the invention overcomes the limitations of the current proteomic approaches and can be employed with MS-based approaches for highly complex mammalian proteomes. Moreover, the ability to selectively label the N-terminus of proteins
15 also provides a straightforward way to monitor the abundance of different N-terminally processed forms of proteins, since each processed form has a different N-terminus. Thus, the method of the invention permits proteomic analysis of N-terminal processing. For instance, one sample of cells is treated with a protease inhibitor, while another is treated with a control inactive compound. The
20 proteins from each sample are labeled according to the methods of the invention, and the products of protease cleavage, i.e., proteins that have been cleaved, can be detected and/or isolated. Identification of these cleavage products can identify the uncleaved protein as a substrate of a particle protease.

Also provided is a method of separating N-terminal modified peptides in
25 a mixture of N-terminal modified peptides. The method includes providing a hydrolysate comprising free peptides and N-terminal transferring moiety-modified peptides. The hydrolysate is prepared by hydrolyzing a plurality of N-terminal transferring moiety-modified proteins obtained by contacting an N-terminal transferring moiety with a sample comprising a plurality of proteins.
30 The plurality of proteins comprise proteins having side chain-modified amines and a free alpha amine group on the amino terminus. The N-terminal transferring moiety comprises a tag, optionally a linker, and a molecule with an amine reactive group. The N-terminal transferring moiety-modified peptides are isolated from the free peptides, and linker is optionally cleaved to generate a

mixture of purified N-terminal modified peptides, or the peptides are isolated based on affinity purification using an affinity tag present on the N-terminal transferring moiety-modified peptides. The purified N-terminal modified peptides may then be separated by standard microfractionation techniques prior to mass spectrometry. These techniques include but are not limited to reverse phase high pressure liquid chromatography on C8, C18, or C4 reverse phase columns, capillary electrophoresis, anion exchange chromatography, or a combination thereof.

Also provided is a method for comparing the amount or level of one or more proteins in at least two samples. The method includes providing a first sample comprising a plurality of N-terminal transferring moiety-modified proteins and a second sample comprising a plurality of N-terminal transferring moiety-modified proteins. The plurality of N-terminal transferring moiety-modified proteins in the first sample is prepared by contacting a first plurality of proteins comprising side chain-modified amines and a free alpha amine group on the amino terminus with a first N-terminal transferring moiety comprising a tag, a linker and a molecule which comprises an amine reactive group, which N-terminal transferring moiety contains a natural isotope of one or more elements, e.g., an abundant natural isotope. The plurality of N-terminal transferring moiety-modified proteins in the second sample is prepared by contacting a sample comprising a second plurality of proteins comprising side chain-modified amines and a free alpha amine group on the amino terminus, with a second N-terminal transferring moiety comprising a tag, a linker, and a molecule which comprises an amine reactive group, which N-terminal transferring moiety that is isotopically heavy relative to the naturally (abundant) isotopic form, e.g., one corresponding to the natural isotope. The plurality of N-terminal transferring moiety-modified proteins in the first and second samples is hydrolyzed and N-terminal transferring moiety-modified peptides in the first and second samples are isolated from free peptides. In one embodiment, the isolated N-terminal transferring moiety-modified peptides in the first and second samples are treated with an agent so as to yield purified N-terminal modified peptides which lack all or a portion of the linker. In another embodiment, the isolated N-terminal modified peptides are purified by selective elution from an affinity matrix. The purified N-terminal modified peptides are combined, then separated by

chromatography or capillary electrophoresis and their structure determined by mass spectrometry. Each peptide derived from the second sample has a measured mass that is larger than the same peptide from the first sample. The difference in the measured mass reflects the difference in the mass of the N-terminal transferring moiety that is attached to the proteins in the first and second samples. This mass difference derives from the use of isotopically heavy atoms in the N-terminal transferring moiety that is employed for the second sample. Then the amount or level of at least one separated peptide in the first sample is compared to the amount or level of the corresponding separated peptide in the second sample. The ratio of the signal obtained in mass spectrometry indicates the ratio of the abundance of the peptides in the sample, and, thus, the ratio of the abundance of the protein from which the N-terminal peptide was derived in the original samples.

Also provided is a method to identify peptides without comparing the amount or level in two samples. The method includes providing a sample comprising a plurality of N-terminal transferring moiety-modified proteins and a second sample comprising a plurality of N-terminal transferring moiety-modified proteins. The plurality of N-terminal transferring moiety-modified proteins in the first sample is prepared by contacting a plurality of proteins comprising side chain-modified amines and a free alpha amine group on the amino terminus with a label transferring moiety comprising a tag, a linker and a molecule with an amine reactive moiety. The plurality of N-terminal transferring moiety-modified proteins in the sample is hydrolyzed and N-terminal transferring moiety-modified peptides in the sample are isolated from free peptides. In one embodiment, the isolated N-terminal transferring moiety-modified peptides in the sample are treated with an agent so as to yield purified N-terminal modified peptides which lack all or a portion of the linker, e.g., which is particularly useful when the tag is a solid support. In another embodiment, the isolated N-terminal peptides are purified by selective elution from an affinity matrix. The purified N-terminal modified peptides are optionally separated by chromatography or capillary electrophoresis and their structure optionally determined by mass spectrometry.

Brief Description of the Drawings

Figure 1 shows Edman chemistry used for sequencing proteins.

Figure 2 shows the structure of a photocleavable (PC)-Leu-NHS resin.

Figure 3 shows steps in the synthesis of a PC-Leu-NHS resin.

5 Figure 4 depicts N-terminal labeling of two samples.

Figure 5A shows blocking of amines. To detect free amines, proteins were incubated with biotin-NHS (an amine-reactive biotinylation agent), and then Western blotted with biotin-specific antibodies. After treatment of BSA with methylisothiocyanate (MITC), amines were virtually completely blocked.

10 Figure 5B depicts amine reactivity following TFA treatment. MITC-treated BSA was treated with TFA, resulting in cyclization of the first amino acid and uncovering of an N-terminal amine.

Figure 6 shows monitoring of binding to and photoelution from a PC-Leu-NHS resin. (1) PC-Leu-NHS resin was pretreated with Tris to block the NHS moiety. MITC-modified BSA was incubated with this resin, the resin was washed, and the photoelute was blotted for BSA. No BSA was detected. (2) MITC-modified BSA was incubated with a PC-Leu-NHS resin, the resin was washed, and incubated in the dark instead of photoeluted. No BSA was detected. (3) MITC-modified BSA was incubated with a PC-Leu-NHS resin, the resin was washed, and photoeluted. BSA was detected in the photoeluate.

15
20

Detailed Description of the Invention

In principle, after a protein mixture has been digested, only one peptide is needed to identify the parent protein. To obtain only one peptide per protein, a label transferring moiety comprising a tag and a molecule having an amine reactive group may be attached to the N-terminus of each protein. Thus, after trypsinization, the tagged peptide can be recovered, resulting in the isolation of a pure population of N-terminally tagged peptides. However, the key problem when tagging the N-terminus of a protein is that any reagent which reacts with the amine group of the N-terminus also reacts with amine groups on lysine side chains and so recovered peptides contain the N-terminal peptides as well as peptides that contained lysine in the unmodified protein. The present method employs the chemistry of Edman degradation, a procedure used for N-terminal sequencing of proteins (Figure 1). Edman degradation involves three steps. In

25
30

the first step, proteins are reacted with phenylisothiocyanate (PITC).

Isothiocyanates react with amines to form thioureas. After PITC treatment, all amines in the protein are modified. In the second step, the modified protein is treated with trifluoroacetic acid (TFA). Upon acid treatment, the N-terminal
5 thiourea undergoes an energetically favorable intramolecular cyclization that results in the liberation of the first amino acid in a molecule containing the PITC atoms and the first amino acid of the protein. No reaction occurs with the thiourea adduct on the ends of lysine side chains because reaction of this adduct with adjacent peptide bonds is conformationally impossible. Thus, after TFA
10 treatment, a new protein is formed that is identical to the original protein except that its first amino acid is removed and it possesses only one amine group, i.e., the one at the N-terminus. The liberated product is separated by HPLC and its identity determined by comparing its mobility with standards containing each of the twenty amino acids. In order to determine the identity of the second amino
15 acid, this procedure is repeated, until the desired number of amino acids are identified.

The invention provides a method which leads to the production of a single peptide per protein in a starting sample, e.g., a biological mixture, resulting in the maximal possible simplification of a peptide mixture. The
20 method involves the incorporation of a moiety at the N-terminus of each protein in a sample using an N-terminal transferring moiety comprising a tag, optionally a linker, and a molecule which has an amine reactive group. Protein mixtures in which lysines are blocked and the N-terminus has an amine group are contacted with the N-terminal transferring moiety yielding a plurality of N-terminal
25 transferring moiety-modified proteins. As used herein, the term "tag" refers to a molecule which is detectable or capable of detection, and preferably which permits isolation of linked molecules, i.e., N-terminal transferring moiety-modified peptides. For instance, such a tag may include an affinity molecule, such as biotin or dinitrophenol, or include a solid support such as a resin or bead.

30 In one embodiment, the N-terminal transferring moiety contains atoms that permits its isotopic composition to be determined by mass spectrometry. Thus, the N-terminal transferring moiety contains atoms that may exist in a form in which all atoms are natural isotopes, or exist in a form in which specific atoms are non-natural atoms. A "non-natural isotope" as used herein refers to an

isotope that is not the predominant naturally-occurring isotope for its element, e.g., deuterium, ^{13}C or ^{15}N . These isotopic differences do not significantly affect physiochemical properties of the molecule, e.g., its mobility on HPLC, but are detectable by mass spectrometry. Detection could be, for instance, by mass difference, or by binding to a ligand, e.g., a biotin-containing moiety binding to avidin or streptavidin. In one embodiment of the invention, the N-terminal transferring moiety comprises a photocleavable linker, which allows N-terminal modified peptides to be selectively eluted from a support which binds the tag, e.g., from a solid support, with 300-350 nm illumination.

10 N-terminal transferring moiety-modified proteins are then hydrolysed either chemically or with a protease such as trypsin to generate a mixture of peptides including N-terminal transferring moiety-modified peptides and free peptides. Preferably, the proteins are cleaved at specific residues, e.g., by a specific protease such as trypsin, V8 protease, endoprotease Glu-C or
15 endoprotease Arg-C. In some cases, an N-terminal peptide generated by proteolysis with a particular protease, e.g., trypsinolysis, may be too small to provide definitive identifying information about the protein from which it was derived. Thus, to increase coverage of the proteome, two samples may each be contacted with a different protease, e.g., one sample is contacted with V8
20 protease, which cuts after Asp and Glu residues, and another sample with trypsin.

The N-terminal transferring moiety-modified peptides are then isolated from the free (unlabeled) peptides, e.g., by washing a support which binds the tag in the N-terminal transferring moiety, for instance, washing a bead or resin,
25 e.g., with organic solvents or salt containing buffers. In one embodiment, the isolated N-terminal transferring moiety-modified peptides are removed from the support by cleaving the linker, and the resulting N-terminal modified peptides which lack all or a part of the linker collected and separated by MS, and, optionally, identified, e.g., by sequence analysis. As used herein "identify"
30 refers to characterizing a peptide by its molecular weight, amino acid composition and/or amino acid sequence. The identified peptides may then be correlated to a protein which was present in the sample. In one embodiment, prior to MS, the labeled peptides are fractionated. Fractionation can be by, for instance, any type of chromatography or electrophoresis, including but not

limited to reverse phase HPLC, ion exchange chromatography, capillary electrophoresis, or a combination thereof.

The ability to label the N-terminus of proteins thus provides a straightforward way to perform proteome analysis. Thus, the method of the invention overcomes the limitations of current proteomic approaches and can be employed with MS-based approaches for highly complex mammalian proteomes. The method of the invention can be used to profile protein expression in any sample, or a plurality of samples, to identify protein changes associated with a particular disease or condition, or associated with treatment with an exogenous agent(s). Moreover, the present method can be used with whole tissue or physiological fluid to characterize, at a systems level, protein and to identify prognostic markers for particular conditions including various human or animal diseases.

The methods of the invention are particularly useful to compare the expression levels of large numbers of proteins derived from two different samples, e.g., two different cellular sources or tissues, for example, samples differentially exposed to an agent, using electrospray ionization mass spectrometry. In order to perform comparative analysis, two or more samples are distinguished based on the corresponding N-terminus of the proteins in each sample. For instance, the N-terminal modification for peptides in for one sample comprises a non-natural isotope, e.g., a deuterated tag, ^{13}C or ^{15}N , while the N-terminal modification for peptides in the other sample does not comprise the non-natural isotope, i.e., ^1H , ^{12}C or ^{14}N .

The methods of the invention are also particularly useful to determine if protease activity differs in two samples. In a sample with increased protease activity relative to another sample, there are more N-termini that are derived from the proteins that are generated by proteolytic cleavage of a precursor protein. The N-terminal peptides derived from proteins that are produced following proteolytic cleavage are more abundant in samples with increased proteolytic activity. Samples which differ in proteolytic activity may be due to a disease process, treatment of one sample or tissue with an agent that binds to a cellular component, such as a receptor or a cellular signaling molecule, or treatment of a cell or tissue with a protease inhibitor. Proteolytic activity may also differ as a result of introduction of a protease to one sample. For example, a

recombinant or a purified protease can be added to a tissue lysate. Alternatively, a protease can be introduced in a sample by inducing the expression of such a protease as a result of introduction of a nucleotide sequence that results in the cellular expression of the desired protease. In order to perform comparative
5 analysis, two or more samples are distinguished based on the corresponding N-terminus of the proteins in each sample. For instance, the N-terminal label transferring moiety for one sample comprises a non-natural isotope, e.g., a deuterated tag, e.g., ^{13}C or ^{15}N , while the label for the other sample does not comprise the non-natural isotope, e.g., ^1H , ^{12}C or ^{14}N .

10 In one embodiment of the invention, the method is employed to detect proteins whose expression level is up- or down-regulated in comparison to control environments. Samples of proteins from control and experimental conditions (e.g., from cells contacted with a drug or from physiological fluid of a patient with a particular disease) are obtained. Proteins in one sample are N-terminally modified with a moiety which does not comprise a non-natural
15 isotope. Proteins in the other sample are N-terminally modified with a moiety comprising one or more non-natural isotopes. In one embodiment, the two samples are mixed prior to hydrolysis. After hydrolysis, e.g., with a protease, the N-terminal modified peptides, obtained by separating the N-terminal
20 modified peptides from free peptides, are purified, and the purified peptides are subjected to analysis by MS. Preferably, before analysis by MS, the labeled peptides are fractionated. Fractionation can be by, for instance, any type of chromatography or electrophoresis, or any combination thereof. Most commonly, fractionation is accomplished by reverse phase HPLC, ion exchange
25 chromatography, or capillary electrophoresis. Ideally, the majority of N-terminally modified peptides in one sample are chemically equivalent but isotopically distinct from the labeled peptides in the other sample. In this context, chemical equivalence is defined by substantially identical chromatographic or electrophoretic behavior during the fractionation step.
30 Corresponding N-terminal modified peptides of the two samples will differ in mass by the difference in mass of the labels for the two samples. The relative amounts of the corresponding peptides can be determined by the ratio of their peak height or area in MS. For example, most pairs of corresponding peptides will have the same ratio of peak height or area in MS. Those with a different

ratio, it can be concluded, are derived from proteins that are up- or down-regulated in the experimental sample relative to the control.

In one embodiment, a resin is included in a N-terminal transferring moiety (Figure 2). In one embodiment, the resin is linked to a *N*-hydroxysuccinimide (NHS)-containing amine reactive leucine via a
5 photocleavable (PC) linker. Protein from a sample is treated with methylisothiocyanate (MITC) or PITC or other isothiocyanate, resulting in complete blockade of all amine groups. The protein is precipitated, dried, and resuspended in a reagent, such as trifluoroacetic acid or hydrochloric acid, to
10 cleave the first amino acid, leaving a free alpha amine group at the N-terminus. Proteins having amines at their N-terminus and modified lysines are incubated with the N-terminal transferring moiety so as to covalently couple the proteins to the N-terminal transferring moiety via the amine reactive group in the N-terminal transferring moiety. In one embodiment, the N-terminal transferring
15 moiety contains an affinity marker, such as biotin. In this embodiment, the N-terminal transferring moiety-modified proteins are hydrolyzed and the peptides are attached to a solid support by an affinity interaction of the N-terminal transferring moiety-modified proteins with avidin agarose or a related support. In another embodiment, the N-terminal transferring moiety contains a solid
20 support such that modification of the N-terminus results in simultaneous covalent immobilization to a solid support. In this embodiment, the resin-bound protein is then subjected to trypsinolysis and only the N-terminal peptides remain bound to the resin. After washing the resin, the peptides are eluted. For example, after washing the resin to remove nonspecifically bound peptides, N-terminal peptides may be modified such that they have leucine as the new N-
25 terminal residue. These peptides are photoeluted with 315 nm UV light, and separated by LC-ESI-MS for quantitation and automated MS/MS for identification of proteins (Figure 4). For comparison of two samples, a PC-Leu-NHS resin with leucine composed of all ^{12}C atoms, and another with leucine
30 composed of six ^{13}C s is prepared. A peptide coupled to the latter resin has a 6 Da molecular weight shift compared to the corresponding peptide that is eluted off of the former resin.

Alternatively, if the proteins are coupled to a N-terminal transferring moiety containing an affinity marker, after cleavage of the labeled proteins with

a protease or a chemical that causes cleavage of proteins after specific residues, the N-terminal modified peptides can be isolated from the other peptides by contacting the mixture with an immobilized ligand for the affinity marker, such as the ligand avidin for the affinity marker biotin, and washing the unlabeled peptides from the immobilized ligand-affinity marker-N-terminal peptide complex. For example, if the immobilized ligand is monomeric avidin, the peptide may be eluted under acidic conditions, e.g., 0.1% TFA, 0.1% acetic acid or the like. If the affinity marker is desthiobiotin, the immobilized ligand may be avidin, neutravidin, or streptavidin, and the peptide may be eluted under acidic conditions, e.g., 0.1% TFA, 0.1% acetic acid or the like, or may be eluted by affinity elution, such as in buffers containing biotin, e.g., 2 mM biotin, or may be eluted in high salt-containing buffers such as 400 mM sodium chloride. In another embodiment, a linker may be included between an affinity marker and a molecule with an isotopic label which is covalently coupled to the N-terminal amine. For instance, such a linker may be cleavable by light or specific chemicals, e.g., a linker may contain a disulfide linkage which is cleavable by reducing agents including but not limited to 2-mercaptoethanol or dithiothreitol.

The invention will be further illustrated with the following non-limiting examples.

20

Example I

Synthesis of a Photocleavable-Leu-NHS Resin

Photocleavable (PC) leucine-N-hydroxysuccinimide (PC-Leu-NHS) resin has an NHS-activated leucine ester and a photolabile group fixed on a bead support. The resin reacts with nucleophilic molecules such as primary amines on proteins. The resin was synthesized as follows (see Figure 4).

25

Step I: Formation of PC-NHS Carbonate Resin

1. 21.6 μ L (156 μ mole) of neat (7.21 M) triethylamine was mixed with 5 mL of dimethylformamide (0.0312 M TEA/DMF).
2. 0.2 g (50 μ mole substitution) of PC-OH Resin (Nova Biochem, 01-64-0118) was suspended in 2 mL of TEA/DMF solution.
3. 133.2 mg (520 μ mole) of disuccinimidyl carbonate (DSC) was dissolved in 3 mL of TEA/DMF solution (0.52 M DSC in 0.0312 M TEA/DMF).

30

4. 3 mL of DSC solution was added to 2 mL of PC-OH Resin suspension.
5. The mixture was incubated at 37°C for 3 hours with constant agitation.
- 5 6. The resin was gently spun down.
7. The resin was washed with 3 mL of 0.0312 M TEA/DMF solution three times.
8. The resin was resuspended in 2 mL of 0.0312 M TEA/DMF.
9. 3 mL of fresh 0.173 M DSC in 0.0312 M TEA/DMF was added to
10 the resin.
10. Steps 5 through 9 were repeated three times.
11. After three 3 hour incubations, the resin was washed with 0.0312 M TEA/DMF three times, followed by replenishing of the DSC solution, and overnight incubation at 37°C with constant agitation.
- 15 12. The resin was washed with 5 mL of DMF three times and then vacuum filtered.

Step II: Coupling of Leu to PC-Carbonate-NHS Resin

1. To 0.2 g (50 μ mole substitution) of PC-carbonate-NHS resin, 3 mL of
20 DMF was added to swell the resin.
2. 36.7 mg (200 μ mole) of Leu-O-Me HCl was dissolved in 2 mL of DMF.
3. 209.8 μ L (1.2 mmole) of neat diisopropylethyamine (DIEA) (5.72 M) was added to the Leu-O-Me solution.
4. The resin mixture was incubated at 25°C for 48 hours on a rotator.
- 25 5. 125 μ L of 2 M Et-NH₂/THF (Aldrich, 39,507-2) (250 μ mole) was added to the resin.
6. The resin mixture was incubated at 37°C for 2 hours.
7. The resin was washed with 5 mL of DMF three times.
8. The resin was vacuum filtered.
- 30 9. The resin was resuspended in 1 mL of DMF.
10. 3 mL of MeOH and 1 mL of 10 M NaOH (1 mmole) was added to the resin.
11. The resin was incubated at 25°C for 2 hours.
12. The resin was washed with 5 mL of MeOH three times.

13. The resin was washed with 5 mL of ddH₂O three times.
14. The resin was washed with 5 mL of DMF three times.
15. The resin was then vacuum filtered.

5 Step III: Formation of PC-Leu-NHS Ester Resin

1. 0.2 mg (52 μ mole substitution) of PC-Leu Resin was resuspended in 5 mL of 0.0312 M TEA/DMF.
2. 82.242 μ L (520 μ mole) of neat diisopropylcarbodiimide was added to the resin suspension.
- 10 3. The resin was incubated at 25°C for 15 minutes on a rotator.
4. 240 mg (2 mmole) of NHS (N-hydroxysuccinimide) was dissolved in 760 μ L of DMF. The final volume was adjusted to 1 mL by adding 80 μ L of DMF (final concentration = 2 M).
5. 40 mg of 1-hydroxybenzotriazole (HOBt) was dissolved in 296 μ L of
15 DMF (final concentration = 1 M).
6. 78 μ L (78 μ mole) of 1 M HOBt/DMF was added to the resin.
7. 390 μ L (780 μ mole) of 2 M NHS/DMF was added to the resin.
8. The reaction vessel was flushed with nitrogen.
9. The resin mixture was incubated 37°C for 3 hours with constant
20 agitation.
10. The resin was washed with 5 mL of DMF three times.
11. The incubation was repeated with fresh reagents three times.
12. After three 3 hour incubations, the resin was incubated with fresh reagents overnight 37°C with constant agitation.
- 25 13. The resin was washed with 5 mL of DMF three times and then vacuum filtered.

In order to make a "heavy" Leu resin, ¹³C₆-leucine is used in place of regular leucine.

Example II

Mass Spectrometry of N-Terminally Modified Peptides

Step I: Generate Proteins with a Free N-Terminal Amine and Protected Side Chain Amines

- 5 1. Denature 5 mg of protein mixture (either a single protein, a defined mixture of proteins, or a biologically derived mixture of proteins) in 1 ml of 50 mM bicine pH 9.5, 2% sodium dodecylsulfate (SDS), 2.5 mM dithiothreitol (DTT), and incubate for 20 minutes at 50°C to reduce disulfides, yielding denatured and reduced proteins.
- 10 2. Dilute protein sample with 50 mM bicine by 10-fold.
3. Add 400 µl of neat (9 M) phenylisothiocyanate (PITC) to 2.1 ml of DMSO and vortex to dissolve.
4. Add 2.5 ml of PITC/DMSO solution to 10 mL of a diluted protein sample to block all primary amines, i.e., amines on lysines and the N-terminal amine.
- 15 5. Incubate at 40°C with constant shaking overnight.
6. Add 2.5 volumes of acetone and 1 volume of diethyl ether to precipitate proteins and to remove residual PITC.
7. Incubate at -20°C for 20 minutes.
- 20 8. Centrifuge at 4°C for 15 minutes and wash the precipitated protein 3 times with cold acetone.
9. Dry pellet in SpeedVac for 20 minutes.
10. To the pellet, add 500 µL of anhydrous TFA (protein sequencing grade) to generate an N-terminal amine (Edman cleavage).
- 25 11. Incubate at 25°C for 60 minutes.
12. SpeedVac dry at 4°C.
13. Add 1 mL of acetone and wash 3 times.
14. Dry the pellet.
15. Resuspend the dried sample in 1 mL of 2% SDS/50 mM bicine pH 9.5.

30

Step II: Couple Protein to Photocleavable Resin Using the Newly-Generated N-Terminal Amine

1. Add protein sample to PC-leucine-NHS resin (approximately 1 mg of protein in starting mixture to 20 mg of resin). Chose resin which has either a "light" or "heavy" leucine.
2. Incubate overnight at 37°C with shaking
3. Transfer resin to a plastic column containing a frit, and wash the resin as follows to remove unbound protein: 5 x 5 ml buffer 1 (50 mM Tris HCl, pH 7.4, 1 M NaCl, 1 mM EDTA, 2% SDS); 5 x 5 ml buffer 2 (50 mM Tris HCl, 6 M guanidinium hydrochloride); 5 x 5 ml buffer 3 (50 mM acetic acid); 5 x 5 ml buffer 4 (50 mM HEPES, pH 7.0).

Step III: Trypsin Step: Cleavage Of Protein, Leaving N-Terminal Peptide Bound to Resin

1. Prepare 1 mg/ml solution of trypsin in 50 mM HEPES and add DMF to 10%.
2. Add to trypsin solution resin and incubate at 37°C for 24 hours.
3. Wash resin as above to remove unbound peptides.

Step IV: Obtain N-Terminally Modified Peptide

1. Incubate resin in 20 mM Hepes pH 7.0, 100 mM NaCl, 1 mM EDTA, and 20% DMF.
2. Place resin in a UV transparent cuvet and irradiate resin with 365 nm UV light from a distance of 5 cm for 30 minutes.
3. Gently spin down resin in a centrifuge.
4. Collect supernatant which contain N-terminal peptides.

Step V: Mass Spectrometry and Identification of Proteins

1. Obtain two protein samples labeled on their N-terminus. One sample has peptides labeled with light Leu and the other sample has peptides labeled with heavy Leu.
2. Subject peptide mixtures to MS. Optionally, separate peptides by reverse phase chromatography (RPLC) or capillary electrophoresis (CE) prior to

MS. Additional chromatographic separations prior to RP-LC or CE can be done, e.g., anion-exchange chromatography.

3. Doublets seen by MS likely represent N-terminal peptides from heavy and light Leu-labeled peptides if the doublet is separated by 6 atomic mass units (the difference in weight of heavy and light Leu).
4. Determine the sequence of the peptide by MS/MS of either peak. Manual or computer-assisted analysis of the MS/MS spectrum allows determination of the amino acid sequence of the peptide. Comparison of the relative peak intensities of the two peaks provides information on the relative abundance of the two peptides, and thus the protein from which the peptides were derived.

Example III

N-Terminal Modified Peptide Recovery, and Mass Spectrometer of an Exemplary Protein (BSA)

Step I: Generate Protein with a Free N-Terminal Amine and Protected Side Chain Amines

1. Denature 5 mg of bovine serum albumin (BSA) in 1 ml of 50 mM bicine pH 9.5, 2% sodium dodecylsulfate (SDS), 2.5 mM dithiothreitol (DTT), and incubate for 20 minutes at 50°C to reduce disulfides, yielding denatured and reduced proteins.
2. Dilute protein sample with 50 mM bicine by 10-fold.
3. Add 2.5 ml of 1 M MITC solution to 10 mL of a diluted protein sample to block all primary amines, i.e., amines on lysines and the N-terminal amine.
4. Incubate at 40°C with constant shaking for 16 hours.
5. Add 2.5 volumes of acetone and 1 volume of diethyl ether to precipitate protein and to remove residual MITC.
6. Incubate at -20°C for 20 minutes.
7. Centrifuge at 4°C for 15 minutes and wash the precipitated protein 3 times with cold acetone.
8. Dry pellet in by rotary evaporation for 20 minutes.
9. To the pellet, add 500 µL of anhydrous TFA (protein sequencing grade) to generate an N-terminal amine (Edman cleavage).

10. Incubate at 25°C for 60 minutes.
11. SpeedVac dry at 4°C.
12. Add 1 mL of acetone and wash 3 times.
13. Dry the pellet.
- 5 14. Resuspend the dried sample in 1 mL of 2% SDS/50 mM bicine pH 9.5.
15. To verify that the amines of BSA are blocked, and then the N-terminal amine is recovered western blot each sample. To each sample, add 10 µl of 100 µM biotin-N-hydroxysuccinimide to label all amines in the protein. Perform SDS-polyacrylamide gel electrophoresis, and detect
- 10 biotin-modified BSA by immunoblotting with biotin-specific antibodies (Figure 5).

Step II: Couple Protein to Photocleavable Resin Using the Newly-Generated N-Terminal Amine

- 15 1. Add protein sample to PC-leucine-NHS resin (approximately 1 mg of protein in starting mixture to 20 mg of resin).
2. Incubate overnight at 37°C with shaking
3. Transfer resin to a plastic column containing a frit, and wash the resin as follows to remove unbound protein: 5 x 5 ml buffer 1 (50 mM Tris HCl, pH 7.4, 1 M NaCl, 1 mM EDTA, 2% SDS); 5 x 5 ml buffer 2 (50 mM
- 20 Tris HCl, 6 M guanidinium hydrochloride); 5 x 5 ml buffer 3 (50 mM acetic acid); 5 x 5 ml buffer 4 (50 mM HEPES, pH 7.0). The resin binds proteins that have an N-terminal amine and does not exhibit nonspecific binding (Figure 6).

25

Step III: Trypsin Step: Cleavage Of Protein, Leaving N-Terminal Peptide Bound to Resin

1. Prepare 1 mg/ml solution of trypsin in 50 mM HEPES and add DMF to 10%.
- 30 2. Add to trypsin solution resin and incubate at 37°C for 24 hours.
3. Wash resin as above to remove unbound peptides.

Step IV: Obtain N-Terminally Modified Peptide

1. Incubate resin in 20 mM Hepes pH 7.0, 100 mM NaCl, 1 mM EDTA, and 20% DMF.
2. Place resin in a UV transparent cuvet and irradiate resin with 365 nm UV light from a distance of 5 cm for 30 minutes.
3. Gently spin down resin in a centrifuge.
4. Collect supernatant which contain N-terminal peptides.

Step V: Mass Spectrometry and Identification of Proteins

1. Obtain two protein samples labeled on their N-terminus. One sample has peptides labeled with "light" Leu (having an abundant naturally-occurring isotope) and the other sample has peptides labeled with "heavy" Leu (having a non-naturally occurring isotope).
2. Subject peptide mixtures to MS. Optionally, separate peptides by reverse phase chromatography (RPLC) or capillary electrophoresis (CE) prior to MS. Additional chromatographic separations prior to RP-LC or CE can be done, e.g., anion-exchange chromatography.
3. Doublets seen by MS likely represent N-terminal peptides from heavy and light Leu-labeled peptides if the doublet is separated by 6 atomic mass units (the difference in weight of heavy and light Leu).
4. Determine the sequence of the peptide by MS/MS of either peak. Manual or computer-assisted analysis of the MS/MS spectrum allows determination of the amino acid sequence of the peptide. Comparison of the relative peak intensities of the two peaks provides information on the relative abundance of the two peptides, and thus the protein from which the peptides were derived.

Example IVComparison of ICAT and N-Terminally Modified Peptide Method

- A set of ten commercially available proteins are selected which contain unmodified N-termini by Edman degradation (U. Texas, Galveston, Peptide Core facility). Two samples are prepared, each containing known quantities of each of the ten proteins. Samples are subjected to the ICAT method and the method described herein, and peptides are recovered and analyzed on a

Micromass Quatro-II triple quadrupole LC-ESI-MS at the Weill-Cornell MS Core facility. The relative quantity of each protein in the two samples is calculated by comparing peak heights in the doublets, and peaks are identified by MS/MS. For instance, MS/MS spectra are analyzed by computer algorithms and a number of potential matches, with percent likelihoods, are generated. More definitive identification of proteins is possible using the method described herein because protein matches are powerfully constrained by the requirement that the predicted match must derive from the N-terminus of the protein.

The ICAT reagent and software for data analysis are available in a kit from Applied Biosystems. Software for the analysis of ICAT data is compatible with the data obtained by the method described herein, and in particular, the expected mass shifts due to Lys and N-terminal modifications. Other software for ICAT and MS/MS data, in particular MASCOT (Matrix Science), Xcaliber and Bioworks 3.1 (Finnegan) are available. MASCOT is able to incorporate molecular weight differences due to modification of Cys residues by the ICAT reagent or Lys modification by MITC as well as replacement of the N-terminal residue by Leu.

Example V

Altering Method Conditions, Resin Testing, and Use of Leu or Nicotinic Acid as Amine Reactive Group

To optimize the labeling of diverse proteins on the N-terminus, some proteins may require longer incubation times with MITC or may have altered behavior on the resin that immobilizes proteins for MITC and TFA treatment. Other proteins may also require acetonitrile washes for complete elution from the resin.

Extraneous signals may derive from incomplete washing of the resins or incomplete amine blockade and subsequent coupling via Lys residues. Washing and blocking steps can be optimized to maximally reduce these extraneous signals. Alternatively, low signals may occur due to nonspecific losses of peptides on resin, plastic, and other surfaces. Certain polymeric carriers, e.g., linear polyacrylamide, polyethylene glycol, and polyvinylpyrrolidone, minimize this phenomenon. The yield of N-terminal amines may also require optimization of the conditions of the TFA step. Solvent conditions may also affect the yield of this cyclization reaction.

The N-terminus of the N-terminal peptides retains an amine group, and in one embodiment the N-terminal residue is a leucine, which is connected to the rest of the peptide through a standard peptide bond. Thus, the isotopic coding is performed with chemical bonds and moieties naturally found in peptides. Additionally, like most tryptic peptides, N-terminal peptides will have a +2 charge under ESI conditions. The only difference between a native peptide and these peptides is that lysine residues will be modified with a methyl thiocarbamyl adduct. If trypsin is employed to hydrolyze the N-terminal transferring moiety-modified proteins, all peptides will likely have a C-terminal Arg residue.

10 In one embodiment, peptides from two parallel samples are labeled on their N-terminus with either $^{12}\text{C}_6\text{-Leu}$ or $^{13}\text{C}_6\text{-Leu}$. For accurate quantitation, peptides that only differ in the isotopic composition of their isotopically-labeled Leu elute from the reverse phase column at the same time. If they do not, i.e., if one isotopic form elutes before another, then they cannot be subjected to simultaneous MS.

15 Resins should be prepared side-by-side to ensure the conditions for preparing each batch of heavy and light resins is identical. $^{13}\text{C}_6\text{-Leu-OMe}$ was synthesized from $^{13}\text{C}_6\text{-Leu}$. To ensure equal coupling, $^{12}\text{C}_6\text{-Leu-OMe}$ is prepared in parallel with $^{13}\text{C}_6\text{-Leu-OMe}$, and the synthesized Leu-OMe derivatives are coupled rather than commercial $^{12}\text{C}_6\text{-LeuOMe}$. To confirm equal coupling, Leu is directly photoeluted, and measured
20 by a fluorescamine assay. After every synthesis, Trp-Trp dipeptide is coupled to the two resins under identical coupling conditions. After coupling, Leu-Trp-Trp tripeptide is photoeluted and the purity is measured by HPLC.

To determine the accuracy of the MS-based quantitation, parallel protein samples are prepared, each containing known amounts of each of ten proteins. The photoeluted N-terminal peptides are subjected to LC-MS to quantitate the ratio of peak intensities and these values compare with the known concentration ratios in the starting samples.

In one embodiment, isotopically labeled nicotinic acid-NHS may be used to label proteins (Munchbach et al., 2000). In this method, instead of using PC-Leu-NHS resin, PC-NHS resin is employed directly. After trypsinization and photoelution, only
30 N-terminal peptides are eluted, with no Leu added to the N-terminus. These peptides are labeled with nicotinic-NHS, a reagent which can be prepared readily in light and heavy forms (Munchbach et al., 2000). Since the peptide has a C-terminal Arg and no primary amine anywhere except the N-terminus (any Lys residues would have a methyl

thiocarbamyl adduct), nicotinic acid-NHS labeling results in exclusive N-terminal labeling. After labeling, the peptides are mixed and desalted of unreacted nicotinic-NHS with a Zip tip, and then subjected to MS. This method retains the positive charge on the N-terminus (since nicotinic acid is a positively charged, pyridine-based
5 compound), as well as its C-terminus, thus ensuring a +2 charge on the peptides.

Example VI

Breast Cell Proteome and Secretome Analyses

The sensitivity of the present method in the context of a mammalian
10 cellular lysate is determined. Cellular lysates from MCF-7, MDA-MB-468, T47D, MDA-231, MDA-435, HS-578T, ZR-75 or BT-474 breast cancer cells, and MCF-10A or HBL-100 cells, which are normal breast cell lines, are subjected to the present method, N-terminal modified peptides are then quantitated by LC-MS, sequenced by MS/MS, and proteins having those
15 peptides, e.g., those preferentially up- or down-regulated in cancer cells, identified.

To identify proteins that are secreted by breast cancer cells, the secretome of breast cancer cell lines as well as normal breast cell lines is identified. Four breast cancer cell lines, MCF-7, MDA-MB-468, T47D, and BT-
20 474 cells, two of which are ER positive and two of which are EGF-R positive, to reflect some of the heterogeneity of the molecular pathogenesis of breast cancer, are employed. In addition, two normal breast cell lines, MCF-10A and HBL-100, are used as controls. Upon confluency, cells are transferred to serum-free media and after 48 hours, the media is filtered and precipitated to recover
25 secreted proteins. Typically 100 µg of protein is needed, which requires at least 20 L of conditioned media, based on previous studies in other cell lines (Miyazaki et al., 1993). The proteins are subjected to the present method, using a control cell line with the light PC-Leu-NHS resin and a breast cancer cell line with the heavy PC-Leu-NHS resin. Proteins whose levels are consistently
30 upregulated compared to normal breast cell lines are considered candidate breast cancer biomarkers. Antibodies are purchased or prepared against these proteins, and a panel of breast cancer cell lines is screened to determine the degree to which these secretion abnormalities exist in different cell lines.

To identify proteins that may serve as biomarkers for breast cancer, breast samples are obtained. Ninety percent of breast cancers are derived from duct cells. Duct cells secrete proteins into the lumen to form duct fluid. In breast cancer, as well as many other types of cancers, exocytic activity (Vega-Salas et al., 1993) and the activity of extracellular proteases (Rocheffort et al., 2001; Duffy et al., 2000) is known to be altered. Thus, one manifestation of cancer may be abnormal protein content in the extracellular fluid. Nipple aspirate fluid, which can be obtained by a modified breast pump, is particularly promising as a fluid that may contain these biomarkers, since its content derives wholly from duct cells. The secretion of nipple aspirate fluid from breast cancer patients is compared to that of control patients using the present method to identify proteins, the expression of which is altered in breast cancer patients.

In summary, the reagents described herein can improve the results of mass spectrometry analysis. Moreover, the present method accelerates the field of proteomics and can have an important impact on the identification of biomarkers that may be useful for the early detection of cancers, including breast cancer.

Example VII

Exemplary Uses of the Methods and Reagents

Cellular substrates of caspase-3 are identified. 10C9 cells are treated with Fas agonist antibody to activate caspase-dependent apoptosis and caspase-3. Caspase-3 activation in these cells is confirmed using DEVD-AMC, a fluorogenic caspase-3 substrate. The N-terminal peptide profile in cells treated with the caspase-3 inhibitor zDEVD-fmk or vehicle is then compared. Cells are lysed, clarified, and then loaded onto a C18 resin. After blocking Lys residues and the N-terminus with MITC, the completeness of amine blockade is assessed using biotin-NHS.

Next, the N-terminus of each protein is liberated with TFA. Generation of amines is confirmed by confirming that the released protein reacts with biotin-NHS. The eluted proteins are coupled to the amine-reactive PC-Leu-NHS resin. Samples from the control cells are labeled with $^{12}\text{C}_6$ -Leu, while the samples from the cells treated with inhibitor are labeled with $^{13}\text{C}_6$ -Leu. After photoelution, the peptides are directly subjected to LC-MS and MS/MS to identify N-terminal peptides.

ESI spectra is obtained on an Agilent Nanoflow Proteomics XCT ion trap mass spectrometer. Data collection are performed in the data-dependent scan mode. The +2 state is defined as the default charge state. A narrow range scan and a MS/MS scan is performed automatically on ion precursors that match prespecified intensity thresholds and isotope ratios. Both the light and heavy peptides are trapped and fragmented individually. The use of an isotopically labeled N-terminal amino acid (i.e., Leu) will likely assist in assignment of the b-ion series (as well as the a- and c-ions), which are separated by a 6-Da mass shift in the two MS/MS spectra, while the y-ion series will be constant between the two. The use of terminal isotope tags for the assistance of ion series assignments has been described previously (Munchbach et al., 2000; Gu et al., 2002; Keough et al., 2000). Most peptides are present as a doublet, of equal intensity in both samples. In cases where there is a discrepancy in the height, a possible protease substrate is identified. In particular, if the heavy-peak is of lower abundance, this peak represents an internal N-terminal peptide. To identify these peptides, SpectrumMill (Millenium), with sequences queried against the NCBI nonredundant protein database using human sequence filtering, is employed.

The degree of peptide complexity simplification by recovering only N-terminal peptides may make 1D chromatography sufficient for analysis of large numbers of peptide doublets. Indeed, the number of N-terminal peptides is expected to be vastly smaller than the number for peptides obtained in the ICAT protocol, as many proteins contain numerous cysteines. For example, bovine serum albumin has 86 tryptic cleavage sites, and generates up to 35 ICAT peptides, but only one peptide using the present N-terminal peptide recovery strategy. However, the accuracy, sensitivity, and amount of peptide identification may be improved by multidimensional chromatography. Peptides are subjected to a strong cation exchange chromatography prior to reversed phase LC-MS as has been described previously (Tam et al., 2004). In these experiments, peptides are fractionated on a 500 μ m i.d. x 15 mm BioSCX column and eluted in 50 μ l fractions with step gradients of ammonium acetate. Eluted peptides are concentrated and desalted on a PepMap nanotrapping column before reversed phase LC-MS.

In the case of N-terminal peptides, a constraint that the peptide is derived from the N-terminus will improve the peptide assignment. This constraint will not apply to internal N-terminal peptides that derive from protease cleavage products. In these cases, the ability to identify these peptides depends on the ability to assign the sequence from the MS/MS spectra. Theoretically, the information in a single peptide should be

sufficient to identify the parent protein. Other methods could be used to improved sequence assignments. For example, mass spectrometers capable of ultrahigh mass accuracy (Li et al., 2001; Kosaka et al., 2000) are likely to improve the accuracy of the assignments.

- 5 Many of the proteins identified will likely include well-known caspase-3 substrates, such as FOXO3a, CDB3, and BAT3. For proteins that are not known to be caspase-3 substrates, these are subjected to cell-based assays for verification.

References

- 10 Boja et al., Anal. Chem., **73**, 3576 (2001).
 Duffy et al., Breast Cancer Research, **2**, 252 (2000).
 Gaston et al., Proc. Natl. Acad. Sci. U.S.A., **90**, 10957 (1993).
 Gu et al., Anal. Chem., **74**:5774 (2002).
 Gustavsson et al., J. Chromatogr., **A937**, 41 (2001).
15 Gygi et al. Nat. Biotechnol., **17**, 994 (1999).
 Hess et al., Nat. Cell Biol., **3**, E46 (2001).
 Jaffrey et al., Nat. Cell Biol., **3**, 193 (2001).
 Kenyon et al., Methods Enzymol., **47**, 407 (1977).
 Keough et al., Electrophoresis, **21**:2252 (2000).
20 Kosaka et al., Anal. Chem., **72**:1179 (2000).
 Kulbe et al., Anal. Biochem., **72**, 123 (1976).
 Laursen, J. Am. Chem. Soc., **88**, 5344 (1966).
 Li et al., Anal. Chem., **73**:3312 (2001).
 Munchbach et al., Anal. Chem., **72**:4047 (2000).
25 Miyazaki et al., Proceed. Natl. Acad. Sci. U.S.A., **90**, 11767 (1993).
 Rocheftort et al., J. Steroid Biochem. & Molecular Biol., **76**, 119 (2001).
 Singh et al., J. Biol. Chem., **271**, 18596 (1996).
 Smith et al., Nitric Oxide, **4**, 57 (2000).
 Stamler et al., Proc. Natl. Acad. Sci. U.S.A., **89**, 444 (1992).
30 Tam et al., PNAS, **101**:6917 (2004).
 Vega-Salas et al., Differentiation, **54**, 131 (1993).
 Zhang et al., Anal. Chem., **73**, 5142 (2001).

All publications, patents, and patent applications cited are incorporated herein by reference. While in the foregoing specification, this invention has been described in relation to certain preferred embodiments thereof, and many details have been set forth for illustration, it will be apparent to those skilled in
5 the art that the invention is susceptible to additional embodiments and that certain of the details herein may be varied considerably without departing from the basic principles of the invention.

WHAT IS CLAIMED IS:

1. A method of preparing a mixture of isolated N-terminal modified peptides, comprising:
 - 5 a) providing a hydrosylate comprising free peptides and N-terminal transferring moiety-modified peptides, wherein the hydrosylate is prepared by hydrolyzing a plurality of N-terminal transferring moiety-modified proteins obtained by contacting an N-terminal transferring moiety with a sample comprising a plurality of proteins comprising side chain-modified amines and a
10 free alpha amine group on the amino terminus, and wherein the N-terminal transferring moiety comprises a tag, a linker, and a molecule with an amine reactive group; and
b) isolating the N-terminal transferring moiety-modified peptides from the free peptides in the hydrosylate.
15
2. The method of claim 1 wherein the linker is cleavable.
3. The method of claim 2 wherein the cleavable linker is a photocleavable
20 linker.
4. The method of claim 2 wherein the cleavable linker is cleavable with an acid or a base.
5. The method of claim 2 further comprising cleaving the linker to generate
25 a mixture of purified N-terminal modified peptides.
6. The method of claim 1 wherein the molecule comprises a natural amino acid, a non-natural amino acid, a natural isotope, or a non-natural isotope, or any combination thereof.
30
7. The method of claim 6 wherein the molecule comprises leucine.
8. The method of claim 1 wherein the tag comprises biotin, desthiobiotin, or iminobiotin.

9. The method of claim 1 wherein the tag comprises a solid support.
10. The method of claim 1 wherein the plurality of proteins comprising side
5 chain-modified amines and a free alpha amine group are prepared by treating a plurality of proteins sequentially with an isothiocyanate and an acid.
11. The method of claim 11 wherein the isothiocyanate is phenylisothiocyanate.
- 10 12. The method of claim 11 wherein the acid is trifluoroacetic acid, hydrochloric acid or hydrofluoric acid.
13. The method of claim 1 wherein the plurality of N-terminal transferring
15 moiety-modified proteins are hydrolyzed with a protease.
14. The method of claim 13 wherein the protease is trypsin.
15. The method of claim 13 wherein the protease is chymotrypsin, pepsin,
20 papain, V8 protease, proteinase K, calpain, subtilisin, endoprotease Glu-C, or endoprotease Arg-C.
16. The method of claim 1 wherein the plurality of N-terminal transferring moiety-modified proteins are chemically hydrolyzed.
- 25 17. The method of claim 16 wherein the plurality of N-terminal transferring moiety-modified proteins are hydrolyzed with cyanogen bromide or 2-(2-nitrophenylsulfenyl)-3-methylindole.
- 30 18. The method of claim 10 wherein the plurality of proteins are denatured prior to treatment with isothiocyanate.
19. The method of claim 5 further comprising separating the purified N-terminal modified peptides by mass spectrometry.

20. The method of claim 19 further comprising, prior to separating the purified N-terminal modified peptides by mass spectrometry, fractionating the mixture of purified N-terminal modified peptides.
- 5
21. The method of claim 20 wherein liquid chromatography, ion-exchange chromatography, or capillary electrophoresis is employed to fractionate the mixture of purified N-terminal modified peptides.
- 10
22. The method of claim 19 further comprising identifying one or more of the separated N-terminal modified peptides.
23. The method of claim 1 wherein the sample is a cellular sample.
- 15
24. The method of claim 1 wherein the sample is a physiological fluid sample.
25. The method of claim 24 wherein the sample is serum, plasma, urine, cerebrospinal fluid or breast nipple aspirate fluid.
- 20
26. The method of claim 24 wherein the sample is a cellular cytosol fraction, cellular organelle fraction, a microsomal fraction, a nuclear fraction, a secreted fraction, or conditioned media.
- 25
27. Isolated N-terminal transferring moiety-modified peptides prepared by the method of claim 1.
28. Purified N-terminal modified peptides prepared by the method of claim 5.
- 30
29. Fractionated N-terminal modified peptides prepared by the method of claim 20.

30. A method of separating N-terminal modified peptides in a mixture comprising:
- a) providing a hydrolysate comprising free peptides and N-terminal transferring moiety-modified peptides, wherein the hydrolysate is prepared by hydrolyzing a plurality of N-terminal transferring moiety-modified proteins obtained by contacting an N-terminal transferring moiety with a sample comprising a plurality of proteins comprising side chain-modified amines and a free alpha amine group on the amino terminus, and wherein the N-terminal transferring moiety comprises a tag, a cleavable linker, and a molecule with an amine reactive group;
 - b) isolating the N-terminal transferring moiety-modified peptides from the free peptides in the hydrolysate; and
 - c) cleaving the linker to generate a mixture of purified N-terminal modified peptides; and
 - d) separating the purified N-terminal modified peptides by mass spectrometry.
31. The method of claim 30 further comprising identifying one or more of the separated peptides.
32. A method of separating peptides in a mixture, comprising:
- providing purified N-terminal modified peptides prepared by the method of claim 5; and
 - separating the purified N-terminal modified peptides by mass spectrometry.
33. The method of claim 32 further comprising identifying one or more of the separated peptides.
34. The method of claim 33 further comprising correlating the one or more identified peptides with one or more proteins to identify one or more proteins in the sample.

35. A method for identifying a protein in a sample comprising a plurality of proteins, comprising:

- a) providing a hydrolysate comprising free peptides and N-terminal transferring moiety-modified peptides, wherein the hydrolysate is prepared by hydrolyzing a plurality of N-terminal transferring moiety-modified proteins obtained by contacting an N-terminal transferring moiety with a sample comprising a plurality of proteins comprising side chain-modified amines and a free alpha amine group on the amino terminus, and wherein the N-terminal transferring moiety comprises a tag, a cleavable linker, and a molecule with an amine reactive group;
- b) isolating the N-terminal transferring moiety-modified peptides from the free peptides in the hydrolysate;
- c) cleaving the linker to generate a mixture of purified N-terminal modified peptides;
- d) separating the purified N-terminal modified peptides by mass spectrometry; and
- e) identifying at least one separated peptide, thereby identifying at least one protein in the sample.

36. A method for comparing the amount or level of one or more proteins in at least two samples, comprising:

- a) providing a first sample comprising a plurality of N-terminal transferring moiety-modified proteins and a second sample comprising a plurality of N-terminal transferring moiety-modified proteins, wherein the plurality of N-terminal transferring moiety-modified proteins in the first sample is prepared by contacting a first plurality of proteins comprising side chain-modified amines and a free alpha amine group on the amino terminus with a first N-terminal transferring moiety comprising a tag, a cleavable linker, and a molecule which comprises an amine reactive group, which first N-terminal transferring moiety comprises a natural isotope, wherein the plurality of N-terminal transferring moiety-modified proteins in the second sample is prepared by contacting a sample comprising a second plurality of proteins comprising side chain-modified amines and a free alpha amine group on the amino terminus with a second N-terminus transferring moiety comprising a tag, a cleavable linker and

a molecule which comprises an amine reactive group, which second N-terminal transferring moiety comprises a non-natural isotope;

b) hydrolyzing the plurality of N-terminal transferring moiety-modified proteins in the first and second samples;

5 c) isolating N-terminal transferring moiety-modified peptides from free peptides in the first and second samples;

d) cleaving the linker in the isolated N-terminal transferring moiety-modified peptides so as to yield purified N-terminal modified peptides;

10 e) separating the purified N-terminal modified peptides by mass spectrometry; and

f) comparing the amount or level of at least one separated peptide in the first sample to the amount or level of the corresponding separated peptide in the second sample.

15 37. A method for comparing the amount or level of one or more proteins in at least two samples, comprising:

a) providing a first sample comprising a plurality of N-terminal transferring moiety-modified proteins and a second sample comprising a plurality of N-terminal transferring moiety-modified proteins, wherein the
20 plurality of N-terminal transferring moiety-modified proteins in the first sample is prepared by contacting a first plurality of proteins comprising side chain-modified amines and a free alpha amine group on the amino terminus with a first N-terminal transferring moiety comprising a tag, a linker and a molecule which comprises an amine reactive group, which first N-terminal transferring moiety
25 comprises a natural isotope, wherein the plurality of N-terminal transferring moiety-modified proteins in the second sample is prepared by contacting a sample comprising a second plurality of proteins comprising side chain-modified amines and a free alpha amine group on the amino terminus, with a second N-terminal transferring moiety comprising a tag, a linker, and a molecule which
30 comprises an amine reactive group, which second N-terminal transferring moiety comprises a non-natural isotope;

b) hydrolyzing the plurality of N-terminal transferring moiety-modified proteins in the first and second samples;

- c) isolating N-terminal transferring moiety-modified peptides from free peptides in the first and second samples;
- d) treating the isolated N-terminal transferring moiety-modified peptides in the first and second samples with an agent so as to yield purified N-terminal modified peptides which lack the linker;
- 5 e) separating the purified N-terminal modified peptides by mass spectrometry; and
- f) comparing the amount or level of at least one separated peptide the first sample to the amount or level of the corresponding separated peptide in
- 10 the second sample.
38. The method of claim 36 or 37 wherein the samples are mixed together prior to step e).
- 15 39. The method of claim 36 or 37 wherein the molecule comprises a natural amino acid, a non-natural amino acid, a natural isotope, a non-natural isotope, or any combination thereof.
40. The method of claim 39 wherein the molecule comprises leucine.
- 20 41. The method of claim 36 or 37 wherein the tag comprises biotin, desthiobiotin, or iminobiotin or a solid support.
42. The method of claim 36 or 37 wherein the plurality of proteins comprising side chain-modified amines and a free alpha amine group are prepared by treating a plurality of proteins sequentially with isothiocyanate and an acid.
- 25 43. The method of claim 42 wherein the isothiocyanate is phenylisothiocyanate or methylisothiocyanate.
- 30 44. The method of claim 36 or 37 wherein the plurality of proteins are hydrolyzed with a protease.

45. The method of claim 36 or 37 wherein the plurality of proteins are chemically hydrolyzed.

46. A mixture comprising a plurality of proteins or peptides, each having at
5 their N-terminus an N-terminal transferring moiety comprising a tag, optionally a cleavable linker, and a molecule with an amine reactive group.

1/5

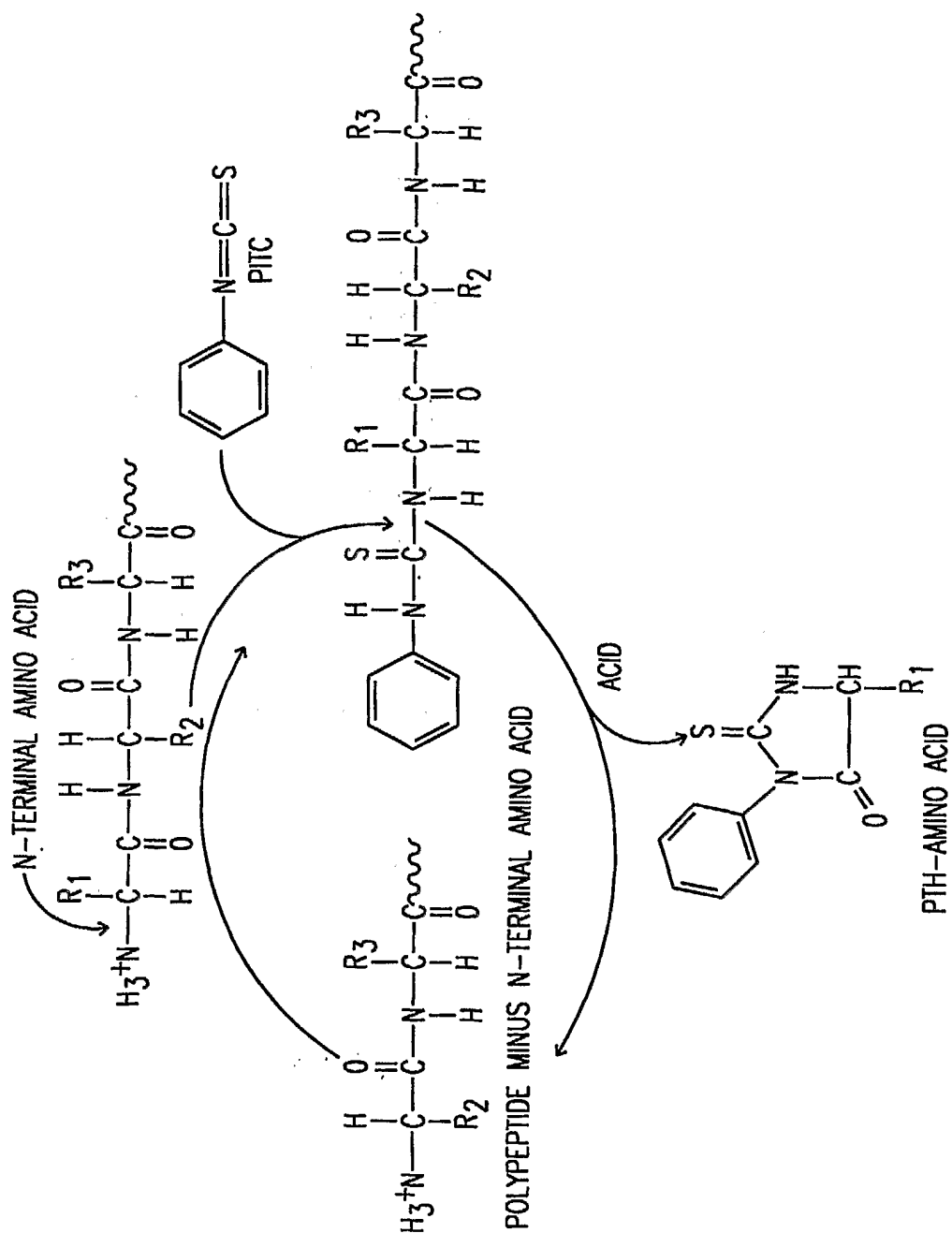


FIG. 1

2/5

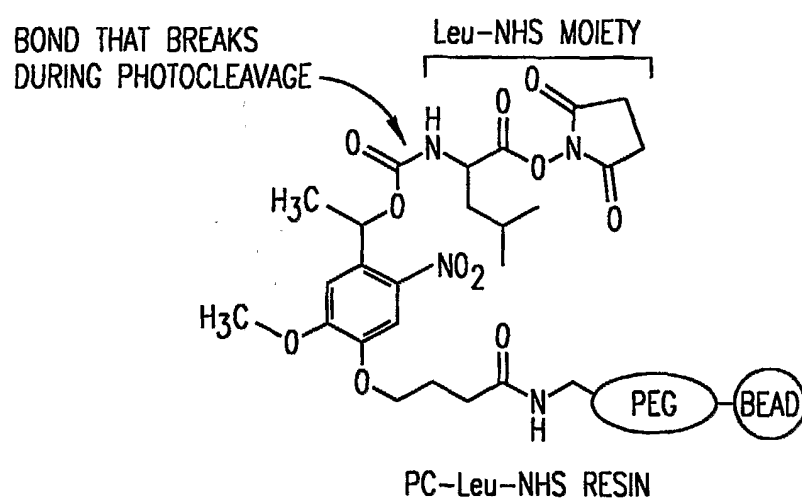
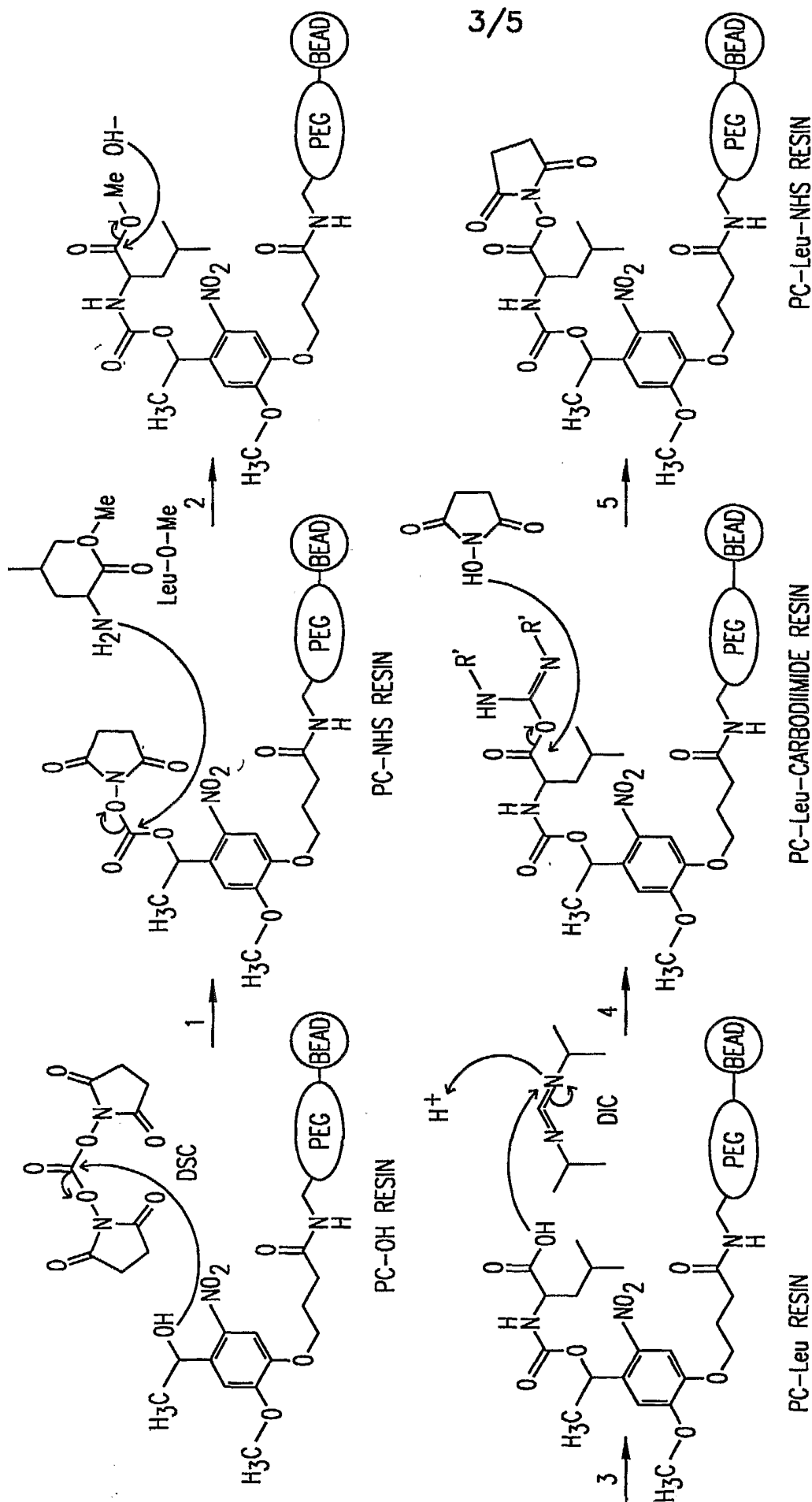


FIG. 2



1. FORMATION OF NHS-CARBONATE ON PC-OH RESIN
2. COUPLING OF LEUCINE METHYLESTER
3. DEPROTECTION OF -OH ON Leu
4. CARBODIMIDE MEDIATED CARBOXYLIC ACID ACTIVATION
5. FORMATION OF NHS ESTER

FIG. 3

4/5

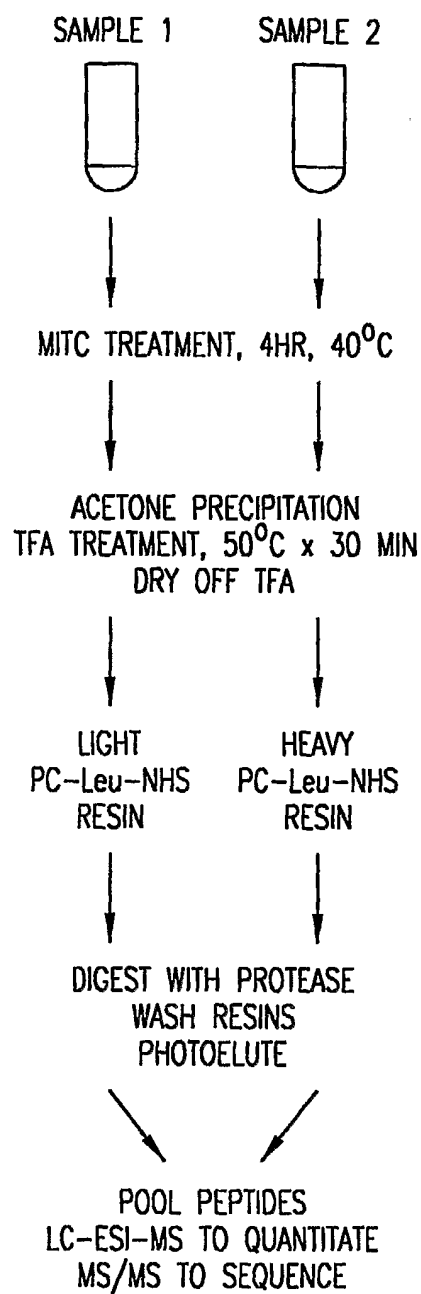


FIG. 4

5/5

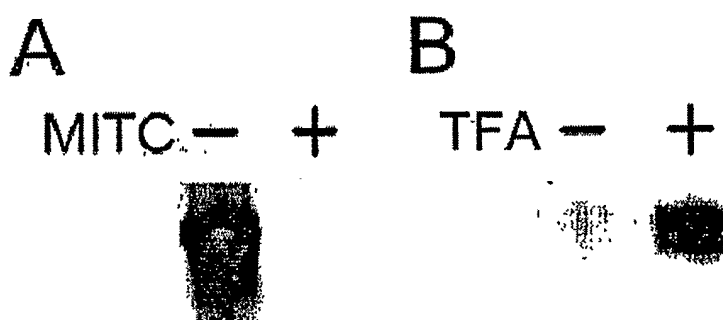


FIG. 5

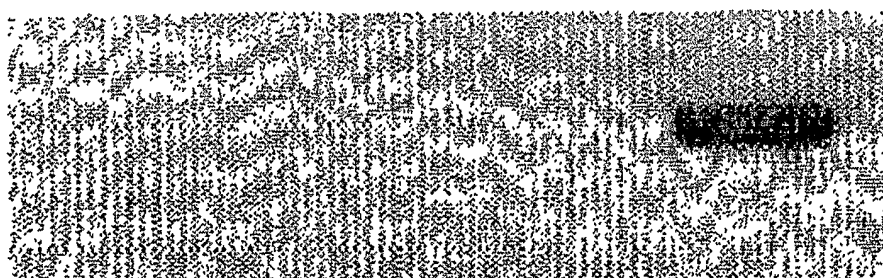


FIG. 6