

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
19 February 2004 (19.02.2004)

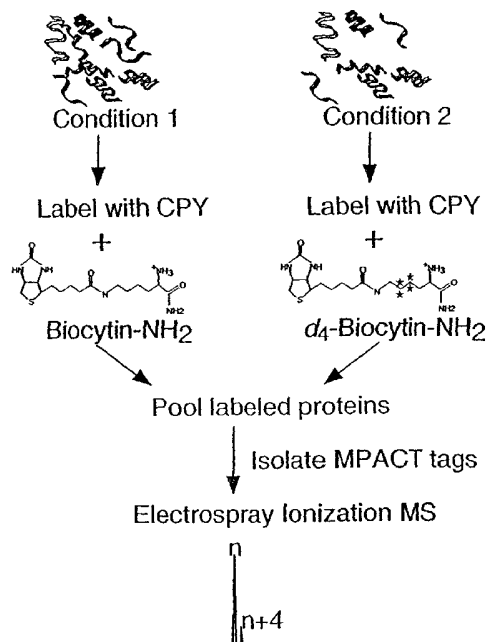
PCT

(10) International Publication Number
WO 2004/015391 A2

- (51) International Patent Classification⁷: **G01N**
- (21) International Application Number:
PCT/US2003/024936
- (22) International Filing Date: 7 August 2003 (07.08.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/402,774 12 August 2002 (12.08.2002) US
- (71) Applicant (for all designated States except US): **CORNELL RESEARCH FOUNDATION, INC.** [US/US]; 20 Thornwood Drive, Suite 105, Ithaca, NY 14850 (US).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): **JAFFREY, Samie, R.** [US/US]; 1675 York Avenue, Apartment 17M, New York, NY 10128 (US).
- (74) Agents: **STEFFEY, Charles, E.** et al.; P.O. Box 2938, Minneapolis, MN 55402 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, 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, 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 (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: MASS SPECTROMETRY-BASED IDENTIFICATION OF PROTEINS



Quantitative MPACT. The use of deuterated (depicted with stars) biotin-containing amino acids in the CPY-dependent labeling reaction results in peptide mixtures that have MPACT tags whose molecular mass is increased by the number of deuterium atoms in the biotin nucleophile. Identical peptides coelute and the relative amounts of each parent protein can be inferred by the relative signals produced by the two peptides.

(57) Abstract: The invention provides a method to C-terminally label proteins in a complex sample and identify those proteins, e.g., using mass spectrometry.



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.

MASS SPECTROMETRY-BASED IDENTIFICATION OF PROTEINS

Background of the Invention

A major goal of preventive medicine is to diagnose diseases in which
5 early intervention can alter the outcome of the disease. One example of such a
disease is pancreatic cancer, a neoplasm which kills 24 000 people in the United
States every year, and whose incidence has increased threefold in the last forty
years (Brand 2001). Pancreatic cancer is highly fatal with only 10% of patients
surviving one year past the time of diagnosis (Brand, 2001). An exception to
10 this high mortality rate for pancreatic cancer is among patients in whom the
cancer is detected early; in rare cases small carcinomas at the head of the
pancreas are associated with hepatobiliary obstruction and painless jaundice,
prompting radiographic studies that identify the cancer at an early, resectable
stage (Brand, 2001). Thus, early detection of pancreatic cancer can have a major
15 impact on the prognosis of the disease. Importantly, early detection of other
cancers, such as cervical cancer by the Pap smear, and prostate cancer, with the
PSA blood test, have had a substantial impact on the mortality of these diseases.
Thus, early detection is a key strategy in the fight against cancer.

However, early detection does not have an impact solely on cancer.
20 Other diseases, such as Parkinson's disease, can be ameliorated with early
intervention. Patients with Parkinson's disease show delayed progression
following treatment with monoamine oxidase inhibitors (Koller, 1997).
Conceivably, early intervention of Parkinson's disease with this class of drugs
can delay the onset of the characteristically debilitating tremor, muscle atrophy,
25 and neurodegeneration associated with this disease.

Ten years ago, a major international effort was initiated to sequence the
entire human genome, which amounts to 3.2 million basepairs. By mid-2000, a
draft of the human genome was announced, comprising a majority of the coding
sequences. With the completion of the sequencing of the genomes of yeast
30 (Mewes et al., 1997), *Caenorhabditis elegans* (The *C. elegans* Sequencing
Consortium, 1998), and *Drosophila* (Adams et al., 2001), and the anticipation of
the complete human genome being fully sequenced by 2003, there is no doubt

that this vast set of knowledge will have a substantial impact on the diagnosis and treatment of human disease.

In particular, the sequencing of the human genome provides data that can be used to identify markers for early stages of disease, e.g., neoplastic and cardiovascular diseases. Once the identity of every protein in the human proteome is known, the levels of each protein can be individually queried in a systematic manner, and compared from blood samples of patients with the disease, and unaffected patients. Epidemiological studies comparing large numbers of afflicted patients with unaffected patients can lead to the identification of proteins whose expression level is increased or decreased only in the case of disease. These proteins can then be detected in routine blood assays in the general population, much in the same way that Caucasian men over 50 years and African-American men over 40 years take a yearly PSA blood test to screen for prostate cancer.

A number of methods have been developed to take advantage of the newly generated genomic data, most notably "gene-chip" technology, that provide a way to simultaneously quantitate transcript levels of numerous genes. While this method is useful, for most experimental purposes, quantitation of protein levels is more relevant than quantitation of transcript levels. Indeed, in many well-documented cases, changes in mRNA levels do not correlate linearly with changes in protein levels (Gygi et al., 1999; Anderson et al., 1997); moreover, the levels of some proteins are regulated at the level of protein stability. For example, the cell cycle is crucially dependent on the stability of cyclins, which are regulated by cell-stage-specific ubiquitination. Additionally, some samples, such as serum, which are particularly useful because they are easily collected from patients for diagnostic purposes, are composed of proteins whose transcripts reside in a variety of diverse tissues. Unfortunately, technology to assess protein levels is remarkably underdeveloped compared to DNA technologies.

The principle technique to study protein levels involves subjecting different samples to two-dimensional gel electrophoresis (2DE) and comparing the pattern of spots that arise following staining. The identities of spots with

different intensities can be determined by excising the spot, trypsinizing the eluted protein, and subjecting the peptides to MALDI/MS. One major limitation of this approach is that only proteins that are excised from the gels are examined, leaving the identities of non-excised spots unknown. Another limitation is that many proteins, especially those which are highly charged or have transmembrane segments, resolve poorly on 2DE gels. Most importantly, since only a relatively small amount of protein can be loaded onto the gels, most proteins are present at levels below the level of detection by MALDI/MS.

Mass spectrometry (MS) is a technique that is capable of generating molecular weight data for peptides and other small molecules, and tandem mass spectrometry (MS/MS) is a technique that provides sequence information for peptides. Because full-length proteins do not produce highly accurate data by MS, samples are first digested with proteases, such as trypsin, and the resulting peptide mixture is subjected to MS and MS/MS. Using data from the completed genome sequencing projects, molecular masses and sequence information obtained by MS and MS/MS can be used to identify parent proteins that were present in the original samples.

Although an approach which employs MS is useful, the complexity of the peptide mixture increases as the number of proteins in the starting material is increased. For example, digestion of the yeast proteome, which comprises 6,000 proteins, results in 344,000 peptides. Peaks seen by MS analysis of these complex mixtures often cannot be interpreted because they represent numerous different peptides with similar molecular masses. Additionally, low abundance peptides are often invisible or lie within or adjacent to peaks of higher abundance. Typically, a liquid chromatography-electrospray mass spectrometer can resolve approximately 30,000 peptides (Gygi et al., 1999). Thus, because of the large number of peptides that would be generated, MS analysis does not appear to be feasible for higher eukaryotes using current technology.

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

Summary of the Invention

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 method involves the incorporation of a label comprising an affinity marker at the C-terminus of each protein in a sample using the reverse reaction of a transpeptidase, e.g., a yeast exopeptidase such as carboxypeptidase Y (CPY), which preferably is specific for the C-terminus. Accordingly, protein mixtures are contacted with a transpeptidase under conditions which favor transpeptidase-mediated ligation of a label comprising an affinity marker, e.g., biotin or biocytin-NH₂, to each protein resulting in C-terminal labeling of each protein. As used herein, the term "label" refers to a molecule comprising an affinity marker, which is detectable or capable of detection, however, the term "affinity marker" does not include a natural amino acid. In one embodiment of the invention, the label comprises a photocleavable affinity marker, such as photocleavable biotin, which allows labeled peptides to be selectively eluted from a ligand for the affinity marker, e.g., from avidin supports with 300-350 nm illumination, as illumination results in cleavage of the peptides from the ligand-bound affinity marker, resulting in elution of peptides that no longer have the affinity marker. In addition to the affinity marker, the label may further comprise an amino acid moiety or a derivative thereof (e.g., an amidated amino acid moiety, such as the amidated lysine moiety of biocytin amide), and/or a non-natural isotope. A "non-natural isotope" as used herein refers to an isotope that is not the predominant isotope for its element, e.g., deuterium, ¹³C or ¹⁵N.

C-terminally labeled proteins are then hydrolysed either chemically or with a protease such as trypsin to generate a mixture of peptides including C-terminally labeled peptides. Preferably, the proteins are cleaved at specific residues, e.g., by a specific protease such as trypsin, endoprotease Glu-C or endoprotease Arg-C. The C-terminal peptides are then isolated using a ligand for the affinity marker, e.g., streptavidin-agarose. In one embodiment, the isolated peptides are collected and separated by MS, and, optionally, identified, e.g., by sequence analysis. As used herein "identify" 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 C-terminus of proteins thus provides a straightforward way to perform proteome analysis and so 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. MPACT can be used to profile protein expression in any sample, e.g., in sarcoplasmic reticulum derived from normal and chronic heart failure (CHF) patients, or a plurality of samples to identify protein changes associated with a particular disease or condition, e.g., CHF and its reversal by LVAD therapy, or associated with treatment with an exogenous agent(s). Moreover, MPACT can be used with whole tissue, e.g., cardiac tissue samples, to characterize, at a systems level, protein and to identify prognostic markers for particular conditions. A combined multidimensional LC-MS/MS and MPACT approach can also be employed to profile protein expression in cardiac tissue to assess protein expression changes in vascular cells, hearts from animal models of septic shock and ischemia-reperfusion injury, as well as cardiac tissue from CHF patients before and after LVAD therapy.

Brief Description of the Drawings

Figure 1 depicts a HPLC analysis of Fmoc-Lys-NH₂.

Figure 2 depicts a HPLC analysis of Fmoc-PCB-NH₂ synthesis.

Figure 3 illustrates a HPLC analysis of the deprotection of Fmoc-PCB-NH₂.

Figure 4 is a schematic of mass spectrometric proteome analysis with C-terminal tags (MPACT). After C-terminal labeling of proteins, the labeled proteins are hydrolyzed and C-terminal peptides recovered, e.g., using

streptavidin-agarose. Biotin is depicted by “*”. Note that only one peptide is recovered per parent protein.

Figure 5. Biotinylation of proteins by CPY and biocytin-NH₂. A) Biotinylation of BSA-methyl ester was tested by incubation of CPY and various biocytin derivatives. Biotinylated BSA was detected by anti-biotin Western blot. Biocytin-NH₂ was a preferred substrate for biotinylation. B) BSA-ME is a substrate for CPY-dependent labeling. BSA prepared with methanol and HCl is methyl esterified, and this is the only preparation that is labeled by Western blot. C) The methyl ester is hydrolyzed by base for the indicated times prior to being subjected to CPY-dependent labeling. Hydrolysis of the ester reduced the ability of BSA to serve as a substrate. D) Biotinylation of various GST-amino acid fusion protein methyl esters. Biotinylation of the indicated GST-fusions was performed and protein were subjected to anti-biotin Western blotting. CPY was able to biotinylate each fusion protein, indicating that all proteins regardless of the C-terminal amino acid are likely to be substrates for CPY-mediated ligation.

Figure 6 illustrates a synthetic strategy for biotinamidoethyl dithiocysteineamide and its similarity to biocytin-NH₂.

Figure 7 shows quantitative MPACT. The use of deuterated (depicted with stars) biotin-containing amino acids in the CPY-dependent labeling reaction results in peptide mixtures that have MPACT tags whose molecular mass is increased by the number of deuterium atoms in the biotin nucleophile. Identical peptides coelute and the relative amounts of each parent protein can be inferred by the relative signals produced by the two peptides.

Detailed Description of the Invention

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 method involves the incorporation of a label comprising an affinity marker at the C-terminus of each protein in a sample using the reverse reaction of a transpeptidase, which preferably is specific for the C-terminus. Accordingly,

protein mixtures are contacted with a transpeptidase under conditions which favor transpeptidase-mediated ligation of a label comprising an affinity marker, e.g., biotin or biocytin-NH₂, to each protein resulting in C-terminal labeling of each protein. In addition to the affinity marker, the label may further comprise
5 an amino acid moiety or a derivative thereof (e.g., an amidated amino acid moiety, such as the amidated lysine moiety of biocytin amide), and/or a non-natural isotope.

C-terminally labeled proteins are then hydrolysed either chemically or with a protease such as trypsin to generate a mixture of peptides including C-terminally labeled peptides. The C-terminal peptides are then isolated using a
10 ligand for the affinity marker, e.g., streptavidin-agarose. In one embodiment, the isolated peptides are collected and separated by MS, and, optionally, identified, e.g., by sequence analysis. The identified peptides may then be correlated to a protein which was present in the sample. In one embodiment, prior to MS, the
15 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.

Isotope-coded affinity tags (ICAT) reagents, which are used for global
20 quantitative proteome analysis, suffer from several problems that limit the utility of the ICAT technique (Gygi et al., 1999). These limitations include (1) copurification of spurious, non-labeled peptides following elution of ICAT-labeled peptides from monomeric avidin; (2) the presence of a large biotin adduct which complicates CID fragmentation spectra interpretation; (3) unequal
25 retention times of heavy and light ICAT reagents on RP-HPLC resulting in heavy and light peptides not being ionized in the MS source simultaneously; and (4) nonspecific reactivity of ICAT reagents with noncysteine nucleophiles in proteins. In contrast, the C-terminal origin of each peptide prepared by the
methods of the invention provides a powerful constraint when its identity is
30 determined, making CID mass spectra interpretation substantially simpler and more reliable. This constraint may make it possible to identify the parent protein based solely on the molecular weight of the peptide. Thus, the mass of a peptide

would be compared to the predicted molecular weights of all C-terminal peptides in a given species' proteome. In cases where the abundance of the peptide is too low to obtain fragmentation data, using the mass of the peptide to identify the parent protein substantially improves the sensitivity compared to approaches, such as ICAT, in which CID mass spectra is essential for peptide identification.

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 MPACT, two or more samples are distinguished based on the corresponding C-terminus of the proteins in each sample. For instance, the label for one sample comprises an affinity marker and a non-natural isotope, e.g., a deuterated tag, ^{13}C or ^{15}N , while the label for the other sample comprises the affinity marker but does not comprise the non-natural isotope.

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) are obtained. Proteins in one sample are C-terminally labeled with a label which does not comprise a non-natural isotope. Proteins in the other sample are C-terminally labeled with a label comprising one or more non-natural isotopes. In one embodiment, the two labeled samples are mixed prior to hydrolysis. After hydrolysis, e.g., with a protease, the C-terminally labeled peptides or C-terminal peptides, obtained by removing the affinity marker from the terminally labeled 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. Most commonly, fractionation is accomplished by reverse phase HPLC, ion exchange chromatography, or capillary electrophoresis. Ideally, the majority of labeled 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. Corresponding C-terminal 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.

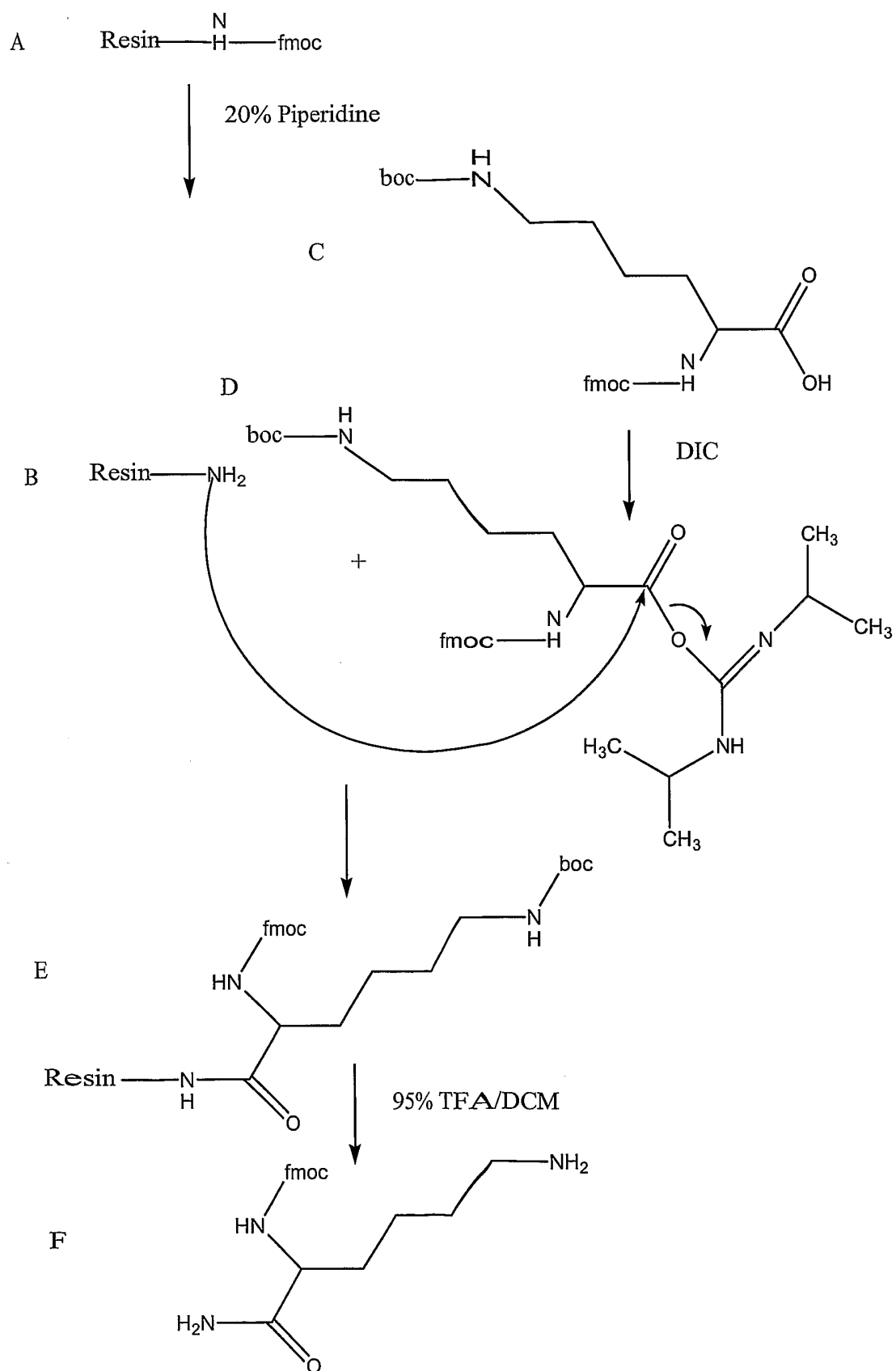
The invention will be further described by the following non-limiting examples.

Example 1

Synthesis and Purification of N α -Fmoc-Lys-NH₂

Step 1: Synthesis of N α -Fmoc-Lys-NH₂

N α -Fmoc-Lys-NH₂ is used to synthesize PCB-NH₂. The presence of the Fmoc protecting group allows PCB-NH₂ to react with the N ϵ of Lys instead of N α .



A. NHS-PC-LC Biotin coupling reaction with Fmoc-Lys-NH₂.

B. Fmoc-PCB-NH₂

C. PCB-NH₂

5

Method

1. Removal of Fmoc from Knorr Resin

- a. Place 125 mg (capacity = 0.1 mmole) of Knorr Resin (Advanced ChemTech) into a 10 mL peptide synthesis vessel.
- b. Turn the stopcock of the vessel to the closed position.
- c. Add 1.25 mL of 20% piperidine/DMF.
- d. Slowly bubble N₂ through to agitate the mixture for 30 minutes at RT.
- e. Vacuum-filter the resin and wash three times with 5 mL DMF.

10

2. Attachment of N α -Fmoc-N ϵ -Boc-Lys to the resin

- a. Suspend the resin in 1.875 mL of DMF in peptide synthesis vessel.
- b. Slowly bubble N₂ through.
- c. Prepare lysine solution:
 - i. In a 20 mL vial, dissolve 117 mg (0.25 mmole) of N α -Fmoc-N ϵ -Boc-Lys (Sigma) in 3 mL of DMF.
 - ii. Add 47.2 mg (0.35 mmole) of hydroxybenzotriazole (ICN).
 - iii. Tap the bottom of the vial until HOBt is completely dissolved.
 - iv. Add 39.1 μ L (0.25 mmole) of diisopropylcarbodiimide (Sigma).
 - v. Incubate for 10 minutes at room temperature.
- d. Add lysine solution to resin.
- e. Allow the reaction to proceed, with nitrogen bubbling through, for 3 hours at room temperature.

20

25

30

3. Blocking of unreacted resin (optional)

- a. Add 18.9 μ L of acetic anhydride (0.2 mmole) to the reaction vessel.
- b. Add 16.2 μ L of pyridine (0.1 mmole).
- 5 c. Allow blocking reaction to proceed for 30 minutes at room temperature.

4. Cleavage of N α -Fmoc-Lys-NH₂

- a. Once the blocking reaction is completed, drain the solvent by vacuum filtration.
- 10 b. Wash the resin three times with 5 mL of DMF.
- c. Wash the resin three times with 5 mL DCM.
- d. Turn the nitrogen off.
- e. Add 3-5 mL of 95% TFA/DCM.
- f. Allow the mixture to stand in RT for 30 minutes.
- 15 g. Filter and wash the resin with 5 mL of DCM.
- h. Combine the filtrate.

Step II. Purification of N α -Fmoc-Lys-NH₂

1. Vacuum-dry the filtrate to obtain red sticky tar.
- 20 2. Add 10-15 mL of cold (-20°C) ether to the dried sample. Red tar becomes white/ivory.
3. Incubate the sample at -20°C for 30 minutes.
4. Transfer the mixture into a 50 mL conical.
5. Spin the mixture at 3K rpm for 10 minutes.
- 25 6. Pipet away the supernatant.
7. Wash the precipitant with cold ether.
8. Repeat steps 5 through 7 twice.
9. Transfer the mixture to a scintillation vial.
10. Dry off the ether using speedvac.
- 30 11. Store the product dry at -20°C.
- a. HPLC parameters
 - i. Flow : 4 mL/minutes

- ii. UV : 280 nm
- iii. Inject : 100 μ L
- iv. Gradient
 - 1. 0% D and 100% A (t = 0 minutes)
 - 2. 0% D and 100% A (t = 5 minutes)
 - 3. 100% D and 100% A (t = 25 minutes)
 - 4. 100% D and 100% A (t = 30 minutes)

*A=0.1% TFA/H₂O

*A=95% ACN, 0.1% TFA, 4.9% H₂O

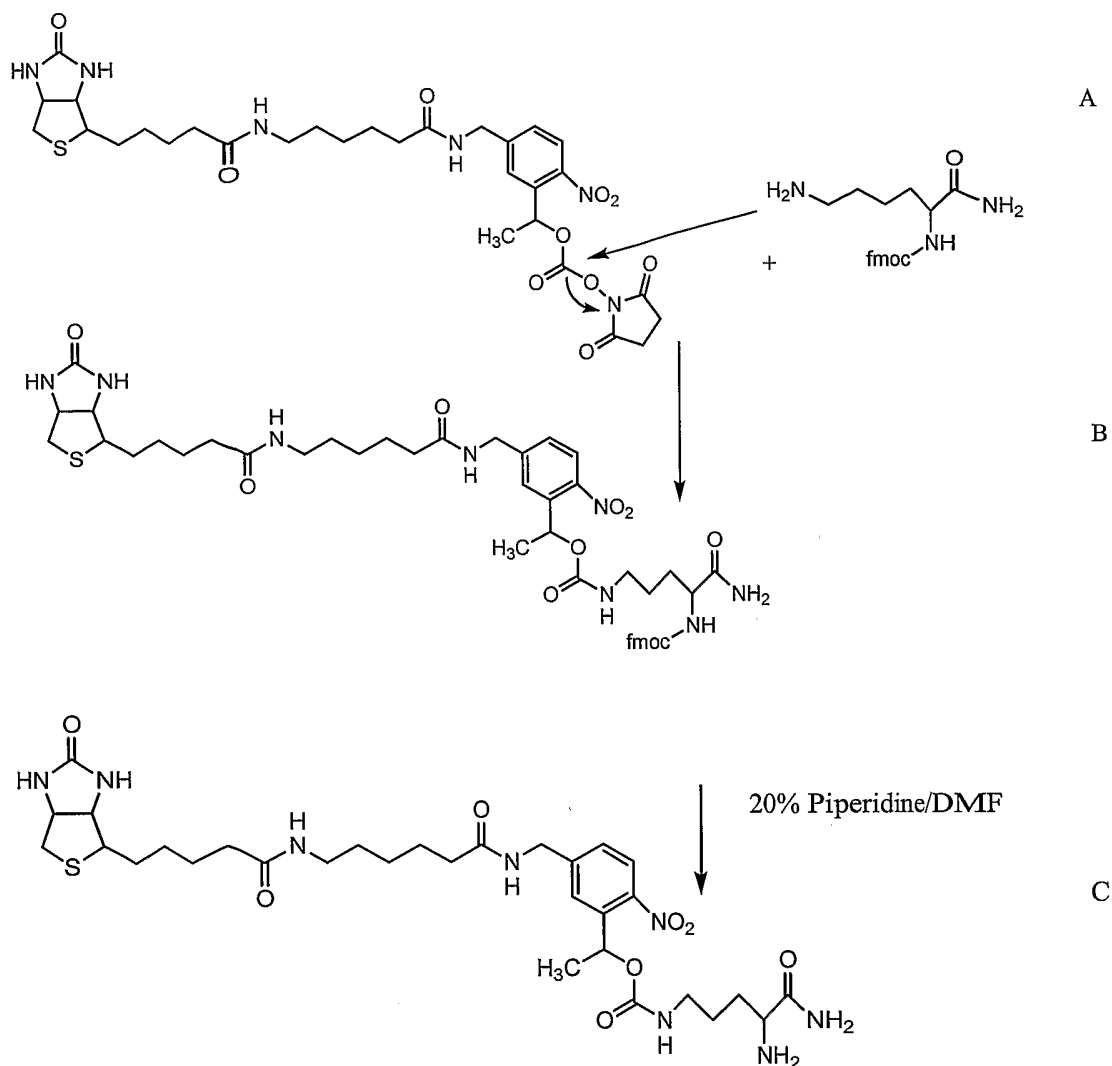
Figure 1 is a purity analysis of Fmoc-Lys-NH₂ by HPLC. 2 mg of Fmoc-Lys-NH₂ was dissolved in 300 μ L of water. To prepare the sample for HPLC, particulate materials were removed by spinning down the sample at 14k rpm for 5 minutes. 10 μ L of supernatant was injected to HPLC, 9.4 x 250 mm C-18 semi-prep column (5 μ m).

Example 2

Synthesis and Purification of Photocleavable Biocytinamide (PCB-NH₂)

Step I: Synthesis of PCB-NH₂

PCB-NH₂ is a molecule which can be used in CPY-dependent C-terminal labeling reaction which reaction products can release protein-Lys-NH₂ when irradiated with UV light (300-360 nm) and so PCB-NH₂ is a useful reagent in purifying peptides and/or proteins. PCB-NH₂ can be synthesized from NHS-PC-LC-Biotin (Pierce, 21332) and Fmoc-Lys-NH₂, and the product purified by HPLC.



- 5 A. NHS-PC-LC Biotin coupling reaction with Fmoc-Lys-NH₂
 B. Fmoc-PCB-NH₂
 C. PCB-NH₂

Methods

- 10 1. NHS-PC-LC-Biotin Solution (MW = 676.74) (Avoid light exposure)
- Dissolve 2 mg (3.0 μ moles) of NHS-PC-LC Biotin (Pierce, #21332) in 40 μ L of DMF.
 - Add 6.28 μ L (45 μ moles) of TEA (EM Sciences)

2. Fmoc-Lys-NH₂ Solution (MW = 367.45)
 - a. Dissolve 5.9 mg (16 μ moles) of Fmoc-Lys-NH₂ (synthesized in the lab) in 233 μ L of DMF.
3. Reaction (Avoid light exposure)
 - 5 a. Put 40 μ L of NHS-PC-LC-Biotin solution in a 1.5 mL microcentrifuge tube.
 - b. Add 233 μ L of Fmoc-Lys-NH₂ solution (drop wise, over 2 minutes, with trituration) to NHS-PC-LC-Biotin solution.
 - c. Allow the reaction to proceed for 2-3 hours at room temperature,
10 in the dark.

Step II: Purification of Fmoc-PCB-NH₂

1. To prepare the sample for HPLC, remove particulate materials by spinning down the synthesis product at 14,000 rpm for 5 minutes.
- 15 2. Take 100 μ L of supernatant and run it on HPLC, C-18 semi-prep column (9.4 x 250 mm; 5 μ m).
 - a. HPLC parameters
 - i. Flow: 4 mL/minute
 - ii. UV: 280 nm
 - 20 iii. Inject: 100 μ L
 - iv. Gradient
 1. 0% D and 100% A (t = 0 minutes)
 2. 0% D and 100% A (t = 5 minutes)
 3. 100% D and 100% A (t = 25 minutes)
 - 25 4. 100% D and 100% A (t = 30 minutes)

* A = 0.1% Formic Acid/H₂O

*D = 95% ACN/0.1 % Formic Acid/H₂O

 3. Turn off the UV at 18.6 minutes to prevent photolysis of the compound.
 4. Collect the fraction from 18.7 minutes to 19.8 minutes in a 20 mL
30 scintillation vial covered with aluminum foil.
 5. Repeat the HPLC run twice more to purify the entire sample.
 6. Dry down the purified sample using Speedvac at 4°C overnight.

HPLC analysis of Fmoc-PCB-NH₂ is shown in Figure 2. The fraction at 19.3 minutes corresponds to Fmoc-PCB-NH₂, and the fraction at 13.7 minutes corresponds to PCB-NH₂. The photocleavable biotin may attach to either N α or N ϵ of Lys-NH₂.

5

Step III. Deprotection of Fmoc-PCB-NH₂

Methods

1. Dissolve 0.5 mg (0.7 μ L) of Fmoc-PCB-NH₂ in 400 μ L of DMF.
2. Add 100 μ L of piperidine (Sigma) to make 20% piperidine/DMF solution.
3. Incubate for 30 minutes at room temperature.
4. Add 42.5 μ L of formic acid (ICN) to neutralize piperidine.
5. To prepare the sample for HPLC, remove particulate materials by spinning down the synthesis product at 14,000 rpm for 5 minutes.
- 15 6. Take 100 μ L of supernatant and run it on HPLC, C-18 semi-prep column (9.4 x 250 mm; 5 μ m).
 - a. HPLC parameters
 - i. Flow: 4 mL/minute
 - ii. UV: 280 nm
 - 20 iii. Inject: 100 μ L
 - iv. Gradient:
 1. 0% D and 100% A (t = 0 minutes)
 2. 0% D and 100% A (t = 5 minutes)
 3. 100% D and 0% A (t = 25 minutes)
 - 25 4. 100% D and 0% A (t = 30 minutes)
- * A = 0.1 % Formic Acid/H₂O
- * D = 95% ACN/0.1 % Formic Acid/H₂O
7. Turn off the UV at 14.2 minutes to prevent photolysis of the compound.
8. Collect the fraction from 14.3 minutes to 15.3 minutes in to 20 mL scintillation vial covered with aluminum foil.
- 30 9. Repeat the HPLC run four times more to purify the entire sample.
10. Dry down the purified sample using Speedvac at 4°C overnight.

Deprotection of Fmoc may also be conducted immediately after the synthesis.

Figure 3 shows the deprotection of Fmoc-PCB-NH₂.

Example 3

Preparation of C-terminally Labeled Peptides using PCB- NH₂

1. Protein Methyl Esterification.

- Make 25 mM HCl solution in methanol.
 - Use concentrated HCl to minimize H₂O in sample.
- Dissolve 5 mg of protein in 1 mL of 25 mM HCl/MeOH
- Incubate at the protein solution at 4°C on a rocking platform for 48 to 72 hours.

2. Removal of MeOH.

- Make 25 mM NaOAc/H₂O from a 3 M stock, pH 5.2
 - Samples incubated less than 48 hours or more than 72 hours produced less biotinylated BSA signals on Western blots.
- Add 1 mL of 25 mM NaOAc solution to 1 mL of protein methyl ester sample (from step #1)
- Rotary-evaporate the above solution to remove methanol (The final volume should be 1 mL)
 - Since the pump pulls relatively strong vacuum and methanol is very volatile, place the reaction flask in ice water bath.
 - Place the collecting flask in ice.
 - BSA-methyl-ester produced less biotinylated BSA signals when sulfuric acid was used instead of hydrochloric acid. This effect was seen only when rotary evaporation was used to remove methanol from the sample. This effect was not seen when acetone precipitation was used.

3. CPY labeling (performed in the dark)

- Make stock solution of 5.4 mM PCB-NH₂ in H₂O.
- Make 1 M Na₂CO₃ fresh, 2 mg of CPY in 1 mL of 1:1 PBS:Glycerol, and 20% Triton X-100.

- Keep all the samples/reaction mixtures in ice.
- Stock CPY solution is diluted 10x with 1:1 PBS:Glycerol.
10% stock has been found to produce better labeling.
- The enzymatic activity of CPY in PBS Glycerol solution did
5 not diminish even after two months in 4°C.
- Make batch solution (for 75 µL of methanol-free protein-methyl
ester) of Na₂CO₃, TritonX-100, and PCB-NH₂
 - Put 7.5 µL of 1 M Na₂CO₃, 7.5 µL of 20% TritonX-100, and
10 70 µL of 5.4 mM PCB-NH₂/H₂O into an eppendorf tube.
Triturate the solution.
- Put 8.5 µL of batch solution to the bottom of an eppendorf tube.
- Put 7.5 µL of rotary vaporized protein methyl ester solution (from
step # 2) onto the inside of the cap of the eppendorf tube.
- Put 1 µL of 10% CPY reagent on the inside wall of the eppendorf
15 tube.
- This order of addition keeps CPY and protein-methyl ester
from the harsh basic condition of sodium carbonate until the
reaction/incubation.
- Spin the samples (3000 rpm for 15 seconds)
- 20 - Tap the bottom of the eppendorf tube to mix the reagents. Spin
the tubes again.
- Add 0.42 µL of 20% SDS to make 0.4 - 0.5% SDS solution.
 - To avoid pipeting less than 1 µL, prepare samples in large
volumes 10x or larger.
- 25 • SDS concentration less than 0.4% causes precipitation.
- Keeping SDS containing sample in ice causes precipitation.
- Furthermore, addition of SDS to cold sample causes precipitation.
- Incubate the tube in 37°C for 20 minutes.
 - Time course experiment of CPY labeling reaction showed
30 20 minutes of incubation to be the optimal incubation time.

*volumes indicated in this section are designed for one sample on a SDS-PAGE.

4. Isolation of Biotinylated BSA

- Put the reaction mixture from step #3 into a 10,000 MW, Pierce Slide-A-Lyzer cassette.
- It took roughly 4 hours for 500 μ L BSA sample (10 mg/mL BSA, 0.2% SDS, 100 mM NaCl, 50 mM Hepes pH 7.7, 5 mM B-NH₂) to lose about 99% of PCB-NH₂ into 500 μ L of PBS.
- 4°C dialysis of biotinylated BSA in CPY reaction mixture showed 6 hour dialysis as preferred dialysis period.
- Dialyze each cassette against 3 L of PBS in cold room for 6 hours. Repeat this dialysis for 4 cycles. (Collect 10 μ L sample for TLC analysis, using fluorescamine to detect free amines on PCB-NH₂, at the end of each cycle).

5. Trypsinization of Biotinylated Protein

- Dissolve 4.2 mg of trypsin (Sigma T-8802) in 5.6 mL of cold 50 mM Hepes, pH 7.7.
- Keep trypsin in ice to minimize the loss enzymatic activity of trypsin.
- Dilute above trypsin solution by 10 times, with 50 mM Hepes pH 7.7, to obtain 5% (mass ratio against protein-methyl ester) trypsin solution.
- Incubate dialyzed, biotinylated proteins (from step #4) with 5% trypsin at 37°C overnight. (The volume ratio between protein sample and 5% trypsin is 17:50).
- 5% trypsin (relative to substrate mass) produces enhanced digestion of biotinylated BSA.

6. Inactivation of Trypsin

- Pack NeutrAvidin resin on a fritted column.
- NeutrAvidin resin has binding capacity of 20 μ g or 82 nmoles of Biotin/mL of resin.

- Experiments with pyruvate carboxylase showed NeutrAvidin resin captures more biotinylated molecules when multiple column loading (10 times or more reloading) was used rather than 1-2 hours of batch loading.
- 5 - Equilibrate the resin with 10x the resin volume of 50 mM Hepes pH 7.7. Repeat 3 times.
- Load the sample.
- Collect the flow through and reload onto the column.
- Repeat the reloading step 20 times.
- 10 - Wash the resin with 10x the resin volume of 50 mM Hepes pH 7.7 three times.
- Wash the resin the 10x the resin volume of 50 mM Hepes pH 7.7 that has 1 M NaCl.
- Wash the resin with 10x the resin volume of 50 mM Hepes pH
- 15 7.7.
- Wash the resin with 10x the resin volume of 1 M Guanidine HCl/100 mM Hepes pH 7.7.
- Wash the resin with 10x the resin volume of 50 mM Hepes pH
- 7.7.
- 20 - Wash the resin with 10x the resin volume of H₂O three times.
- Resuspend the resin in 3x the resin volume of 50 mM Hepes pH
- 7.7.
- Transfer the resin and the 50 mM Hepes pH 7.7 solution to a quartz cuvet.
- 25 - Irradiate the cuvet for 2 hrs with 365 nm UV lamp while gently mixing the solution.
- Transfer the resin and the 50 mM Hepes pH 7.7 solution to an eppendorf tube.
- Spin the resin down and collect the supernatant.
- 30 - Resuspend the resin with 2x the resin volume of 50 mM Hepes 7.7 and spin down the resin. Collect the supernatant.

8. Removal of Salts from the Peptide Solution

- Condition the Waters OASIS HLB 1 cc Extraction Cartridge with 1 mL of acetonitrile.
- Equilibrate the column with 1 mL H₂O twice. Use a syringe and a reservoir adapter (Waters) to force the liquid down.
- Load the sample (from step #7).
- Wash the column with 1 mL of H₂O four times. Use a syringe and a reservoir adapter (Waters) to force the liquid down.
- Elute the sample in 1 mL of 0.1 % TFA in acetonitrile.
- Collect the eluate and elute with another 1 mL of 0.1% TFA in acetonitrile.
- Dry down the eluate in speedvac at 4°C.

9. Purification of Peptides

- Resuspend the dried down sample in minimal ACN (usually 20 - 30 µL)
- Dilute the sample with water. Reduce the ACN concentration down to 10% or less.
- To prepare the sample for HPLC, remove particulate materials by spinning down the sample at 14 k RPM for 5 minutes.
- HPLC purify the peptides
 - HPLC parameters
 1. UV : 20
 2. Flow : 4 mL/minute
 3. Column : C-18 Semi-Prep 9.6 x 250 mm (5 µm)
 4. Gradient (A=0.1 % TFA/H₂O D=95% ACN, 0.1% TFA, 4.9% H₂O)
 - a. 0% D and 100% A at t = 0 minutes
 - b. 0% D and 100% A at t = 5 minutes
 - c. 100% D and 0% A at t = 25 minutes
 - d. 100% D and 0% A at t = 30 minutes
- Collect peptide fractions in tared glass vials and dry them down.
- Weigh the vials that contain dry samples.

- Calculate the number of moles in the sample.
 - Redissolve the samples in ACN.
10. Analyze the sample using MALDI.

5

Example 4

Mass Spectrometric Proteome Analysis with C-terminal Tags (MPACT)

A peptide mixture obtained from the trypsinization of a protein mixture could be maximally simplified if just one peptide was recovered for each protein in the starting material. If each protein in the starting material could be modified with a single label, the digested protein would produce several peptides, only one of which would possess the label, and the label could then be used to retrieve the single peptide. One approach would be to label one end of the protein, for example, the C-terminus. The incorporated label could be biotin because modified proteins could be detected by Western blotting with biotin-specific antibodies and streptavidin-agarose provides a way to recover the labeled proteins.

The ability to label and recover C-terminal amino acids provides a straightforward way to perform proteome analysis (Figure 4). In this method protein mixtures are labeled on their C-termini with a label comprising an affinity marker, e.g., biotin, and then digested with a specific protease to generate short peptides. The C-terminal peptides are selectively recovered using a ligand for the affinity marker, e.g., streptavidin-agarose, and their masses determined by liquid chromatography-coupled electrospray ionization (LC-ESI) MS. To correlate the peptide mass with the protein from which they were derived, the molecular mass of C-terminal peptides, in most cases, can be used uniquely to identify the protein from which the peptide was derived. Thus, a 2337.82 Da peptide is uniquely associated with catalase. These molecular masses can be calculated from the amino acid sequence of each protein identified in the genome sequencing project of that organism.

The size distribution and overlap of labels useful in the methods of the invention was tested using the yeast genomic sequence obtained from GenBank. A computer simulation of a proteolytic digestion by endoprotease Glu-C was

generated. Following a virtual digest of the yeast proteome with this protease, 63.5% of proteins produces C-terminal peptides of appropriate size for high resolution (6-15 residues in length), unambiguous identification by ESI/MS.

Use of endoprotease Glu-C in one digest followed by endoprotease Arg-C in the next digest results in 85.7% of the proteome producing MPACT tags appropriate for identification by ESI-MS. Typically, proteins that will produce a peptide suitable for MS analysis will be known prior to MS by inspection of the genome sequence.

MPACT provides a major advance over other MS-based studies of complex mixtures. By limiting analysis of a protein mixture to its C-terminal peptides, a yeast-derived lysate, which typically contains 6,000 proteins, will produce 344,000 tryptic peptides (Gygi et al., 1999), but will be simplified to a maximum of about 6,000 distinct peptides by MPACT. Identification of these peptides is thus powerfully constrained by the C-terminal origin of these peptides. Because MPACT produces one peptide per protein, it provides the most complete simplification of complex mixtures for proteome analysis. Also, because peptides rather than proteins are analyzed by MPACT, substantially more starting material can be used for preparing tagged C-terminal peptides than can be used for 2D-electrophoresis or other techniques that have been used for proteomics.

To avoid the simultaneous modification of carboxyl groups in glutamic and aspartic residues along with the C-terminus of proteins, a transpeptidase, CPY, an enzyme which binds to the C-termini of proteins as part of its catalytic activity, was employed. CPY is a C-terminal exopeptidase that hydrolyzes C-terminal amino acids using water as a nucleophile. CYP is also a C-terminal esterase, hydrolyzing C-terminal esters in proteins whose carboxy-terminus has been esterified (Hayashi et al., 1975). The exopeptidase activity is inhibited above pH 8, while the esterase activity is unaffected (Hayashi et al., 1975).

Surprisingly, the exopeptidase activity of CPY can utilize amino acids as nucleophiles in place of water (Berne et al., 1990). Thus, in the presence of high concentrations of, for example, alanine, the alpha-amino group of alanine can attack the peptide bond between the penultimate and last amino acid, resulting in

a replacement of the last amino acid with alanine. This reaction occurs in competition with water, and only a small fraction of proteins undergo the alanine-mediated reaction, while most undergo the water-mediated reaction (Buckler et al., 1993).

5 As the CPY-dependent amino acid incorporation reaction exists in competition with water, the C-terminal residue can be hydrolyzed several times before an amino acid is incorporated. To test if biotin-containing nucleophiles could be used as an alternative nucleophile in the esterase reaction of CPY, bovine serum albumin (BSA) was derivitized such that all the carboxyl groups,
10 including the C-terminal carboxyl group, were methyl esterified to generate a substrate for the esterase reaction. The nucleophile that was used was biocytin, a derivative of biotin found in nature. The methyl esterified protein was incubated with CPY and several biocytin analogs, including biocytin (biocytin-OH), a biocytin methyl ester (biocytin-Ome), and a carboxyamidated biocytin (biocytin-
15 NH₂). CPY was found to label methyl esterified BSA (BSA-ME), and that the carboxyamidated derivative, biocytin-NH₂, is a much more efficient substrate than the other biocytin derivatives in the CPY-dependent biotinylation reaction (Figure 5).

To enhance the access of the C-terminus of the proteins to CPY, proteins
20 are initially denatured in the presence of SDS, followed by the addition of Triton X-100. CPY is added following the addition of Triton X-100. The presence of this nondenaturing detergent sequesters free SDS monomers and prevents SDS from inhibiting the enzyme.

The reaction involves the esterase activity of CPY, as BSA-ME, and not
25 unmodified BSA, was the substrate for the reaction (Figure 5B). Furthermore, removal of the methyl ester by gentle base hydrolysis prevented the incorporation of biotin (Figure 5C). Thus, CPY is capable of biotinylating BSA-ME. Trypsinolysis, and purification of C-terminal peptides using streptavidin-agarose, followed by MALDI-MS based detection of BSA-labeled protein is
30 consistent with a C-terminal labeling reaction.

To determine if CPY was capable of modifying a variety of proteins, regardless of the C-terminal amino acid, twenty glutathione-S-transferase (GST)

fusions containing GST followed by each of the twenty common amino acids were prepared, and it was found that each was a substrate for CPY (Figure 5D), indicating that CPY could label a variety of proteins, and each C-terminal amino acid could be used in a labeling reaction.

5 In other experiments, enzyme, protein substrate, and nucleophile concentrations for the CPY-dependent labeling reaction were altered. Identical results were obtained for a variety of other selected proteins, including superoxide dismutase, glyceraldehydes-1-3-phosphate dehydrogenase catalase, and calmodulin. Additionally, whole rat liver lysate and yeast mitochondrial
10 preparations were similarly biotinylated by CPY in a biocytin-NH₂-dependent fashion.

Example 5

Biotinamidoethyl Dithiocysteineamide (BADTC)

15 As described above, prior to MALDI-MS, peptides were purified on streptavidin-agarose and eluted with guanidine HCl. The peptides were then desalted using reverse phase chromatography. The unusually high affinity of the biotin-streptavidin interaction (k_D of about 10^{-15}) resulted in less than optimal elution of the peptides from the streptavidin-agarose. To enhance the recovery
20 of C-terminal peptides, a biotinylated amino acid, biotinamidoethyl dithiocysteineamide (BADTC), was prepared (Figure 6). BADTC is similar to biotin-NH₂ in chemical structure, except that BADTC has a disulfide bond, which can be reduced with 2-mercaptoethanol (2-ME). BADTC-labeled peptides were purified on streptavidin-agarose and then eluted from the resin
25 with 2-ME (a mild elution protocol). The eluted peptides each contain a C-terminal cysteineamide, however, biotin remains on the streptavidin-agarose. 2-ME is removed by lyophilization prior to ESI-MS. To confirm that BADTC-labeling is compatible with a variety of proteins as well as GST fusions containing each of the twenty common C-terminal amino acids at its C-terminus,
30 various fusions are prepared in *E. coli* in a manner identical to the GST-X experiments described in Example 4 and analyzed.

Analogues of BADTC may be better substrates for CPY ligation. In particular, the mercaptoalkylamine component of BADTC can be modified by using various mercaptoalkylamines with different numbers of methylene groups. The role of distance between the alpha-carbon and the disulfide may be
5 evaluated using homocysteine, which has an additional methylene group between the alpha-carbon and the sulfur atom, rather than cysteine. Structures are confirmed using ¹H-NMR and mass spectrometry.

In order to perform comparative MPACT (Figure 7), a strategy in MS involving the use of labels comprising non-natural isotopes such as deuterated
10 tags, is employed. *d*₃-BADTC, containing three deuterium ions in the cysteinamide moiety of BADTC, is synthesized, using commercially available *d*₃-cysteine. To compare two samples, proteins from one sample, e.g., condition-1, are labeled with BADTC, while proteins from sample 2, e.g., condition-2, are labeled with *d*₃-BADTC (Figure 7). The samples are pooled and trypsinized
15 together, and peptides comprising the affinity marker are recovered, e.g., using avidin-agarose and 2-ME, and simultaneously subjected to LC-ESI MS. Because each peptide has either a cysteinamide or a *d*₃-cysteinamide, the MS spectrum will contain a series of doublets separated by 3 atomic mass units; the lighter peak corresponding to a peptide derived from condition-1, and the
20 heavier peptide corresponding to condition-2. The relative expression level is determined by the ratio of the signal intensity between the two peaks.

As an alternative approach, a PCB-NHS ester is condensed with ¹³C₄- or *d*₄- lysinamide to generate a photocleavable lysinamide non-natural isotope containing affinity marker (PLICAT), and purified by HPLC. PLICAT synthesis
25 is verified by NMR and MS. Although structurally longer than biocytin-NH₂, photocleavage of PLICAT results in the incorporation of a C-terminal lysinamide, imparting an additional positive charge to the peptide fragment that, in some cases, facilitates detection after CID. PLICAT derivatives using lysinamide derivatives are also prepared.

30 MPACT can thus be used to examine more complex proteomes, including whole yeast, and human biological specimens, including physiological

fluids include CSF and serum, as well as other extracellular samples, and subcellular, e.g., membrane, fractions.

References

- 5 Adams et al., Science, 287:2185 (2000).
Alexandre et al., FEBS Lett., 498:98 (2001).
Anderson et al., Electrophoresis, 18:533 (1997).
Berne et al., J. Biol. Chem., 265:19551 (1990).
Brand, Cancer J., 7:287 (2001).
- 10 Buckler et al., Anal. Biochem., 209:20 (1993).
Endrizzi et al., Biochemistry, 33:11106 (1994).
Futcher et al., Mol. Cell. Biol., 19:7357 (1999).
Gygi et al., Mol. Cell. Biol., 19:1720 (1999).
Gygi et al., Nat. Biotechnol., 17:994 (1999).
- 15 Haurie et al., J. Biol. Chem., 276:76 (2001).
Hayashi et al., J. Biochem. (Tokyo), 77:69 (1975).
Jaffrey et al., Nat. Cell. Biol., 3:193 (2001).
Koller, Exp. Neurol., 144:24 (1997).
Mewes et al., Nature, 387:7 (1997).
- 20 Mortensen et al., Biochemistry, 33:508 (1994).
O'Farrel et al., J. Biol. Chem., 250:4007 (1975).
Olesen et al., Biochemistry, 36:12235 (1997).
Porath et al., Nature, 258:598 (1992).
The *C. elegans* Sequencing Consortium, Science, 282:2012 (1998).
- 25 Sechi et al., Anal. Chem., 70:5150 (1998).
Sorenson et al., Protein Sci., 6:2227 (1997).

All publications, patents and patent applications are incorporated herein by reference. While in the foregoing specification, this invention has been
30 described in relation to certain preferred embodiments thereof, and many details have been set forth for purposes of illustration, it will be apparent to those skilled in 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 C-terminally labeled peptides, comprising:
 - a) providing a hydrosylate comprising a mixture of peptides obtained by subjecting a plurality of C-terminally labeled proteins to hydrolysis, wherein the label comprises an affinity marker, and wherein the C-terminally labeled proteins are prepared by contacting a sample comprising a plurality of proteins with the label and a transpeptidase; and
 - b) contacting the hydrosylate with a ligand for the affinity marker so as to form a complex, thereby isolating C-terminally labeled peptides.
2. The method of claim 1 wherein the affinity marker is biotin, biocytin, a peptide antigen, polyhistidine, dinitrophenol, an oligonucleotide or a peptide nucleic acid.
3. The method of claim 1 wherein the label is photocleavable.
4. The method of claim 1 wherein the transpeptidase is carboxypeptidase Y.
5. The method of claim 1 wherein the label further comprises a non-natural isotope.
6. The method of claim 1 wherein the plurality of proteins in the sample are esterified prior to contact with the transpeptidase.
7. The method of claim 1 wherein the ligand is streptavidin, avidin, neutravidin, or monomeric avidin.
8. The method of claim 1 wherein the ligand is coupled to a solid support.

9. The method of claim 1 wherein the hydrolysate is prepared by contacting the plurality of C-terminally labeled proteins with a protease.
10. The method of claim 9 wherein the protease is trypsin, chymotrypsin, pepsin, papain, proteinase K, calpain, subtilisin endoprotease Glu-C, or endoprotease Arg-C.
11. The method of claim 1 wherein the hydrolysate is prepared by chemical hydrolysis.
12. The method of claim 11 wherein the hydrolysate is prepared by contacting the plurality of C-terminally labeled proteins with cyanogen bromide.
13. The method of claim 1 wherein the plurality of proteins are denatured prior to contact with the transpeptidase.
14. The method of claim 13 wherein the plurality of proteins are denatured with SDS.
15. The method of claim 1 further comprising disrupting the complex formed by the binding of the ligand to the affinity marker to yield purified C-terminally labeled peptides.
16. The method of claim 15 further comprising separating the purified C-terminally labeled peptides by mass spectrometry.
17. The method of claim 16 further comprising, prior to separating the purified C-terminally labeled peptides by mass spectrometry, fractionating the purified C-terminally labeled peptides.

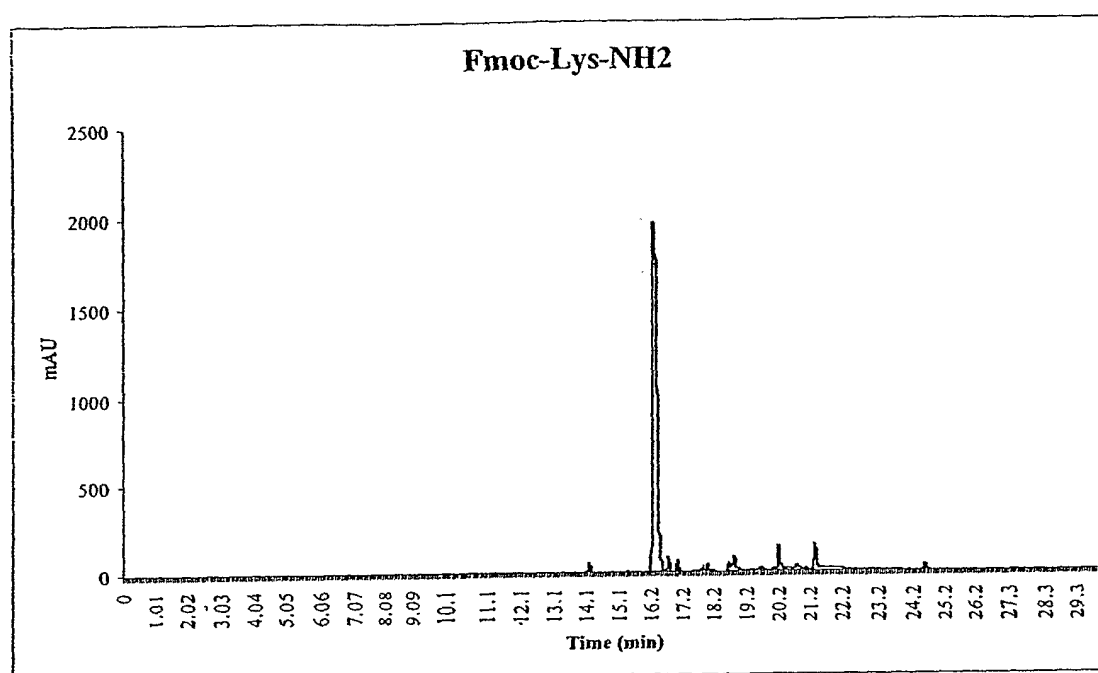
18. The method of claim 17 wherein liquid chromatography, ion exchange chromatography or capillary electrophoresis is employed to fractionate the purified C-terminally labeled peptides.
19. The method of claim 16 further comprising identifying one or more of the separated C-terminally labeled peptides.
20. The method of claim 1 further comprising treating the isolated C-terminally labeled peptides with an agent so as to yield C-terminal peptides which lack the affinity marker.
21. The method of claim 20 further comprising purifying the C-terminal peptides which lack the affinity marker.
22. The method of claim 21 further comprising separating the purified C-terminal peptides by mass spectrometry.
23. The method of claim 22 further comprising, prior to separating the purified C-terminal peptides by mass spectrometry, fractionating the purified C-terminal peptides.
24. The method of claim 23 wherein liquid chromatography, ion exchange chromatography or capillary electrophoresis is employed to fractionate the purified C-terminally labeled peptides.
25. The method of claim 22 further comprising identifying one or more of the separated C-terminal peptides.
26. The method of claim 1 wherein the sample is a cellular sample.
27. The method of claim 1 wherein the sample is a physiological fluid sample.

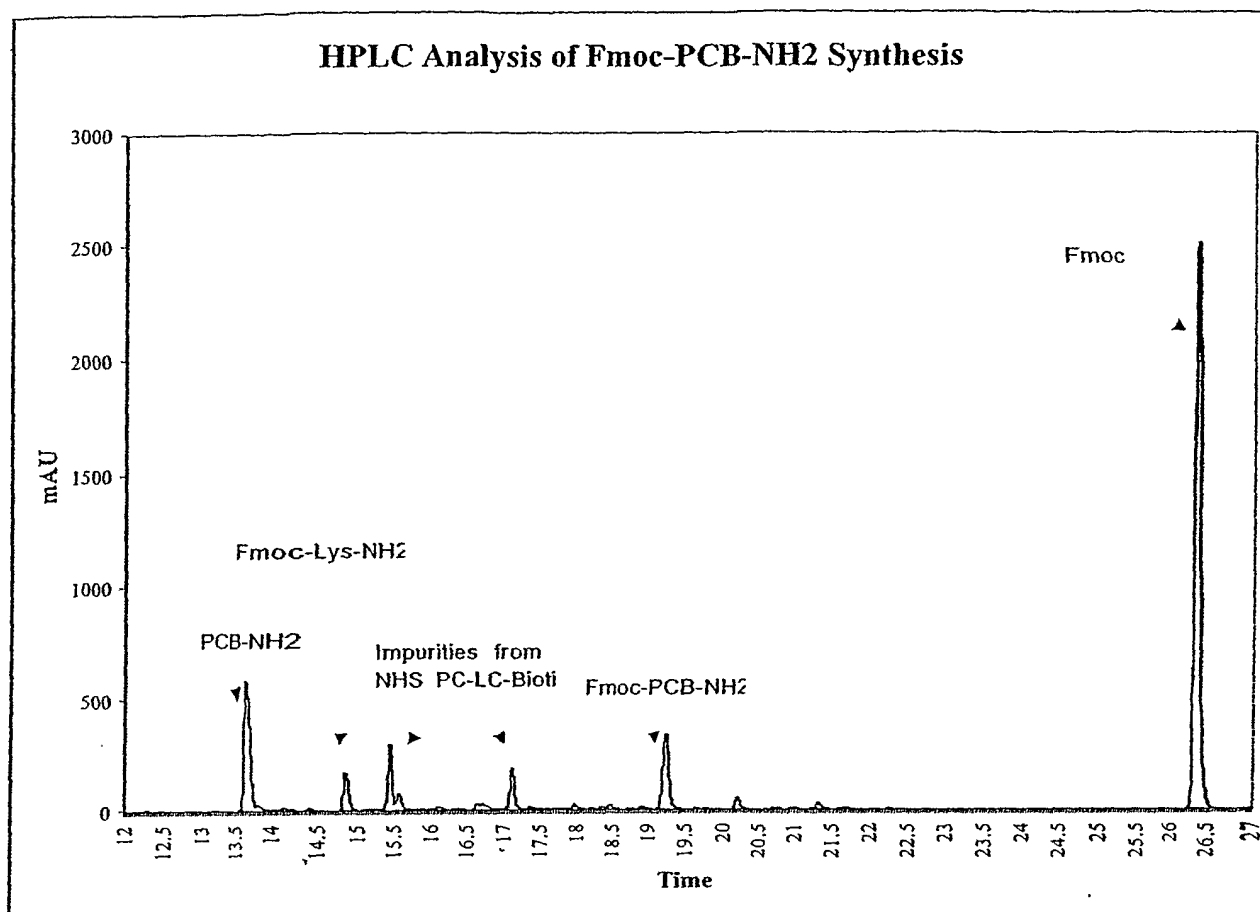
28. The method of claim 27 wherein the sample is serum, plasma, urine or cerebrospinal fluid.
29. The method of claim 1 wherein the sample is a subcellular fraction.
30. A method of separating C-terminally labeled peptides in a mixture, comprising:
- providing a hydrosylate comprising a mixture of peptides obtained by subjecting a plurality of C-terminally labeled proteins to hydrolysis, wherein the label comprises an affinity marker, and wherein the C-terminally labeled proteins are prepared by contacting a sample comprising a plurality of proteins with the label and a transpeptidase;
 - contacting the hydrosylate with a ligand for the affinity marker so as to form a complex, thereby isolating C-terminally labeled peptides;
 - collecting the isolated C-terminally labeled peptides; and
 - separating the isolated C-terminally labeled peptides by mass spectrometry.
31. The method of claim 30 further comprising identifying one or more of the separated peptides, thereby identifying at least one protein in the sample.
32. A method of separating peptides in a mixture, comprising:
- providing purified C-terminally labeled peptides prepared by the method of claim 15; and
 - separating the purified peptides by mass spectrometry.
33. A method of separating peptides in a mixture, comprising:

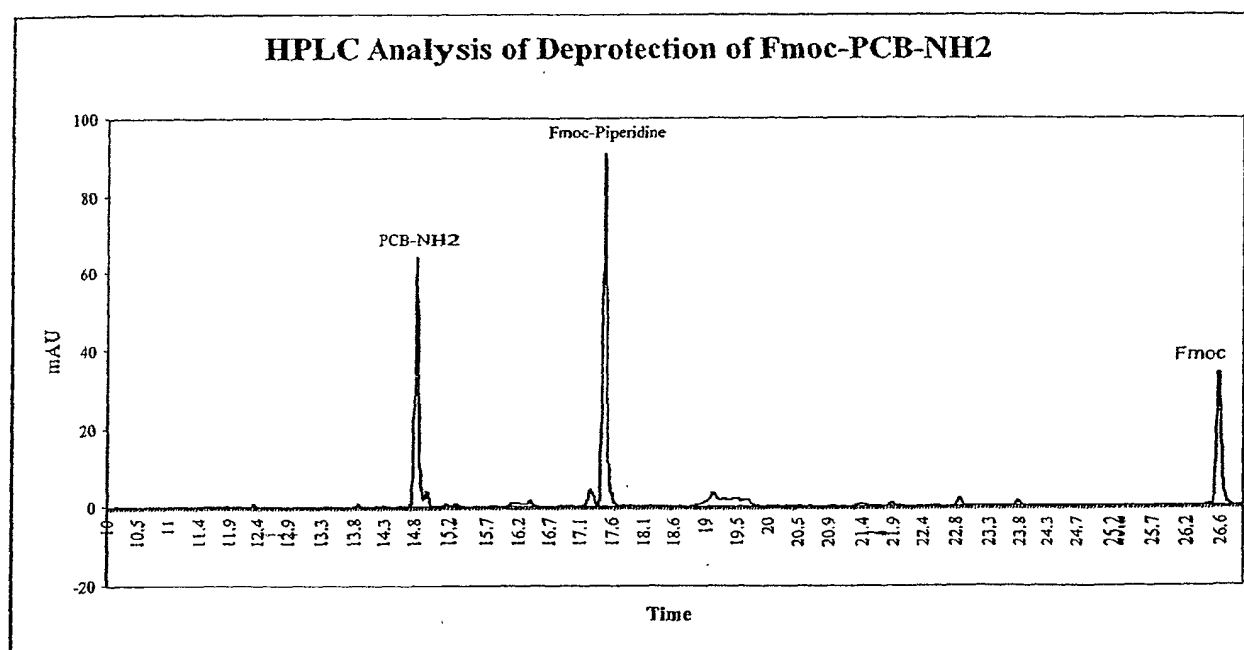
- a) providing purified C-terminal peptides prepared by the method of claim 21; and
 - b) separating the purified peptides by mass spectrometry.
34. The method of claim 32 or 33 further comprising identifying one or more of the separated peptides.
35. A method for identifying proteins in a sample comprising a plurality of proteins, comprising:
- a) contacting a hydrosylate comprising a mixture of peptides obtained by subjecting a plurality of C-terminally labeled proteins comprising an affinity marker to hydrolysis, with a ligand for the affinity marker so as to isolate C-terminally labeled peptides, wherein the C-terminally labeled proteins are prepared by contacting a sample comprising a plurality of proteins with a label comprising the affinity marker and a transpeptidase;
 - b) purifying the isolated C-terminally labeled peptides;
 - c) separating the purified C-terminally labeled peptides by mass spectrometry; and
 - d) identifying at least one separated peptide, thereby identifying at least one protein in the sample.
36. A method for identifying proteins in a sample comprising a plurality of proteins, comprising:
- a) contacting a hydrosylate comprising a mixture of peptides obtained by subjecting a plurality of C-terminally labeled proteins comprising an affinity marker to hydrolysis with a protease, with a ligand for the affinity marker so as to isolate C-terminally labeled peptides, wherein the C-terminally labeled proteins are prepared by contacting a sample comprising a plurality of proteins with a label comprising the affinity marker and a transpeptidase;

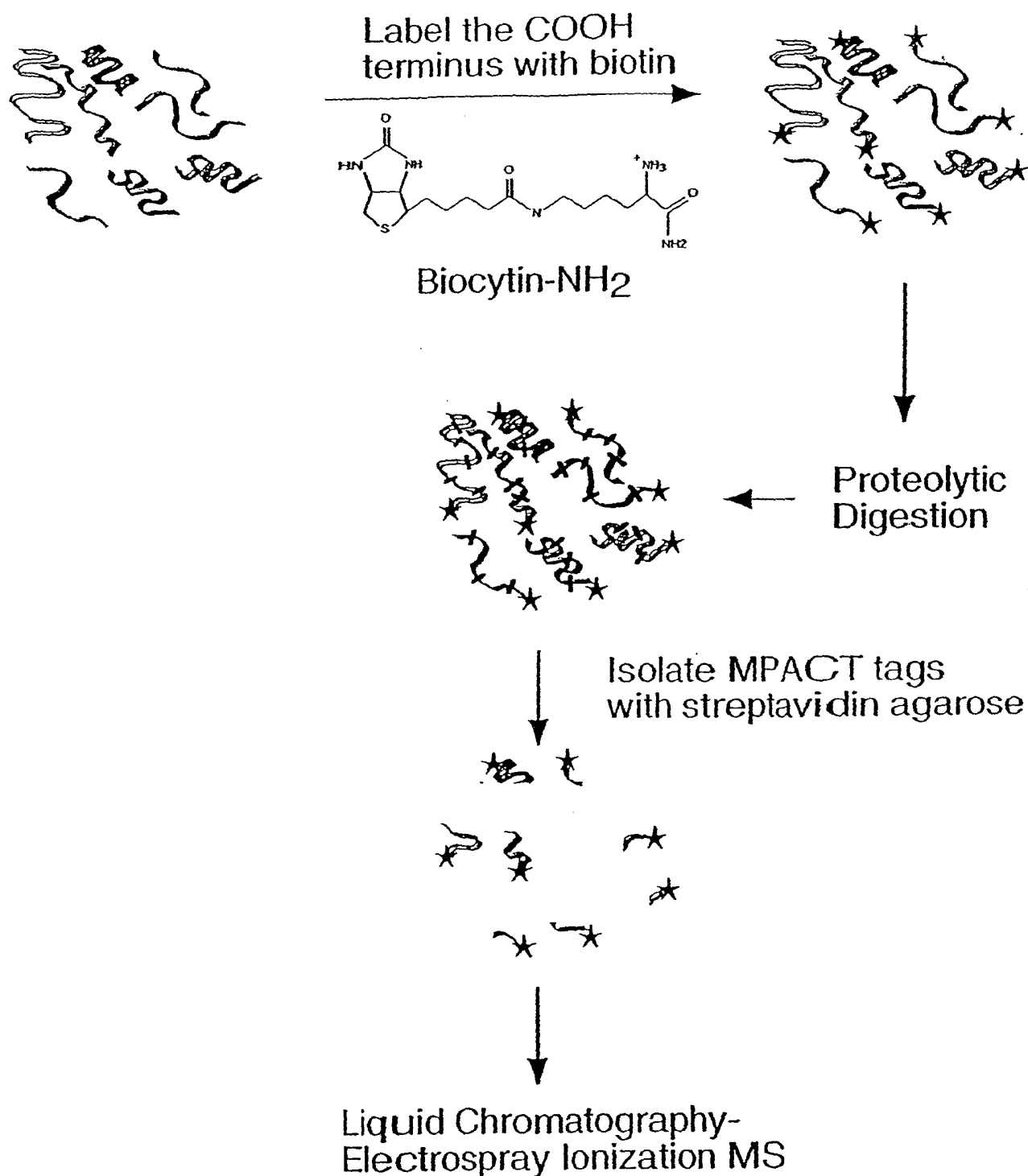
- b) treating the isolated C-terminally labeled peptides with an agent so as to yield purified C-terminal peptides which lack the affinity marker;
 - c) separating the purified C-terminal peptides by mass spectrometry; and
 - d) identifying at least one separated peptide, thereby identifying at least one protein in the sample.
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 C-terminally labeled proteins and a second sample comprising a plurality of C-terminally labeled proteins, wherein the label for the first sample comprises an affinity marker but not a non-natural isotope, wherein the label for the second sample comprises the affinity marker and a non-natural isotope, wherein the C-terminally labeled proteins in the first sample are prepared by contacting the first sample with the label comprising the affinity marker but not a non-natural isotope with a transpeptidase, and wherein the C-terminally labeled proteins in the second sample are prepared by contacting the second sample with the label comprising the affinity marker and a non-natural isotope with a transpeptidase;
 - b) hydrolysing the C-terminally labeled proteins in the first and the second sample;
 - c) contacting the hydrolysate with a ligand for the affinity marker so as to isolate C-terminally labeled peptides;
 - d) purifying the isolated C-terminally labeled peptides;
 - e) separating the purified C-terminally labeled peptides by mass spectrometry; and
 - f) comparing the amount or level of at least one separated peptide from the first sample to the amount or level of the corresponding separated peptide from the second sample.

38. 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 C-terminally labeled proteins and a second sample comprising a plurality of C-terminally labeled proteins, wherein the label for the first sample comprises an affinity marker but not a non-natural isotope, wherein the label for the second sample comprises the affinity marker and a non-natural isotope, wherein the C-terminally labeled proteins in the first sample are prepared by contacting the first sample with the label comprising the affinity marker but not a non-natural isotope with a transpeptidase, and wherein the C-terminally labeled proteins in the second sample are prepared by contacting the second sample with the label comprising the affinity marker and a non-natural isotope with a transpeptidase;
 - b) hydrolysing the C-terminally labeled proteins in the first and the second sample;
 - c) contacting the hydrolysate with a ligand for the affinity marker so as to isolate C-terminally labeled peptides;
 - d) treating the isolated C-terminally labeled peptides with an agent so as to yield C-terminal peptides which lack the affinity marker;
 - e) separating the C-terminal peptides which lack the affinity mark 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 from the second sample.
39. The method of claim 37 or 38 wherein the samples are mixed together prior to step f).

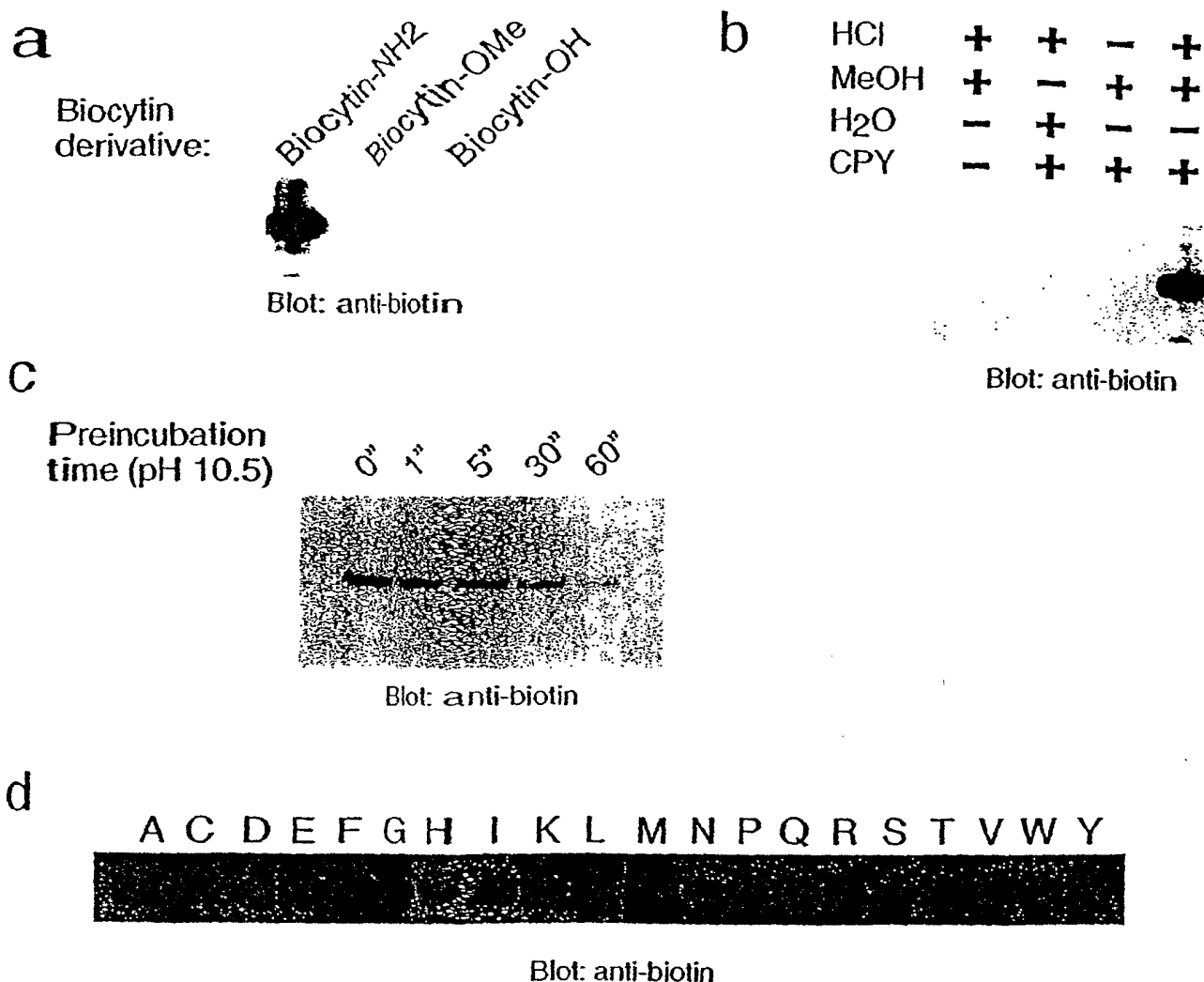


Fr₂

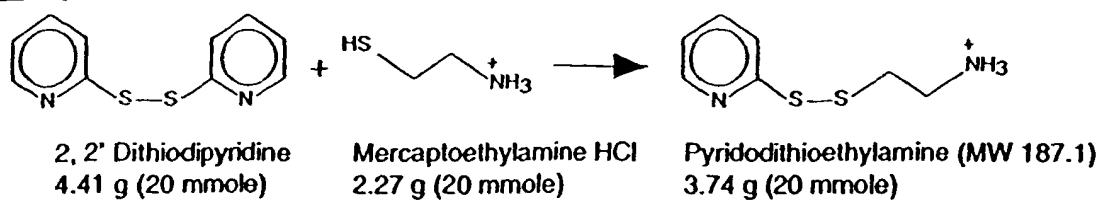
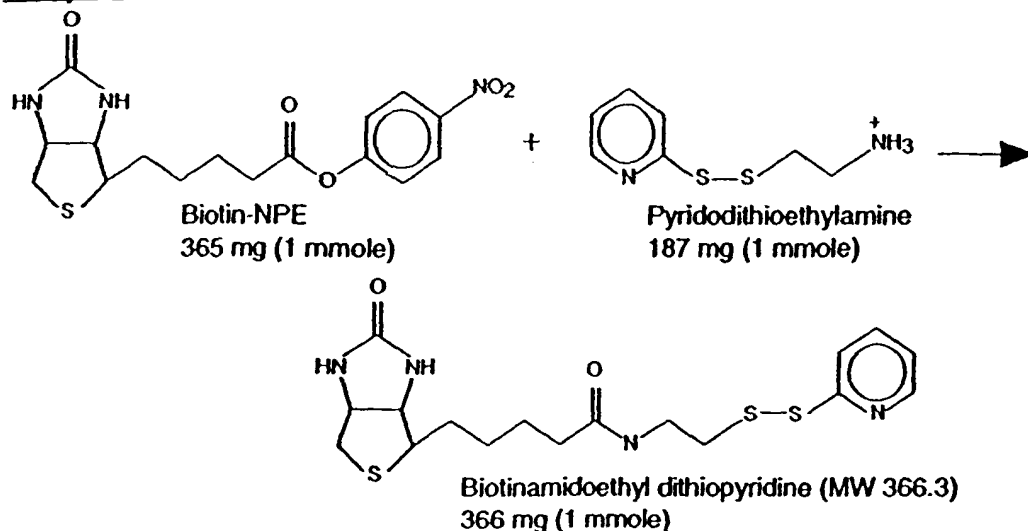
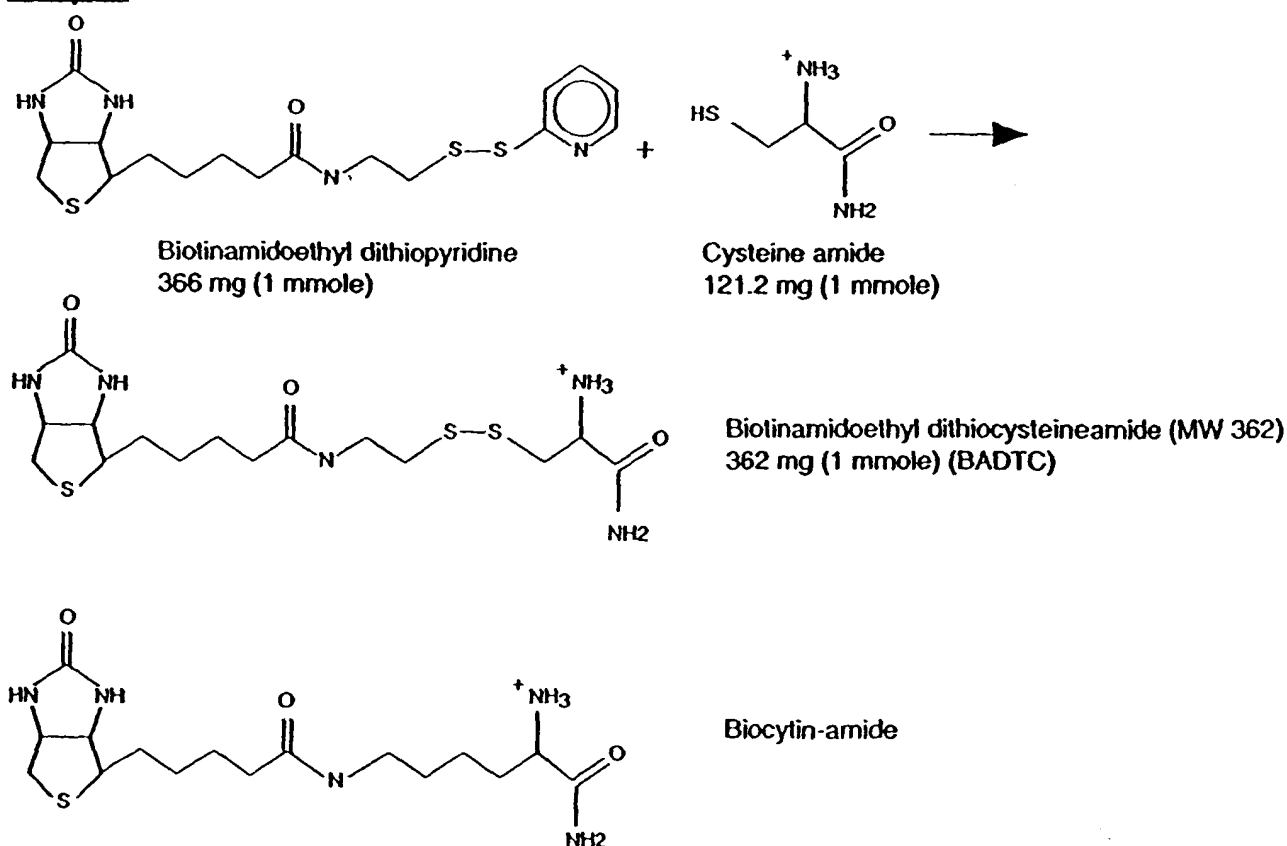




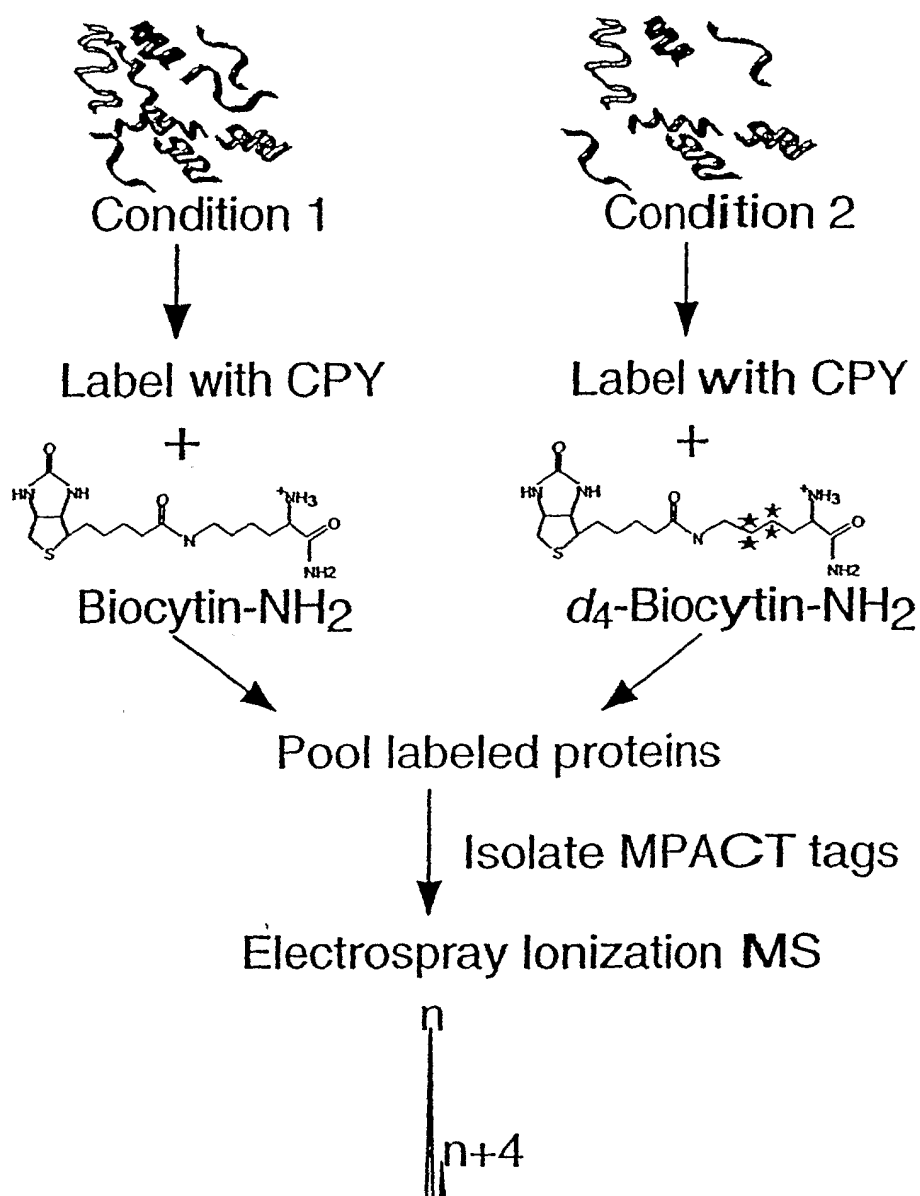
Schematic of the MPACT protocol. MPACT requires the C-terminal labeling of proteins. Subsequently, proteins are proteolyzed, and C-terminal peptides are recovered using streptavidin-agarose. The addition of biotin is indicated schematically with stars. Note that only one peptide is recovered per parent protein.



Biotinylation of proteins by CPY and biocytin-NH₂. (A) Biotinylation of BSA-methyl ester was tested by incubation of CPY and various biocytin derivatives. Biotinylated BSA was detected by anti-biotin Western blot. Biocytin-NH₂ was the preferred substrate for biotinylation. (B) Only BSA-ME is a substrate for CPY-dependent labeling. BSA prepared with methanol and HCl is methyl esterified, and this is the only preparation that is labeled by Western blot. (C) The methyl ester is required for CPY-dependent labeling. The methyl ester is hydrolyzed by base for the indicated times prior to being subjected to CPY-dependent labeling. Hydrolysis of the ester reduced the ability of BSA to serve as a substrate. (D) Biotinylation of various GST-amino acid fusion protein methyl esters. Biotinylation of the indicated GST-fusion was performed and proteins were subjected to anti-biotin Western blotting. CPY was able to biotinylate each fusion protein, indicating that all proteins, regardless of the C-terminal amino acid, are likely to be substrates for CPY-mediated ligation.

Step 1**Step 2****Step 3**

Synthetic strategy for BADTC. Biocytin-NH₂ is included at the bottom of this figure to display its structural similarity to BADTC.



Quantitative MPACT. The use of deuterated (depicted with stars) biotin-containing amino acids in the CPY-dependent labeling reaction results in peptide mixtures that have MPACT tags whose molecular mass is increased by the number of deuterium atoms in the biotin nucleophile. Identical peptides coelute and the relative amounts of each parent protein can be inferred by the relative signals produced by the two peptides.