

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
29 June 2006 (29.06.2006)

PCT

(10) International Publication Number
WO 2006/069203 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2005/046496

(22) International Filing Date:
21 December 2005 (21.12.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/638,856 22 December 2004 (22.12.2004) US
60/668,471 5 April 2005 (05.04.2005) US

(71) Applicant (for all designated States except US): **CALIFORNIA INSTITUTE OF TECHNOLOGY** [US/US];
12000 E. California Blvd, Ms 201-85, Pasadena, CA 91125 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **DIETERICH, Daniela, C.** [DE/US]; 435 S. El Molino Ave., Apt.10, Pasadena, CA 91101 (US). **TIRRELL, David, A.** [US/US]; 714 Arden Road, Pasadena, CA 91106 (US). **SCHUMAN, Erin** [US/US]; 766 Oak Knoll Circle, Pasadena, CA 91106 (US). **LINK, Aaron, J.** [US/US]; 550 E. California Blvd., Apt. 10, Pasadena, CA 91106 (US).

(74) Agents: **VINCENT, Matthew, P.** et al.; Ropes & Gray LLP, One International Place, Boston, MA 02110-2624 (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, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

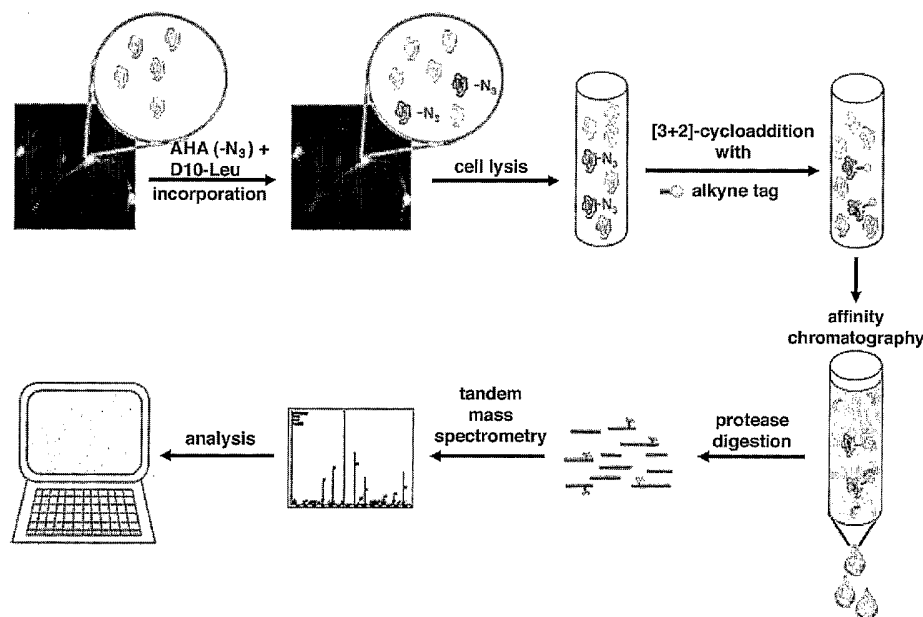
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT,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

Published:

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

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

(54) Title: METHODS FOR PROTEOMIC PROFILING USING NON-NATURAL AMINO ACIDS



(57) Abstract: The invention provides methods, reagents and systems for incorporating nonnatural amino acids into proteins, preferably *in vivo*, using the endogenous protein synthesis machinery of an organism. The incorporated non-natural amino acids contain reactive groups for further chemical reagents, which may serve as a "handle" to enrich the proteins or fragments thereof in a number of uses, such as proteomic analysis, imaging of diseased tissues / cells, etc.

5 Methods for Proteomic Profiling using Non-natural Amino Acids

Reference to Related Applications

 This application claims the benefit of the filing date of U.S. Provisional Applications U.S.S.N. 60/638856, filed on December 22, 2004; and U.S.S.N. 60/668471, filed on April 5, 2005. The teachings of the referenced applications are
10 incorporated herein by reference.

Government Support

 Work described herein was funded, in whole or in part, by Grant No. DAAD19-03-D-0004 (ARO) from the United States Army. The United States government has certain rights in the invention.

Background of the Invention

 Cells respond to fluctuations in their environment by changing the set of proteins they express. Understanding such changes is important for the understanding of cellular processes and for the understanding of how pharmaceuticals alter such expression patterns. Alterations in protein synthesis and
20 degradation enable cells (such as neurons), to adapt to changing external conditions.

 As a specific example, in neurons, there is increasing evidence that local dendritic protein synthesis is used to allow individual synapses to respond dynamically to the environmental changes that accompany the establishment, maintenance and plasticity of synaptic connections. It is now well established that
25 the essential components of the translational machinery as well as mRNAs are present at or near synapses. There is also now ample evidence that dendrites can synthesize new proteins. Nevertheless, only a handful of dendritically translated proteins have been identified thus far, largely through candidate-based approaches inspired by a particular laboratory's interest in a specific protein. The current
30 piecemeal accrual of knowledge about the dendritically synthesized proteome has hindered the understanding of the logic of local translation.

5 Most of the current understanding of the dendritic proteome comes not from protein data, but rather from catalogs of dendritic mRNAs. Yet, the dendritic mRNA population has been explored via either candidate based approaches (does the mRNA for protein “x” localize to dendrites) or mRNA generation from single neuritic process PCR experiments. A major shortcoming of these data is that this
10 approach only provides a rather short list of transcripts, not an exhaustive list of proteins. A focus on translated proteins is essential for an accurate picture of how the synapse and the dendrite can mount a response to local changes in the environment.

 This problem is far from being neuron-specific. In fact, there is a general
15 need to effectively determine the protein expression profiles, of especially newly synthesized proteins, in two samples, such as diseased sample vs. normal sample, a control sample and a drug treated sample, the same sample observed at two different time points, *etc.*

 Proteomic approaches, including differential 2D-gel electrophoresis
20 (Freeman & Hemby, *Proteomics for protein expression profiling in neuroscience. Neurochem Res* **29**: 1065-1081, 2004; Lilley & Friedman, All about DIGE: quantification technology for differential-display 2D-gel proteomics. *Expert Rev Proteomics* **1**: 401-409, 2004) (DIGE), isotope-coded affinity tags (ICAT) (Gygi *et al.*, Quantitative analysis of complex protein mixtures using isotope-coded affinity
25 tags. *Nat Biotechnol* **17**: 994-999, 1999), quantitative proteomic analysis using samples from cells grown in ¹⁴N or ¹⁵N-media (Washburn *et al.*, Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal Chem* **74**: 1650-1657, 2002) and stable isotope labeling by amino acids in cell culture (Andersen *et al.*, Nucleolar proteome dynamics. *Nature* **433**: 77-
30 83, 2005; Ong *et al.*, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**: 376-386, 2002) (SILAC), have been developed to compare the protein expression profiles of cells in different states. In addition, approaches focusing on posttranslational modifications (Chang *et al.*, Analysis of protein phosphorylation by
35 hypothesis-driven multiple-stage mass spectrometry. *Anal Chem* **76**: 4472-4483, 2004; Garcia *et al.*, Analysis of protein phosphorylation by mass spectrometry.

5 *Methods* **35**: 256-264, 2005; Peters *et al.*, Exploring the phosphoproteome with mass spectrometry. *Mini Rev Med Chem* **4**: 313-324, 2004) have emerged to explore the “phosphoproteome” and “glycoproteome” of a cell.

There are, however, drawbacks associated with approaches like DIGE, ICAT and SILAC. For example, 2D-gel electrophoresis has a limited ability to detect
10 proteins of low abundance, membrane proteins, or proteins with unusual isoelectric points or molecular weights. With ICAT, valuable protein information including post-translational modifications can be lost, as only cysteine-containing peptides are analyzed. SILAC critically depends on the ideally complete incorporation of a given “light” or “heavy” form of an essential amino acid into two cell pools, a process that
15 is expensive, time-consuming and restricted to cell lines which can be kept in an artificial medium. Common to all these approaches is a prerequisite for decreasing the sample complexity before mass spectrometrical analysis. This is accomplished by cutting out single bands or spots (in the case of DIGE and SILAC) or by the affinity-purification of biotinylated cysteine-containing peptides (ICAT). Most
20 importantly, however, these approaches lack the ability to enrich for newly synthesized proteins. Enrichment for newly synthesized proteins has at least two advantages: (i) it permits an analysis of the primary protein synthesis response to internal and external cues / stimuli, and, (ii) it decreases the complexity of the sample, potentially enabling the identification of proteins expressed at low levels. In
25 order to understand the primary response of cells to internal and external cues the ability to differentiate the newly synthesized proteome from the pre-existing protein pool is desirable. Currently available approaches using radioactively labeled amino acids can detect newly synthesized proteins but do not enrich for them. Furthermore, it is problematic to directly use radiolabeled samples for mass spectral analyses due
30 to radioactive contamination issues.

Finally, methods focusing specifically on mRNA expression profiles, only indirectly reflect the protein population in a sample, since not all mRNAs in the sample are necessarily translated at any given condition. Even for translated mRNAs, the degree of translation is often not predicted by the transcriptional
35 response.

- 5 Accordingly, there is a need for new methods, systems, and reagents for analyzing (new) protein expression.

Summary of the Invention

One aspect of the invention provides a method for determining the protein expression pattern in a cell or a tissue, comprising: (1) contacting the cell or the
10 tissue with a non-natural amino acid comprising a first reactive group, under a condition where the non-natural amino acid is incorporated into the newly synthesized proteins of the cell or the tissue; (2) contacting proteins or fragments thereof from the cell or the tissue with an affinity reagent comprising a second reactive group and one or more affinity moieties, wherein the first and second
15 reactive groups react to tag the non-natural amino acid with the affinity reagent; (3) isolating the newly synthesized proteins or fragments thereof comprising the non-natural amino acid labeled with the affinity reagent via the affinity moiety; and, (4) determining protein expression pattern by identifying the isolated proteins or fragments thereof.

20 In one embodiment, the method further comprises quantitating each identified / isolated protein or fragments thereof.

In one embodiment, the first reactive group is an azido group. For example, in one embodiment, the non-natural amino acid is azidoalanine, azidohomoalanine (AHA), azidonorvaline, or azidonorleucine.

25 In one embodiment, the first reactive group is a ketone or aldehyde moiety.

In one embodiment, the first reactive group is a diboronic acid moiety.

In one embodiment, the first reactive group is a terminal alkyne moiety.

In one embodiment, in step (1), the non-natural amino acid is incorporated *in vivo* by endogenous protein synthesis machinery of the cell or the tissue.

30 In one embodiment, in step (1), the non-natural amino acid is site-specifically incorporated in place of a natural amino acid selected from methionine or phenylalanine. In other embodiments, the non-natural amino acid is site-specifically incorporated in place of any other natural amino acids.

In one embodiment, the affinity reagent further comprises an antigenic
35 moiety that can be recognized by an antibody. For example, the affinity moiety may

5 be biotin, and the antigenic moiety may be an epitope tag, such as a FLAG tag, an HA tag, a His6 tag, *etc.*

In one embodiment, the FLAG tag comprises one or more cleavage sites for a sequence-specific protease, such as trypsin.

10 In one embodiment, the FLAG tag is situated between the affinity moiety and the second reactive group.

In one embodiment, the second reactive group and the affinity moiety are linked by one or more cleavable functional groups, such as photo-cleavable groups, chemically cleavable groups, or enzymatically cleavable groups.

15 In one embodiment, the second reactive group and the affinity moiety are linked by one or more cleavage sites for a sequence-specific protease, such as Factor Xa or PreScission Protease.

In one embodiment, the second reactive group and the affinity moiety are linked by a photo-cleavable linker.

20 In one embodiment, the fragments of the proteins are generated by protease digestion or chemical fragmentation. For example, the protease digestion may be performed using a protease that generates a C-terminal Arginine or Lysine, such as trypsin or endoproteinase Lys C. The chemical fragmentation may be performed by CNBr.

25 In one embodiment, the fragments of the proteins are generated either before or after the isolation step (2).

In one embodiment, identification of the target protein or fragments thereof is performed by mass spectrometry.

In one embodiment, the fragments are generated either before step (2) or after step (3).

30 In one embodiment, in step (1), the condition is such that the concentration of the non-natural amino acid is optimized to effect a pre-selected amount or percentage of incorporation.

In one embodiment, step (1) is carried out by contacting the tissue or the cell cultured *in vitro*.

35 In one embodiment, step (1) is carried out by administering the non-natural amino acid to an animal.

5 In one embodiment, in step (1), the cell or the tissue is further contacted with a second non-natural amino acid. The second non-natural amino acid may contain an isotope tag, such as a deuterated natural amino acid. The isotope may be one or more of ^2H , ^3H , ^{13}C , ^{14}C , ^{15}N , ^{18}O , or ^{35}S .

Another aspect of the invention provides a method of comparing protein
10 expression profiles of a first sample and a second sample, comprising: (1) using the method of claim 1, determining the protein expression profiles of the first and the second samples, (2) comparing the protein expression profile of the first sample with that of the second sample.

In one embodiment, the first sample is a control sample, and the second
15 sample is the control sample treated by an agent.

In one embodiment, the agent is a pharmaceutical agent.

Another aspect of the invention provides a method for monitoring protein
expression in a cell or a tissue, comprising: (1) contacting the cell or the tissue with
a non-natural amino acid comprising a first reactive group, under a condition where
20 the non-natural amino acid is incorporated into the newly synthesized proteins of the cell or the tissue; (2) contacting the cell or the tissue with a fluorescent reagent comprising a second reactive group and a fluorescent moiety or a reagent that can be subsequently coupled to a fluorescent reagent, wherein the first and second reactive groups react to label the non-natural amino acid with the fluorescent reagent, said
25 fluorescent reagent is substantially more fluorescent after the first and second reactive groups react, or unreacted fluorescent reagent is removed (*e.g.*, washed away) before detection by fluorescence imaging; (3) monitoring the intensity of fluorescence in the cell or the tissue, thereby monitoring protein expression in said cell or said tissue.

30 In one embodiment, the cell or the tissue is in a live animal.

Another aspect of the invention provides a functional reagent comprising: (1)
a reactive group for reacting with a non-natural amino acid incorporated into a
protein; (2) an affinity moiety or a fluorescent moiety; and, (3) one or more
cleavable groups situated between the reactive group and the affinity moiety or the
35 fluorescent moiety.

In one embodiment, the reactive group is an azido group.

- 5 In one embodiment, the reactive group is a ketone or an aldehyde moiety.
 In one embodiment, the reactive group is a terminal alkyne moiety.
 In one embodiment, the functional reagent is sufficiently membrane
permeable to diffuse into a live cell.
 In one embodiment, the functional reagent is substantially non-toxic.
- 10 In one embodiment, the reactive group reacts with a non-natural amino acid
incorporated in a protein.
 In one embodiment, the functional reagent further comprises an antigenic
moiety that can be recognized by an antibody. In one embodiment, the antigenic
moiety is an epitope tag, such as a FLAG tag, an HA tag, a His6 tag, *etc.*
- 15 In one embodiment, the FLAG tag comprises one or more cleavage sites for
a sequence-specific protease, such as trypsin.
 In one embodiment, the FLAG tag is situated between the affinity moiety
and the second reactive group.
 In one embodiment, the second reactive group and the affinity moiety are
- 20 linked by one or more cleavable functional groups, such as photo-cleavable groups,
chemically cleavable groups, or enzymatically cleavable groups.
 In one embodiment, the second reactive group and the affinity moiety are
linked by one or more cleavage sites for a sequence-specific protease, such as Factor
Xa or PreScission Protease.
- 25 In one embodiment, the affinity moiety is biotin.
 In one embodiment, the affinity moiety is biotin, wherein the FLAG tag
comprises two cleavage sites for a sequence-specific protease, and wherein the
FLAG tag is situated between the affinity moiety and the second reactive group.
 In one embodiment, the linker is a photo-cleavable linker.
- 30 The embodiments described above, including those described under different
aspects of the invention, are contemplated to be applicable for all aspects of the
inventions wherever appropriate.

Brief Description of the Figures

- Figure 1: Overview of the protein identification procedure using non-natural
35 amino acids, such as azidohomoalanine (AHA). Cells are challenged in the presence

5 of AHA to allow for protein synthesis with AHA-incorporation (“+/-” represents the presence / absence of a modulator of activity). After incubation, cells are lysed or undergo a subcellular fractionation for biochemical enrichment of specific cellular compartments followed by lysis. Lysates are then coupled to an alkyne-bearing affinity tag, followed by affinity chromatography to enrich for AHA-incorporated
10 proteins. Purified proteins are digested with a protease, and the resulting peptides are analyzed by tandem mass spectrometry to obtain experimental spectra. Different search programs are used to match the acquired spectra to protein sequences.

Figure 2: Generation of a tandem featured alkyne tag for two-step purification of AHA incorporated proteins. Structure of a Biotin-FLAG-alkyne
15 affinity reagent. Biotin (square), alkyne (square circle) as well as the tryptic cleavage sites (scissors) are indicated. The FLAG epitope DYKDDDDK (SEQ ID NO: 1) is separated from the biotin moiety by a short linker (GGA). (B) Western blot analysis for tandem purified biotinylated proteins using the Biotin-FLAG-alkyne tag. Cell lysates from both AHA and methionine incubated HEK293 cells
20 were subjected to [3+2]-cycloaddition with the Biotin-FLAG-alkyne tag and subsequently purified using monomeric avidin and anti-FLAG-M2 affinity matrices. SDS was used in both steps to elute the proteins from the resin. SDS was removed from avidin-eluates by ultrafiltration and neutralization with Triton X-100 before incubation with the FLAG-M2 affinity matrix. Sizes of marker proteins are indicated
25 on the left margin.

Detailed Description of the Invention

1. Overview

The invention provides methods and reagents for *in vitro* and/or *in vivo* labeling, detecting / monitoring, isolating, and/or identifying the translation products
30 in a cell or tissue. The invention can be easily applied to many proteomic profiling questions, regardless of tissue type or cellular context.

According to the methods of the invention, new proteins synthesized in a cell are labeled by contacting the cell with a non-natural amino acid that can be incorporated into proteins in the cell in place of a naturally occurring amino acid.
35 The non-natural amino acid generally has a first reactive moiety that can be used to

5 label proteins that incorporate such an amino acid residue. Generally, the first reactive moiety has a functional group that can be specifically reacted with a second reactive moiety on a reagent, which reagent further comprises a functional moiety, such as a fluorescent moiety, an affinity moiety, *etc.* Once the first reactive moiety on the non-natural amino acid reacts with the second reactive moiety on the reagent,
10 the functional group becomes associated with the protein containing the non-natural amino acid, thereby facilitating the labeling, detecting / monitoring, isolating, and/or identifying the translation products in a cell or tissue.

Preferably, the first and the second reactive moieties react with each other with relatively high specificity, such that the reaction may occur without undesirable
15 side-reactions or interferences due to the presence of other proteins without the non-natural amino acids (*e.g.*, reactions between the reagent and the unlabeled proteins).

The cells may be cultured *in vitro*, and the non-natural amino acid and the reagent are provided directly to the cell culture medium. The cells may even be lysed or partially permeabilized to facilitate the accessibility of the non-natural amino
20 acid and the reagent to the translation machinery. Alternatively, the cells may be cells of a tissue or a live organism, wherein the non-natural amino acid and the reagent are administered to the organism (*e.g.*, by feeding, injection, *etc.*).

In one embodiment, the non-natural amino acid is incorporated into proteins by the endogenous translation machinery of cells or tissue, including the endogenous
25 amino-acyl tRNA synthetases (AARS) that can take the non-natural amino acid as a substrate, and charge it to an endogenous tRNA; and the endogenous ribosome. Preferably, the AARS and/or the tRNA are for a natural amino acid to which the non-natural amino acid is a structural homolog.

Alternatively, the non-natural amino acid is incorporated into proteins of the
30 cells / tissue with at least partial aid from non-endogenous translation machinery. For example, the non-natural amino acid may be charged to a tRNA (endogenous, modified endogenous, or tRNA from a different species, *etc.*) by a non-endogenous AARS, which AARS may be engineered specifically to take the non-natural amino acid as a substrate. See, for example, US 2004-0053390 A1, US 2004-0058415 A1,
35 WO 03/073238 A2, U.S. Pat. No. 6,586,207 (all incorporated herein by reference). The charged non-natural amino acid – tRNA may be produced *in vitro* and provided

5 to the cell or tissue. Alternatively, the modified AARS and/or tRNA may be provided to the cells / tissue, for example, by introducing into the cells a polynucleotide encoding the AARS and/or tRNA, to charge the non-natural amino acid *in vivo*.

The non-natural amino acid preferably contains one or more of the following
10 desirable characteristics: (1) relatively permeable through bio-membranes, such as plasma membranes, such that it can be directly provided in tissue culture medium or administered to a live organism for direct uptake by the cell; (2) relatively stable *in vitro* and *in vivo*; (3) being a structural homolog of one or more natural amino acids, such as methionine or phenylalanine (In certain embodiments, the natural amino
15 acid is an essential amino acid, so that the cells can be made auxotrophic for that natural amino acid; in certain embodiments, the natural amino acid is relatively rarely used in proteins); (4) capable of being charged directly to an endogenous tRNA by an endogenous AARS; (5) the charged non-natural amino acid – tRNA complex can be readily incorporated into proteins by endogenous ribosomes; (6) the
20 structure of the non-natural amino acid is such that, upon incorporation into the proteins, it does not substantially affect the folding and/or biological function of any proteins incorporating the non-natural amino acid; and (7) being substantially non-toxic and biologically inert.

In certain embodiments, the first reactive moiety on the non-natural amino
25 acid preferably can react with the second reactive moiety of another reagent under relatively mild conditions, such as physiological conditions with a relatively neutral / physiological pH and temperature. One or more catalysts (preferably non-toxic) may also be provided to facilitate the reaction under such conditions.

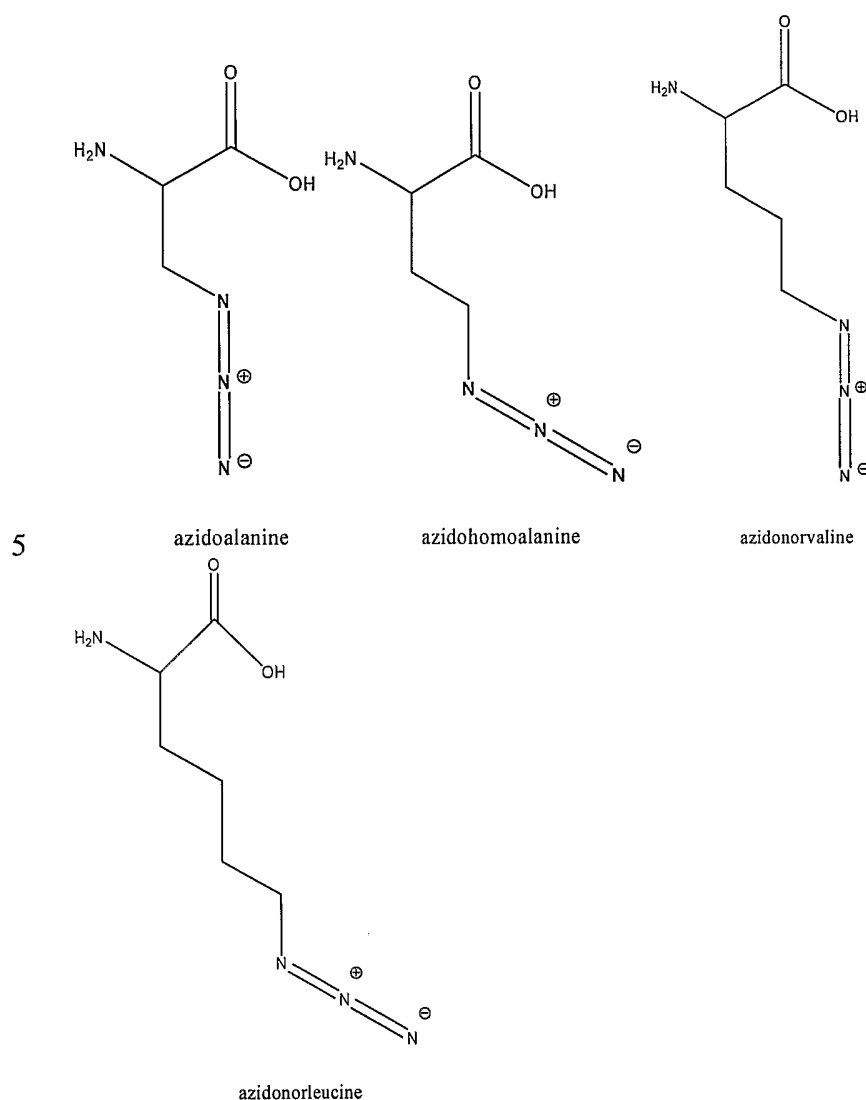
However, in certain other embodiments, the reaction may occur in more
30 harsh conditions, with any type of suitable catalyst, since proteins labeled by the non-natural amino acid may be isolated or be present in a cell lysate or an *in vitro* translation system before the reaction between the first and second reactive moieties occur.

In certain embodiments, cells / tissues incorporated with non-natural amino
35 acids may be fixed or permeated after the incorporation step, before the agent with the second reactive group and functional moieties are provided (*e.g.*, in the case of

- 5 immunostaining, *etc.*). Since the reaction between the first and the second reactive groups occur only after the fixation of cells / tissues, there are fewer limitations regarding the type of second reactive groups and/or catalysts that may be used in the methods of the invention.

- In certain embodiments, the methods of the invention can be used to
10 incorporate non-natural amino acids (*e.g.*, those with azide moiety, such as AHA, or those with terminal alkyne groups, *etc.*) into proteins / polypeptides translated *in vitro*. In these embodiments of the invention, either endogenous or ortholog AARS / tRNA pairs may be used to charge the non-natural amino acids. Since there is no issue of membrane permeability, and much less of an issue of toxicity (if any at all),
15 the range of non-natural amino acids and the types of catalysts that can be used in these embodiments are wider than those for intact cell / tissue use.

- In one embodiment, the amino acid used is azidohomoalanine (AHA) or an analog thereof. AHA is an analog of the essential natural amino acid methionine, which is relatively rarely used in proteins. Other methionine / AHA homologs that
20 can be used in the invention include: azidoalanine, azidonorvaline, and azidonorleucine, *etc.* See Link *et al.*, *J. Am. Chem. Soc.* **126**: 10598-10602, 2004 (incorporated herein by reference).



In another embodiment, the non-natural amino acid has a ketone moiety. In yet another embodiment, the non-natural amino acid has a diboronic acid moiety.

The concentration of the non-natural amino acid used to contact cells in culture (or the amount administered to an animal) can be selected to optimize the detection of newly synthesized proteins.

Once the non-natural amino acid is incorporated into the proteins of the cells, proteins from the cells are then treated with a reagent that can become associated with the non-natural amino acid. The reagent comprises a second reactive moiety that can react with the first reactive moiety of the non-natural amino acid. For example, when the non-natural amino acid has an azido moiety, the reagent will generally have an alkyne moiety, and vice versa.

5 In certain embodiments, the reagent is an affinity reagent that also comprises at least one affinity moiety, such as biotin, that can be used to isolate the non-natural amino acid-labeled proteins. In one embodiment, the affinity reagent further comprises a second affinity moiety, such as an antigenic moiety that can be recognized by an antibody. In a preferred embodiment, the affinity reagent has a
10 FLAG epitope that can also be used for affinity purification. In a preferred embodiment, the affinity reagent has both a FLAG tag and a biotin moiety. In such embodiments, the affinity moieties may be used independently or in combination (*e.g.*, sequential) to effect optimum isolation of the labeled protein through the affinity moieties.

15 As used herein, "affinity moiety" includes any moiety that may be used for affinity binding / isolation / purification purpose. Common affinity moieties include antigens (*e.g.*, epitope tags, such as FLAG tag, HA tag, His6 tag, *etc.*) for certain antibodies, a member of a ligand-receptor pair, *etc.* The affinity moiety may be polypeptide, nucleic acid, polysaccharide, lipids, vitamin (*e.g.*, biotin, *etc.*), or
20 molecules of any other chemical nature.

 In certain embodiments, the reagent is a fluorescent reagent that comprises the second reactive moiety and one or more fluorescent moieties. Preferably, the fluorescent moiety becomes more fluorescent when the first and the second reactive moieties react with each other, or the unreacted fluorescent reagent can be easily
25 removed (*e.g.*, washed away).

 In certain embodiments, the reagents may be cleavable. In one embodiment, the reagent harbors, in addition to any of the above-described components (such as an antigenic component), a cleavage site for a sequence-specific protease, such as Factor Xa or PreScission Protease. When used as an affinity tag, a first purification
30 can be performed that will select for the presence of the affinity reagent. For example, the proteins can be purified using a FLAG binding resin. Upon treatment with the specific protease, only proteins that have the non-natural amino acid will be released for subsequent analyses, while proteins that bind non-specifically to the affinity reagent or the resin / column will likely stay bound.

35 In another embodiment, the reagent may comprise a photo-cleavable linker between the second reactive moiety and the functional group. For example, a photo-

5 cleavable affinity reagent includes a photo-labile linker between the second reactive moiety and the one or more affinity moieties. Upon light exposure, non-natural amino acid-bearing molecules are specifically released and can be subsequently identified in various detection assays, such as mass spectrometric analyses.

Proteins or fragments thereof having the non-natural amino acid moiety can
10 be isolated using the one or more affinity moieties, if such moieties are present in the reagent comprising the second reactive moiety. The isolated proteins can be analyzed further. For example, such proteins or fragments can be identified by mass spectrometry techniques, with or without obtaining the sequences of the proteins / fragments. In certain embodiments, the isolated proteins can be digested using a
15 protease, such as trypsin, and the resulting peptides can be analyzed by mass spectrometry. Alternatively, proteins from the cell may be digested with protease before the resulting fragments are reacted with the second reactive groups. The resulting peptide fragment – affinity reagent complex can then be isolated through affinity column. Protein digestion may also be carried out after the reaction with the
20 second reactive moiety, but before the affinity purification / isolation step.

In certain embodiments, the cells may be treated with a second non-natural amino acid, and/or a natural amino acid derivative. In one embodiment, the natural amino acid derivative contains one or more isotopes such that the overall molecular weight of the amino acid derivative is different from that of a wild-type natural
25 amino acid. For example, the natural amino acid derivative may be a deuterated amino acid (alternatively, an amino acid that contains ^3H), which gives a mass shift in mass spectrum. In certain embodiments, the non-natural amino acid may also contain isotopes, such as ^{13}C , ^{15}N or ^{18}O . The presence of such isotopes may help to differentiate newly synthesized proteins from different samples, thus enabling
30 simultaneous analysis of multiple samples in, for example, high throughput experiments.

The methods of the invention can be used in general to profile the expression pattern of new proteins in two or more samples / cell lines / tissues, and compare such expression patterns.

35 In one embodiment, the methods of the invention may be used to study pharmaceutical impact on protein expression in a cell. According to such a method,

5 the protein expression pattern is determined according to the methods of the invention in the presence or absence of a pharmaceutical agent or a candidate drug.

In one embodiment, the methods of the invention may be used to monitor the changes in expression pattern, if any, of a sample over time. For example, a zero time-point expression pattern may be obtained before the sample is subject to certain
10 treatment. Expression patterns at later time points may be obtained and compared to the zero time point. Pulse labeling of new proteins using the method of the invention may be used to obtain samples from different time points.

The methods of the invention can also be used to metabolically label new protein synthesis in cells or tissues. Such method is particularly useful for real-time
15 imaging of local protein synthesis in cells / tissues. For example, in one embodiment, protein synthesis can be monitored within a living cell, such as one in a live organism or in tissue culture. According to such methods, the cell is contacted with a non-natural amino acid, such as AHA, that is incorporated in a protein in place of a naturally occurring amino acid by the cell's endogenous machinery. The
20 cells are then treated with a fluorescent reagent that has a second reactive moiety capable of reacting with the first reactive moiety on the non-natural amino acid residue to form a covalent bond or, alternatively, the cells are treated with a reagent that allows for sequential coupling of a fluorescent moiety. The reaction of the fluorescent tags with the proteins will result in fluorescent-labeled proteins. In a
25 preferred embodiment, the fluorescent reagent is substantially more fluorescent after reaction with the non-natural amino acid residue.

The invention also includes various reagents for use with the methods of the invention. In certain embodiment, the invention provides an affinity reagent that has a first reactive moiety capable of reacting with a non-natural amino acid, at least one
30 affinity group /moiety, and at least one cleavage site that allows the separation of the second reactive moiety and the affinity moiety. In a preferred embodiment, the affinity reagent has two affinity groups and two distinct cleavage sites. In one such embodiment, the affinity reagent has two affinity moieties, including a biotin group and an immunological / epitope tag, and a two peptide cleavage sites (such as two
35 protease cleavage sites for trypsin or trypsin-like proteases).

5 As used herein, the term “physiological conditions” is meant to encompass those conditions compatible with living cells, *e.g.*, predominantly aqueous conditions of a temperature, pH, salinity, *etc.* that are compatible with living cells.

 The term “aryl” as used herein means 5- and 6-membered single-aromatic radicals which may include from zero to four heteroatoms. Representative aryls
10 include phenyl, thienyl, furanyl, pyridinyl, (is)oxazolyl and the like.

 The term “lower alkyl”, alone or in combination, generally means an acyclic alkyl radical containing from 1 to about 10, preferably from 1 to about 8 carbon atoms and more preferably 1 to about 6 carbon atoms. Examples of such radicals include methyl, ethyl, n-propyl, isopropyl, n-butyl, isobutyl, sec-butyl, tert-butyl,
15 pentyl, iso-amyl, hexyl, octyl and the like.

2. **Non-natural Amino Acids**

 Numerous non-natural amino acids have been successfully incorporated into protein either *in vitro* (such as *in vitro* translation) or *in vivo* (*e.g.*, in live cells or animals). These non-natural amino acids usually contain one or more functional
20 groups not present in the twenty naturally occurring amino acids, thus incorporation of non-natural amino acids into proteins *in vivo* can provide biological materials with new chemical functions and improved physical properties. Examples include new posttranslational modification chemistry by introducing azide and ketone moieties into recombinant proteins, and novel strategies for engineering hyper-stable
25 proteins by incorporating fluorinated side chains.

 However, implementing such methods generally requires certain manipulation of the protein biosynthesis machinery. This may be accomplished by using a cell’s translational machinery in either a residue-specific (Link *et al.*, *Curr. Opin. Biotechnol.* **14**: 603–609, 2003) or a site-specific manner (Wang & Schultz,
30 *Angew. Chem. Int. Edn. Engl.* **44**: 34–66, 2004).

 Thus in one embodiment, residue-specific incorporation of non-natural amino acids into proteins involves replacement of a natural residue with a conservatively modified analog. The translational machinery is sufficiently tolerant of altered substrates that, especially in the absence of competing natural substrates,
35 the modified residue is converted to an aminoacyl tRNA that is subsequently used

5 by the ribosome. Alternatively, the specificity of certain endogenous or designed / engineered exogenous AARS may be altered to catalyze the attachment of the non-natural amino acids to suitable tRNAs.

By these mechanisms, non-natural amino acids bearing bioorthogonal chemical moieties can be introduced into proteins that are overexpressed in a host
10 cell. To avoid competition with the endogenous amino acid, the host cell (*e.g.*, bacterial strain) may be rendered auxotrophic for the natural amino acid. Proteins cannot be overexpressed unless the cells are supplemented with either that residue or a closely related unnatural analog. For example, a phenylalanine auxotroph was used to express proteins in which all phenylalanine residues were replaced with p-
15 azidophenylalanine or p-acetylphenylalanine (a keto derivative) (Datta *et al.*, *J. Am. Chem. Soc.* **124**: 5652–5653, 2002; Kirshenbaum, *ChemBioChem* **3**: 235–237, 2002). Similarly, a methionine auxotroph was used for production of proteins that contained homopropargylglycine or azidohomoalanine at sites that encode for methionine (van Hest, *J. Am. Chem. Soc.* **122**: 1282–1288, 2000; Kiick *et al.*, *Proc.*
20 *Natl. Acad. Sci. USA* **99**: 19–24, 2002). Applicants have also extended this work to the labeling of bacterial cell surfaces (Link, *J. Am. Chem. Soc.* **126**: 10598–10602, 2004; Link & Tirrell, *J. Am. Chem. Soc.* **125**: 11164–11165, 2003). Azido amino acids were installed in outer membrane protein C (OmpC) of an *E. coli* methionine auxotroph and the cell surface azides were then ligated with alkyne probes through
25 both copper(I)-mediated and strain-promoted [3+2] cycloaddition.

Residue-specific metabolic labeling can produce proteins with multiple copies of a bioorthogonal functional group, and is particularly useful for the proteome-wise expression profiling. However, this has only limited application in cases where a chemical moiety is desired at a single position within the protein.
30 Thus site-specific insertion of a bioorthogonal amino acid may be achieved using nonsense suppression techniques (Wang & Schultz, *Angew. Chem. Int. Edn. Engl.* **44**: 34–66, 2004). In this approach, a mutually selective tRNA and aminoacyl-tRNA synthetase are developed so that the non-natural amino acid can be uniquely activated by the tRNA *in vivo*. The tRNA's anticodon is engineered to complement a
35 rare stop codon, which is co-opted to encode the non-natural amino acid in the corresponding DNA (and intermediate mRNA). Cells transfected with genes

5 encoding the engineered tRNA, aminoacyl-tRNA synthetase and target protein will produce the modified protein when supplemented with the non-natural amino acid.

The non-natural amino acid mutagenesis method has been used to introduce chemical moieties into proteins in both *E. coli* and yeast. For example, m-acetylphenylalanine was site-specifically incorporated into LamB, an outer-
10 membrane protein of *E. coli*, and subsequently labeled with membrane-impermeant hydrazide dyes (Zhang *et al.*, *Biochemistry* **42**: 6735–6746, 2003). Similarly, azido and alkynyl amino acids related to tyrosine were installed in proteins within both *E. coli* and yeast. After cell lysis, the derivatized proteins were tagged by copper-catalyzed [3+2] cycloaddition.

15 Regardless of how the non-natural amino acid is incorporated into proteins, in certain embodiments, it is preferred that the non-natural amino acids contain one or more of the following desirable characteristics to facilitate easy incorporation into live cells *in vitro* or *in vivo*: (1) relatively permeable through bio-membranes, such as plasma membranes, such that it can be directly provided in tissue culture medium
20 or administered to a live organism for direct uptake by the cell; (2) relatively stable *in vitro* and *in vivo*; (3) being a structural homolog of one or more natural amino acids, such as methionine or phenylalanine (In certain embodiments, the natural amino acid is an essential amino acid, so that the cells are auxotrophic for that natural amino acid; in certain embodiments, the natural amino acid is relatively
25 rarely used in proteins); (4) capable of being charged directly to an endogenous tRNA by an endogenous AARS; (5) the charged non-natural amino acid – tRNA complex can be readily incorporated into proteins by endogenous ribosomes; (6) the structure of the non-natural amino acid is such that, upon incorporation into the proteins, it does not substantially affect the folding and/or biological function of any
30 proteins incorporating the non-natural amino acid; and (7) being substantially non-toxic and biologically inert. The non-natural amino acid is typically relatively small in size, and contains a single functional reactive moiety.

In one preferred embodiment of the invention, the non-natural amino acid is incorporated in the place of a naturally occurring amino acid. In general, such non-
35 natural amino acid will be an acceptable substrate for an aminoacyl-tRNA synthetase (AARS) that charges a tRNA recognizing a naturally occurring codon.

5 Preferably, the AARS and/or the tRNA are endogenous AARS and endogenous tRNA.

For *in vivo* use, or *in vitro* use in live cells, the non-natural amino acid chosen should not be significantly toxic to cells.

10 However, in certain other embodiments, the methods of the invention also can be utilized in cells, tissues or organisms engineered to express an aminoacyl-tRNA that charges a natural tRNA with the non-natural amino acid. The methods of the invention may further be used in *in vitro* translation systems where cell lysates are used to support protein synthesis. In those embodiments, the requirement of membrane permeability becomes largely irrelevant, and one or more ortholog AARS
15 / tRNA pairs may be provided to the lysate to support incorporation of the non-natural amino acid.

There are several preferred reactive moiety pairs that are particularly suitable for the invention, which are illustrated below as non-limiting examples.

(A) Azide – Alkyne Ligation

20 In one specific embodiment, the protein tagging is based on the azide-alkyne ligation. In a typical embodiment, the side chain of the non-natural amino acid contains the azide group, while the second reactive moiety on the reagent contains the terminal alkyne moiety. Alternatively, the side-chain of the non-natural amino acid contains the terminal alkyne, while the second reactive moiety contains the
25 azide moiety.

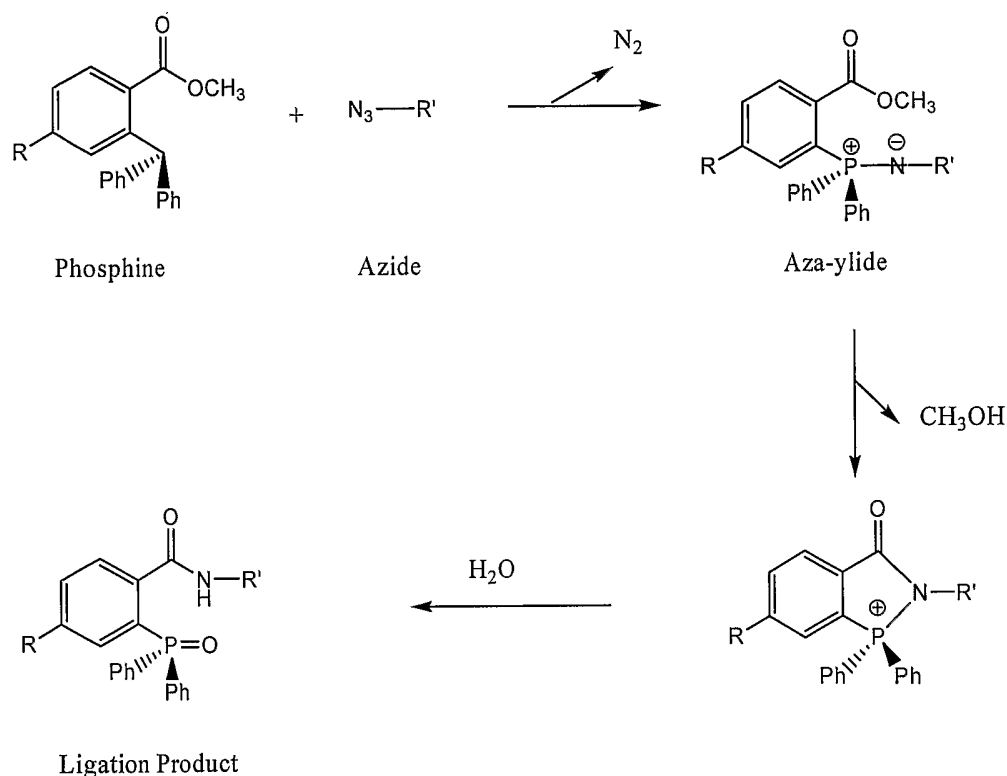
Several non-natural amino acids contain the azide group and are suitable for the instant invention. For example, azidoalanine, azidohomoalanine (AHA), azidonorvaline, or azidonorleucine are all azide-containing methionine homologs that can be incorporated into proteins using the endogenous protein translation
30 machinery. AHA, which serves as a surrogate for the essential amino acid methionine during protein synthesis, exhibits excellent membrane permeability and is not toxic even in primary neuronal cell cultures. Studies in *E. coli* and experiments in mammalian cells have shown no evidence of increased protein degradation upon introduction of AHA, indicating that the modified amino acid is
35 not likely to cause severe protein misfolding.

5 Azides are suitable for labeling all classes of biomolecules (including proteins) in any biological locale. This versatile functional group is absent from nearly all naturally occurring species. (Only one naturally occurring azido metabolite has been reported to date, isolated from unialgal cultures. See Griffin, *Prog. Med. Chem.* **31**: 121–232, 1994). Azides do not react appreciably with water
10 and are resistant to oxidation. Additionally, azides are mild electrophiles that do not react with amines or the other “hard” nucleophiles that are abundant in biological systems. Rather, they require “soft” nucleophiles for reaction. Although azides are susceptible to reduction by free thiols, including the ubiquitous cellular reductant, glutathione, reactions between monothiods and alkyl azides typically require
15 vigorous heating (100°C for several hours) or auxiliary catalysts.

Azides are prone to decomposition at elevated temperatures, but they are quite stable at physiological temperatures (See Griffin, *Prog. Med. Chem.* **31**: 121–232, 1994). Whereas aryl azides are well-known photocrosslinkers, alkyl azides do not photodecompose in the presence of ambient light. Finally, although azide anion
20 (for example, in the form of NaN_3) is a widely used cytotoxin, organic azides have no intrinsic toxicity. Indeed, organic azides are components of clinically approved drugs such as AZT (See Griffin, *Prog. Med. Chem.* **31**: 121–232, 1994).

Although kinetically stable, azides are predisposed to unique modes of reactivity owing to their large intrinsic energy content. This feature has been
25 exploited for the development of bioorthogonal reactions, including the Staudinger ligation of azides with functionalized phosphines and the [3+2] cycloaddition of azides with activated alkynes. These reactions can be used for the selective labeling of azide-functionalized biomolecules.

Staudinger ligation. In 1919, Hermann Staudinger reported that azides react
30 with triphenylphosphines (soft nucleophiles) under mild conditions to produce azaylide intermediates (Staudinger & Meyer, *Helv. Chim. Acta* **2**: 635–646, 1919). Bertozzi *et al.* modified the classic Staudinger reaction by introduction of an intramolecular trap into the phosphine (Saxon & Bertozzi, *Science* **287**: 2007–2010, 2000). Now known as the Staudinger ligation, this transformation ultimately
35 produces a covalent link between one nitrogen atom of the azide and the triarylphosphine scaffold (see below).



5

The Staudinger Ligation

The Staudinger ligation can be used to covalently attach probes to azide-bearing biomolecules. Like the azide, phosphines do not react appreciably with biological functional groups and are therefore also bioorthogonal. Additionally, the reaction proceeds readily at pH 7 with no apparent toxic effects. These and other applications of the Staudinger ligation have been recently reviewed (see Kohn & Breinbauer, *The Staudinger ligation—a gift to chemical biology. Angew. Chem. Int. Edn. Engl.* **43**: 3106–3116, 2004).

Copper-catalyzed [3+2] azide-alkyne cycloaddition. In the context of the Staudinger ligation, the azide serves as an electrophile subject to reaction with soft nucleophiles. Azides are also 1,3-dipoles that can undergo reactions with dipolarophiles such as activated alkynes. These π -systems are both extremely rare and inert in biological systems, further enhancing the bioorthogonality of the azide along this reaction trajectory.

The [3+2] cycloaddition between azides and terminal alkynes to provide stable triazole adducts was first described by Huisgen more than four decades ago (Huisgen, *Angew. Chem. Int. Edn. Engl.* **2**: 565–598, 1963). The reaction is

5 thermodynamically favorable by 30–35 kcal/mol. The reaction typically requires alkyne activation, which process requires elevated temperatures or pressures that are not compatible with living systems. Although this may not be a concern for *in vitro* use, activation may be achieved with catalysts for use *in vitro* as well as *in vivo*.

In one embodiment, the activation may be achieved by the addition of
10 electron-withdrawing groups, such as esters, to the alkyne.

In another embodiment, a Cu(I)-based catalyst may be used to accelerate the rate of cycloaddition between azides and alkynes by about 10⁶-fold (Rostovtsev *et al.*, *Angew. Chem. Int. Edn. Engl.* **41**: 2596–2599, 2002; Tornøe, *J. Org. Chem.* **67**: 3057–3064, 2002). This copper-catalyzed reaction, termed “click” chemistry,
15 proceeds readily at physiological temperatures and in the presence of biological materials to provide 1,4-disubstituted triazoles with nearly complete regioselectivity (Kolb & Sharpless, *Drug Discov. Today* **8**: 1128–1137, 2003). The copper-mediated reaction has been used to tag azides installed within virus particles (Wang *et al.*, *J. Am. Chem. Soc.* **125**: 3192–3193, 2003), nucleic acids (Seo *et al.*, *Proc. Natl. Acad. Sci. USA* **101**: 5488–5493, 2004) and proteins from complex tissue lysates (Speers &
20 Cravatt, *ChemBioChem* **5**: 41–47, 2004) with virtually no background labeling.

It should be noted that the same reaction can be carried out in reverse - using the alkyne as the first reactive moiety on the non-natural amino acid. Like the azide, a terminal alkyne consists of a mere three atoms.

25 The primary advantage of the catalyzed azide-alkyne cycloaddition over the Staudinger ligation is its faster rate. The copper-catalyzed reaction of azides with alkynes reportedly proceeds at least 25 times faster than the reaction of azides with triarylphosphines in cell lysates. Accordingly, “click” chemistry is preferably used in situations that require detection of very small quantities of azide-labeled
30 biomolecules.

In certain embodiments, an improved protocol for copper-catalyzed triazole formation may be used to further increase the efficiency of reaction by about 10-fold (Link *et al.*, *J. Am. Chem. Soc.* **126**: 10598 – 10602, 2004, incorporated herein by reference).

5 According to this embodiment of the invention, instead of generating Cu(I) *in situ* by combining Cu(II) and a reducing agent (such as 200 μ M CuSO₄ and 200 μ M of tris-(carboxyethyl)phosphine (TCEP)), ultra-pure Cu(I) is added directly to the reaction mixture as catalyst. In a preferred embodiment, the ultra-pure Cu(I) is CuBr, with at least 99.999% purity (such as those obtained from Aldrich).

10 **Strain-promoted cycloaddition.** An alternative means of activating alkynes for catalyst-free [3+2] cycloaddition with azides involves the use of ring strain (Agard *et al.*, *J. Am. Chem. Soc.* **126**: 15046–15047, 2004). Constraining the alkyne within an eight-membered ring creates ~18 kcal/mol of strain, much of which is released in the transition state upon [3+2] cycloaddition with an azide (Turner *et al.*,
15 *J. Am. Chem. Soc.* **95**: 790–792, 1972). As a consequence, cyclooctynes react with azides at room temperature, without the need for a catalyst. This strain-promoted cycloaddition has been used to label biomolecules both *in vitro* and on cell surfaces without observable toxic effects (Agard *et al.*, *J. Am. Chem. Soc.* **126**: 15046–15047, 2004). The rate of the strain-promoted cycloaddition can be increased by appending
20 electron-withdrawing groups to the octyne ring.

(B) **Ketones and Aldehydes**

 A second group of non-natural amino acids that can be incorporated into proteins according to the method of the instant invention includes those with a ketone or aldehyde side chain. Comprising only a handful of atoms, ketones and
25 aldehydes are bioorthogonal chemical moieties that can tag not only proteins, but also glycols and other secondary metabolites. These mild electrophiles are attractive choices for modifying biomolecules as they are readily introduced into diverse scaffolds, absent from endogenous biopolymers and essentially inert to the reactive moieties normally found in proteins, lipids and other macromolecules. Although
30 these carbonyl compounds can form reversible Schiff bases with primary amines such as lysine side chains, the equilibrium in water favors the carbonyl. By contrast, the stabilized Schiff bases with hydrazide and amino groups (hydrazones and oxides, respectively) are favored in water and are quite stable under physiological conditions (Jencks, *J. Am. Chem. Soc.* **81**: 475–481, 1959).

35 In one embodiment, ketone (and aldehyde) condensations are used for chemical modifications in the presence of cultured cells.

5 In one embodiment, ketone (and aldehyde) condensations are used to label proteins on cell surfaces or in the extra cellular environment.

In other embodiments, ketone (and aldehyde) condensations are used in the context of selectively labeling tissues in living organisms.

3. Reagents with Second Reactive Moiety

10 The reagents with the second reactive moiety of the invention comprise at least two moieties: a second reactive group that can specifically react with the first reactive moiety of the non-natural amino acid, and a functional moiety that becomes attached to the incorporated non-natural amino acid after the reaction. Optionally, a cleavable functional group is situated between the second reactive moiety and the
15 functional moiety to allow separation of these two moieties under controlled conditions (*e.g.*, protease digestion, photo-cleavage, *etc.*).

A number of possible functional moieties can be linked to the second reactive moiety. For example, the functional moiety may be an affinity moiety that can be used to isolate the labeled proteins or fragments thereof.

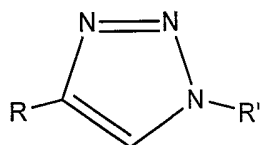
20 Alternatively, any other functional groups, such as a radioactive moiety (*e.g.*, radioisotope), a fluorescent moiety, *etc.*, may also be attached to the labeled protein or fragments thereof containing the non-natural amino acids.

(A) Second Reactive Moiety

In one embodiment, when the non-natural amino acid used is AHA or
25 another azide-bearing non-natural amino acid, proteins bearing one or more azide groups can subsequently be tagged for enrichment by treatment with an alkyne-bearing affinity reagent. Alternatively, as described above, the affinity reagent can have a phosphine group for reacting with the azide moiety via a Staudinger reaction (Saxon and Bertozzi, Cell Surface Engineering by a Modified Staudinger Reaction, *Science* **287**: 2007-2010, 2000, incorporated herein by reference).
30

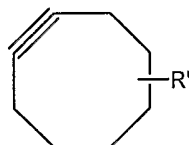
As used herein, "azide" group refers to $R-N_3$, wherein R represents the rest of the azide-containing molecule (comprising the non-natural amino acid side chain).

"Terminal alkyne" refers to $R'-C\equiv C$, wherein the [3+2] cycloaddition
35 product of the $R-N_3 + R'-C\equiv C$ reaction has a general formula of:

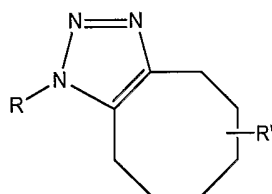


5

When “strain-promoted cycloaddition is used, the terminal alkyne in the above reaction is replaced by the following reactive partner:



In this case, the final ligation product of the cycloaddition is:



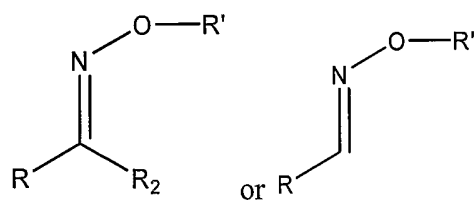
10

In other embodiments, where the non-natural amino-acid has a ketone group, the affinity reagent will preferably have an aminooxy or hydrazide moiety ($\text{NH}_2\text{-NH-CO-R'}$) (See, for example, U.S. Pat. Appln. No. 20020016003 to Bertozzi).

As used herein, “ketone” group refers to the general formula: R-CO-R_2 ;

15 “aldehyde” group refers to the general formula R-CHO .

“Aminooxy” moiety refers to $\text{NH}_2\text{-O-R'}$. The reaction product between the aminooxy group and the ketone / aldehyde group is:



(B) Functional Moiety

20 In one embodiment, the functional moiety is an affinity moiety, and thus the reagent is an affinity reagent. The affinity reagent has at least one affinity moiety that can be used to purify labeled proteins. In one embodiment, the affinity group is a biotin moiety. In another embodiment, the affinity group is an immunological or epitope tag. In a preferred embodiment, the affinity reagent has two or more affinity groups, thus allowing tandem purification of the labeled proteins. Tandem

25

5 purification decreases contamination by proteins that non-specifically adhere to the matrices or support material (such as resin) used for purification.

The affinity reagent can have one or more cleavable sites that allow for release of the purified proteins from the affinity matrix. The cleavable sites can be peptide bonds cleavable by proteases, or can be chemical cleavage or photo-labile
10 sites built into the affinity reagent. In a preferred embodiment, the cleavage sites are peptide sequences that can be specifically cleaved by proteases such as trypsin.

In certain embodiments, the cleavable sites are at or adjacent to the epitope tag. For example, in Figure 2, the FLAG tag sequence contains two trypsin cleavage sites that allow for the cleavage of the FLAG sequence, thereby releasing the
15 attached protein or fragment incorporating the non-natural amino acid.

Other exemplary functional moieties include, but are not necessarily limited to, fluorescent molecules or tags (*e.g.*, auto-fluorescent molecules, molecules that fluoresce upon contact with a reagent, *etc.*), radioactive labels (*e.g.*, ^{111}In , ^{125}I , ^{131}I , ^{212}B , ^{90}Y , ^{186}Rh , and the like); biotin (*e.g.*, to be detected through reaction of biotin
20 and avidin); imaging reagents (*e.g.*, those described in U.S. Pat. No. 4,741,900 and U.S. Pat. No. 5,326,856), and the like. Functional moieties may also include peptides or polypeptides that can be detected by antibody binding, *e.g.*, by binding of a detectably labeled antibody or by detection of bound antibody through a sandwich-type assay.

25 In one embodiment, an engineered phosphine may act as an azide-selective phosphine dye. In this embodiment, the dye remains substantially undetectable until reaction with an azide. In general, the phosphorous lone pair renders the dye substantially undetectable (*e.g.*, substantially non-fluorescent) until reaction with an azide, when the lone pair is removed from conjugation by formation of a phosphine
30 oxide. The unmasked dye provides for a detectable signal following reaction with the azide of the target molecule according to the invention. This reaction is described in detail in paragraphs [0078] – [0085] of US 2002-0016003 A1 (incorporated herein by reference). Briefly, the masked dye generally comprises an aryl group substituted with R_1 and PR_2R_3 ; wherein R_1 is an electrophonic group to trap (*e.g.*, stabilize) an aza-ylide group, including, but not necessarily limited to, a
35 carboxylic acid, an ester (*e.g.*, alkyl ester (*e.g.*, lower alkyl ester, benzyl ester), aryl

5 ester, substituted aryl ester), aldehyde, amide, *e.g.*, alkyl amide (*e.g.*, lower alkyl amide), aryl amide, alkyl halide (*e.g.*, lower alkyl halide), hoister, colony ester, alkyl ketone (*e.g.*, lower alkyl ketone), aryl ketone, substituted aryl ketone, halosulfonyl, nitrile, nitro and the like; R_2 and R_3 are generally aryl groups, including substituted aryl groups, or cycloalkyl groups (*e.g.*, cyclohexyl groups) where R_2 and R_3 may be
10 the same or different, preferably the same; R_1 and PR_2R_3 are preferably in the ortho position relative to one another on an aryl ring of the dye, or in an equivalent position that provides for the reaction to proceed according to the invention; X represents a target molecule having an azide N_3 ; and where R_1' represents a modified R_1 group following reaction with the azide of the target molecule via the
15 Staudinger ligation reaction described herein.

In one exemplary embodiment, the phosphine dye is a fluorescein derivative, which may be in unacetylated or acetylated (cell permeable) form. Three phosphine dyes are described in detail in paragraphs [0086] – [0089] of US 2002-0016003 A1 (incorporated herein by reference). These phosphine dyes can be used to detect an
20 azide on any molecule, such as those proteins incorporating AHA. The phosphine dyes can be used to detected biomolecules having an azide either at the cell surface or within cells.

(C) Cleavable Functional Group

The reagents with the second reactive moiety of the invention may optionally
25 contain one or more cleavable functional groups to allow separation of the second reactive moiety and the functional moiety under one or more controlled conditions.

As used herein, a “cleavable functional group” or “cleavable linker” is a chemical group that can be cleaved by a variety of methods, including input of energy, a chemical, an enzyme, and the like. For use in methods of the invention, the
30 cleavable functional group is generally specific, that is, one which can be specifically cleaved without altering or damaging the molecule being cleaved or which relatively uniformly alters the molecule in a reproducible manner. For example, the cleavable functional group can be a photo-cleavable group. In such a case, the photo-cleavable group is generally cleaved at a wavelength of light that
35 does not damage the molecule being released, for example, in the ultraviolet to visible range. Exemplary photo-cleavable linkers include, for example, linkers

5 containing o-nitrobenzyl, desyl, trans-cinnamoyl, m-nitrophenyl, benzylsulfonyl groups and the like (see, for example, Dorman and Prestwich, *Trends Biotech.* **18**: 64-77, 2000; Greene and Buts, "Protective Groups," in *Organic Synthesis*, 2nd ed., John Wiley & Sons, New York, 1991; U.S. Patent Nos. 5,143,854; 5,986,076; 5,917,016; 5,489,678; 5,405,783).

10 The cleavable functional group can also be a chemically cleavable group cleavable by a chemical such as an acid or base. If desired, a chemically cleavage reaction can be carried out under relatively mild conditions in which the chemically cleavable group is essentially the only chemical bond cleaved. A chemically
15 can cleave a methionine residue. CNBr can be particularly useful for releasing a molecule if a chemically cleavable group such as methionine has been added to the molecule, particularly in a polypeptide that does not have a methionine residue, or when the methionine residues have been replaced by non-natural amino acids.

Suitable chemically cleavable groups are well known to those skilled in the
20 art (see, for example Wilson and Czarnik, eds., *Combinatorial Chemistry: Synthesis and Application*, John Wiley & Sons, New York, 1997; Merrifield, *J. Am. Chem. Soc.* **85**: 2149, 1964; Bodanszky, *Principles of Peptide Synthesis*, Springer-Verlag, 1984; Houghten, *Proc. Natl. Acad. Sci., USA* **82**: 5131, 1985). Exemplary chemically cleavable linkers can contain a disulfide, which can be cleaved with
25 reducing agents; a diol, which can be cleaved with periodate; a diazo bond, which can be cleaved with dithionate; an ester, which can be cleaved with hydroxylamine; a sulfone, which can be cleaved with base, and the like (see Hermanson, *Bioconjugate Techniques*, Academic Press, San Diego, 1996; Pierce Chemical Co., Rockford IL).

30 The cleavable functional group can also be an enzymatically cleavable group. For example, a protease can be used to cleave a cleavable functional group having a suitable recognition sequence for the protease.

Particularly useful proteases are endopeptidases such as factor Xa, tobacco etch virus (TEV) protease, trypsin, chymotrypsin, *Staphylococcus aureus* protease,
35 submaxillaris protease, and the like. The protease can be selected based on the incorporation of a particular cleavable recognition sequence as a functional group.

5 Other considerations for selecting a protease include the presence or absence of a recognition sequence in the molecule being captured and released. For example, a rare cleaving protease such as TEV protease or factor Xa can be used to cleave a functional group containing the corresponding protease recognition sequence, resulting in release of the captured molecule.

10 Such rare cleaving proteases are particularly useful for releasing an intact polypeptide molecule since the recognition sequence for these proteases would not occur in the vast majority of polypeptides. Alternatively, a polypeptide sample can be treated with a specific protease, and the digested peptides isolated by the methods disclosed herein. In such a case, the captured peptides would not contain a
15 recognition sequence for the protease used for cleavage since the polypeptide has already been digested.

In addition, in certain embodiments, an intact polypeptide can be captured and digested with a protease after binding to the solid support, resulting in the incorporation and release of a label on the peptide fragment of the polypeptide that
20 was captured on the solid support. Thus, protease digestion can be used before or after capture of a sample molecule, in particular polypeptide sample molecules, as desired.

In addition to proteases, a cleavable functional group can be a recognition sequence for an endonuclease such as a restriction enzyme. Thus, an appropriate
25 recognition sequence for a restriction enzyme can be incorporated as a cleavable functional group and cleaved with the respective restriction enzyme. It is understood that such a nucleotide functional group can be useful for capturing and releasing a nucleic acid or a polypeptide, or any other type of molecule, as desired. Similarly, a protease recognition sequence can be useful for capturing and releasing a
30 polypeptide, nucleic acid or any other type of molecule, as desired.

As used herein, the term "isotopic label" or "isotope tag" refers to a chemical group which can be generated in two distinct isotopic forms, for example, heavy and light isotopic versions of the constituent elements making up the chemical group. Such constituent elements include, for example, carbon (*e.g.*, ^{13}C or ^{14}C), oxygen
35 (*e.g.*, ^{18}O), hydrogen (*e.g.*, ^2H or ^3H), nitrogen (*e.g.*, ^{15}N), and sulfur (*e.g.*, ^{35}S). In addition, other elements that are chemically or functionally similar can be

5 substituted for the above naturally occurring elements. For example, selenium can be used as a substitute for sulfur. Particularly useful isotopic labels or tags are those that allow convenient analysis by MS. In such cases, the isotope is preferably non-radioactive or with low radioactivity. For example, heavy and light isotopic versions of an amino acid can be used to differentially isotopically label a polypeptide.

10 4. Detection and Analysis

After tryptic digestion, mass spectral analysis may be used to analyze the eluted non-natural amino acid-containing proteins or fragments thereof.

In one embodiment, the eluted proteins may be analyzed using conventional techniques, such as 2-D electrophoresis.

15 In a preferred embodiment, detection and analysis is achieved by utilizing MudPIT (Multi-dimensional Protein Identification Technology) or related advanced mass spectrometrical techniques. MudPIT is a technique for the separation and identification of complex protein and peptide mixtures. MudPIT is another approach developed to alleviate many of the disadvantages associated with two-dimensional
20 gel electrophoresis. MudPIT uses two chromatography steps interfaced back to back in a fused silica capillary. More specifically, MudPIT uses columns consisting of strong cation exchange (SCX) material back-to-back with reversed phase (RP) material inside fused silica capillaries. The chromatography proceeds in cycles, each comprising an increase in salt concentration to “bump” peptides off of the SCX
25 followed by a gradient of increasing hydrophobicity to progressively elute peptides from the RP into the ion source. In other words, chromatography proceeds in steps with increases in salt concentration used to free peptides from the cation-exchange resin, after which they bind to a reversed phase resin. A typical reversed phase gradient to increasing hydrophobicity is then applied to progressively elute peptides
30 from the reversed phase packing into the mass spectrometer. Typically, this mass spectrometer will be a tandem electrospray (ESI-MS/MS), so peptides are ionized in the liquid phase, separated in a primary mass spectrometer, broken up using collision-induced dissociation (CID) and analyzed again.

The advantage of this process is that the band broadening associated with
35 many chromatographic steps is avoided, thus avoiding loss of resolution, and

5 preventing components from running into one another. In addition, the capillary can be placed directly into the ion source of a mass spectrometer to maximize sensitivity.

Raw mass spectral data are filtered and analyzed using art-recognized methods, such as SEQUEST, DTASelect and Contrast algorithms, which allow for
10 an efficient and comprehensive interpretation and comparison of the proteomic data. Specifically, the mass spectrometer's data-dependent acquisition isolates peptides as they elute, and subjects them to Collision-Induced Dissociation, recording the fragment ions in a tandem mass spectrum. These spectra are matched to database peptide sequences by the SEQUEST algorithm. SEQUEST's peptide identifications
15 are assembled and filtered into protein-level information by the DTASelect algorithm.

In a typical setting, a Sutter Instrument Company P-2000 is used to produce tips from fused silica capillaries, which are usually purchased from Polymicro or Agilent. Apertures are approximately 5 nanometers across from capillaries that have
20 an inner diameter of 100 microns. The material that is loaded into the SCX columns can be made from 5 micron spherical silica beads. For RP, 5 micron C-18 coated beads from a variety of commercial vendors can be used. Agilent 1100 and ThermoFinnigan Surveyor Quaternary pumps may be operated at flow rates of 100-200 microliters/min, with pre-column splitting of the flow to produce about 100-200
25 nL/min flow rates at the column.

SEQUEST is a program that correlates uninterpreted tandem mass spectra of peptides with amino acid sequences from protein and nucleotide databases. SEQUEST will determine the amino acid sequence and thus the protein(s) and organism(s) that correspond to the mass spectrum being analyzed. SEQUEST is
30 distributed by Thermo Finnigan, A division of Thermo Electron Corporation. SEQUEST uses algorithms described in U.S. patents 6,017,693 and 5,538,897.

While SEQUEST is very powerful for matching uninterpreted tandem mass spectra to database peptide sequences, DTASelect is designed to reassemble this peptide information into usable protein information. The DTASelect program can do
35 more than simply report the protein content of a sample; it features many customizable filters to specify which identifications should be kept and which

5 discarded. It also features reports to investigate post-translational modifications, align sequences of peptides to identify poorly sequenced regions, and analyze chromatography efficiency. The software makes the process of analyzing SEQUEST results much faster, more powerful, and more consistent than possible before, even for data sets containing a million spectra or more. By automating SEQUEST
10 analysis, DTASelect enables experiments of far greater scope. Alternatively, programs like PeptideProphet and ProteinProphet can be used for data processing after the SEQUEST algorithm. Programs mentioned above and in the following paragraph are only examples for performing the data analysis.

Differentiating biological samples by protein content is an important
15 application of proteomics. The CONTRAST program uses the filters present in DTASelect to highlight the most important identifications of samples and then compares them. Unlike most relatively simple comparison algorithms, Contrast can differentiate up to 63 different samples at once. Contrast handles differential analysis between experimental and control samples simply and flexibly.

20 Both the DTASelect and the Contrast programs can be licensed for use from The Scripps Research Institute (La Jolla, CA). For reference, see Tabb *et al.*, DTASelect and Contrast: Tools for Assembling and Comparing Protein Identifications from Shotgun Proteomics. *J. Proteome Res.* 1: 21-26, 2002 (incorporated herein by reference).

25 5. Exemplary Uses

To exclude contamination by somatically synthesized proteins, newly synthesized proteins from either synaptoneurosome or isolated dendrites of hippocampal cultures can be analyzed. Isolated dendrites are obtained from a special culture system using polycarbonate nets to separate dendrites from cell bodies.
30 Individual datasets of newly synthesized proteins can be obtained following different parameters of synaptic stimulation. Once compared, these datasets provide a detailed picture of the local molecular changes associated with different patterns of synaptic activity or the presence or absence of different neuromodulators. Furthermore, the approach allows the applicants to assess the general translational
35 capacity of dendrites by quantifying single proteins in comparison to their ¹⁵N-

5 substituted counterparts. Verification of the data can be examined at the mRNA level (in situ hybridization, RT-PCR) as well as at the protein level (immunocytochemistry and immunoprecipitation studies). The use of cell permeable, fluorescent alkyne-tags further expands this technology to *in vivo* real-time imaging of local protein synthesis in small dendritic segments by restricted
10 application of labeling reagents and neuromodulators. This helps to address the question of whether spatial specificity of locally synthesized proteins are locally synthesized proteins delivered only to the primary site of activation, or are they delivered to neighboring synapses as well.

Other possible applications, for instance, include proteomic profiling of
15 transgenic animals in comparison to their wild-type littermates or profiling of pathological states and changes during development of a tissue and even a whole organism. Knowing the proteomic differences of two cellular compartments, like somata and dendrites, or a brain region challenged either by a missing particular gene or a neuropathological state, helps to understand the molecular and subcellular
20 mechanisms of learning and memory formation, and of neurodegenerative diseases.

The invention circumvents the various problems associated with traditional approaches, by providing an approach to protein identification that unites techniques from organic chemistry, high throughput mass spectrometry and bioinformatics. For example, the procedure can be used to label newly synthesized dendritic proteins.

25 **Examples**

The following examples are for illustrative purpose only, and should in no way be construed to be limiting in any respect of the claimed invention.

Example 1 Culturing Neurons

30 Dissociated hippocampal neuron cultures were prepared from newborn rat pups (PO) as outlined in (Banker and Goslin, 1990), Neurons were plated at a density of 15,000 - 45,000 cells/cm² onto poly-D-lysine coated cell culture dishes or onto poly-D-lysine and growth factor reduced matrigel (BD Biosciences) coated polycarbonate nets with a pore size of 3 µm (Transwell, Corning) for the preparation

5 of isolated dendrites. The cultures were maintained and allowed to mature in growth medium (Neurobasal-A supplemented with B27 and GlutaMAX-1) for 14 to 21 days before use. The use of this growth medium suppressed glial proliferation, which was even more reduced by application of Ara-C (5 μ M final concentration). To supplement glial neurotrophic factors, 25% conditioned growth media from glial and
10 cortical cultures were added to the growth media. Isolated dendrites were obtained from Transwell cultures by removing the cell body layer with a sterile cell lifter after change to HBS. For immunolabeling after AHA-incorporation and cycloaddition with BiotinCyclooctyne, neurons are fixed at room temperature with 4% paraformaldehyde for 20 minutes. Fixed cultures are then treated sequentially with
15 PBS, blocking solution (0.2% Triton X-100, 2 mg/ml BSA, 5% Sucrose, 10% normal horse serum in PBS), primary Ab in blocking solution at 4°C overnight, PBS-Tx (0.2% Triton X-100 in PBS), Alexa488- or Alexa568-conjugated secondary Ab in blocking solution, PBS-Tx and PBS. Immunostained specimens are imaged in PBS with an Olympus confocal microscope or two-photon microscope through
20 either a 10 $\times$ or a 63 $\times$ oil-immersion lens. Alexa568 is excited at 543 nm and Alexa488 at 488 nm. Images are recorded through standard emission filters at contrast settings for which the crossover between the two channels is negligible.

25 **Example 2 [3+2] Cycloaddition Chemistry and Purification of Tagged Proteins**

In embodiments using [3+2] Cycloaddition, the following exemplary protocol may be used. Briefly, AHA, Biotin-PEO-Propargylamide and the triazole ligand were prepared as described previously [3, 46, 47]. The tandem featured alkyne tag was synthesized by GenScript Corporation. Biotin-Cyclooctyne is a
30 generous gift from Carolyn Bertozzi [45]. D10-L-leucine was purchased from Sigma. In all culture experiments, growth medium was removed from HEK293 cells, whole neuronal cultures or isolated dendrites were replaced with HEPES-buffered solution (HBS) [48] with 2.86 mM AHA, and 2.86 mM methionine for control experiments. After incubation at 37°C, 5% CO₂, cells were washed with
35 PBS-MC (1 mM MgCl₂, 0.1 mM CaCl₂ in PBS) to remove excess amounts of ABA and methionine. SNS and other biochemical fractions are incubated with 2.86 mM

5 AHA or 2.86 mM methionine under agitation at 37°C and washed in StimBuffer after incubation. Cells and fractions are lysed in 0.5% SDS, 1% Triton X-100 in PBS, with DNA destroyed by syringe and needle use. Lysates are diluted to 0.1% SDS, 0.2% Triton X-100 in PBS, complete protease inhibitor and treated for 10 min at 70°C. After addition of 200 μ M triazole ligand, 50 μ M alkyne tag and 75 μ g/ml
10 CuBr, the reaction is allowed to proceed for 16 hours at 4°C. After conclusion of the reaction, samples are dialyzed in 0.2% Triton X-100, 1 mM EDTA in PBS to remove excess reagents. Tagged proteins are purified using monomeric avidin columns followed by FLAG-M2-antibody immunoprecipitations. Eluted proteins are denatured in 8M Urea and proteolytically digested with endoproteinase Lys-C and
15 trypsin (both Roche) as described in [49].

Example 3 Improved Cu(I) Catalysis

An aliquot of cells expressing recombinant OmpC (1 mL) was centrifuged at 4°C and washed once in 1 mL of PBS (pH 7.4). The cells were centrifuged and
20 resuspended in 1 mL of PBS. Triazole ligand 5 was added to a final concentration of 200 μ M, and biotin-PEO-propargylamide 6 was added to a final concentration of 50 μ M. Addition of the active copper species was accomplished in two different ways. For in situ generation of Cu(I), 100 μ M CuSO₄ and 200 μ M of tris-(carboxyethyl)phosphine (TCEP) were added to the cells. Alternatively, the Cu(I)
25 ion was added directly to the cells in the form of an aqueous suspension of CuBr. Briefly, 10 μ L of a 10 mM suspension of CuBr (99.999% purity, Aldrich) was thoroughly agitated and added to the cells. As discussed in the Results section, the quality of the CuBr is critical for the success of the experiment. All labeling reactions were allowed to continue for 16 h at 4°C and were stopped by washing the
30 cells with PBS.

Example 4 Mass Spectrometric Analysis

Analysis of peptide mixtures by MudPIT was done essentially as described in Graumann *et al.* [49] using a LCQ-DecaXP electrospray ion trap mass
35 spectrometer (ThermoElectron). Raw mass spectra were filtered and analyzed with

5 2to3, SEQUEST, DTASelect and Contrast algorithms [33, 34] using standard and modification settings to find modified amino acid residues, like AHA in its possible tagged (+107 mass units) or failed-tagged (-5.1 mass units) version or D10-deuterated 1-leucine (+10 mass units).

The current state-of-the-art tandem mass spectrometry approaches, such as
10 MudPIT, couple in-line multidimensional chromatographic separations with continuous acquisition of data from a column effluent, allow collecting tens of thousands of spectra in a single experiment. Raw mass spectral data were filtered and analyzed using 2to3, SEQUEST, DTASelect and Contrast algorithms [33, 34], which validate the data and allow for efficient and comprehensive interpretation and
15 comparison with proteomic databases. In addition to the validation parameters set by the algorithms, the incorporation of stable isotope amino acids, for instance a deuterated variant of L-leucine, can serve as another inner validation for newly synthesized proteins [35]. Certain non-limiting exemplary data addressing the above issues are included in the following section.

20

Example 5 Pharmacology

In pharmacology experiments, the following drugs will be applied at the following concentrations: CNQX: 10 μ M (Sigma); APV: 50 μ M (Sigma); Anisomycin: 40 μ M (Sigma); Cycloheximide: 107 μ M (Sigma); DHPG: 50 μ M
25 (Tocris); NMDA: 50 μ M (Sigma); BDNF: 50 ng/ml (Promega); MG132: 50 μ M; glutamate/glycine: 100 μ M/10 μ M (Sigma), KC1: 90 mM (Sigma).

Example 6 Use of a Non-canonical Amino Acid AHA to Label and Identify Newly Synthesized Proteins

30 This example demonstrates the general suitability of the invention for mammalian cells, such as HEK293 cells and dissociated hippocampal neurons. To assess the purification of a target protein, cells were either transfected (HEK293 cells) or infected (dissociated hippocampal neurons) with a construct coding for a destabilized GFP protein. For neuronal infections, the destabilized and myristoylated
35 GFP reporter described in Aakalu *et al.* [4] is used. Meanwhile, HEK293 cells are

5 transfected with the pd2EGFP-N1 plasmid (Clontech) using PolyFect (Qiagen). The use of GFP enables monitoring protein levels in intact cells using fluorescence, as well as monitoring protein levels in cell lysates using Western blot analysis.

Figure 1 is a general overview for the suitability of AHA for use in the methods of the invention. Applicants first examined the specificity of AHA
10 incorporation, and its potential toxicity, using whole neuron morphological measurements as well as measurements of protein degradation. Incubation with methionine was used as a general control in all of the performed experiments. In order to visualize their gross morphological structure, neurons were infected with a destabilized form of the fluorescent protein EGFP, and incubated for 1.5 hours with
15 the same molar concentration of either AHA or methionine.

Dissociated hippocampal cultured neurons (12 DIV) were infected with a destabilized and myristoylated variant of GFP, whose mRNA is targeted into dendrites ([4]). Nine hours post infection, cells were incubated for 2 hrs with equimolar concentrations of AHA or methionine. Neurons expressing the GFP
20 reporter indicating no change in the gross morphology of AHA-incubated neurons compared to methionine controls. Thus AHA is not toxic to neurons, as evidenced by intact neuronal processes and continuous non-blebby expression of GFP throughout the dendritic arbor.

Applicants also confirmed the successful incorporation of AHA into proteins
25 in these cultures by subsequent biotinylation with the alkyne linker and Western blot analysis. Expressed GFP was selectively enriched by avidin chromatography from AHA-treated cells as compared to methionine-treated cells. A Western blot of avidin-purified samples was performed using either anti-biotin-antibody or an anti-GFP antibody. In both experiments, the eluate was enriched for the presence of
30 biotin-labeled proteins, indicating that the purification was successful.

To show the specificity of this technique for the detection of newly synthesized proteins, HEK293 cells were incubated for 2 hrs with AHA in the presence or absence of a protein synthesis inhibitor (Anisomycin or Cycloheximide). After incubation, cells were lysed and subjected to [3+2]cycloaddition with the
35 biotin-bearing alkyne-linker Biotin-PEO-Propargylamide [3], followed by Western blot analysis with an anti-Biotin-antibody. No signal was detected in either the

5 methionine or the protein synthesis inhibitor lanes, indicating the specificity of this technique as well as the membrane permeability of the reagent. The potential for increased protein degradation was measured by examining the ubiquitin signal strength on a Western blot from whole lysates of HEK293 cells, which were treated with AHA, methionine or incubation buffer for two hours. No increased
10 ubiquitination was observed in AHA treated cells compared to buffer or methionine controls, indicating that the modified amino acid does not cause severe protein misfolding.

Example 7 Purification of AHA-labeled (Biotinylated) Proteins

15 Following protein synthesis, cell lysates were subjected to a [3 + 2]-cycloaddition reaction with a biotin-bearing alkyne linker. To initially address our ability to purify AHA-labeled proteins, we have conducted avidin-affinity chromatography of biotinylated GFP, which was expressed in HEK293 cells or neuronal cultures.

20 In the same system, biotinylated GFP was detected in the eluate of a GFP-immunoprecipitation from AHA-treated cells followed by cycloaddition with the biotin-alkyne tag, but was not detected in methionine-treated control cells. Biotinylated GFP can be immunoprecipitated with a GFP-antibody from AHA-treated neuronal cultures or HEK293 cells, but not from methionine treated cultures.

25

Example 8 Mass Spectrometry

About 200 proteins were identified and validated in a tandem mass spectrometry analysis of single avidin-purified proteins from AHA-treated whole cell lysates of HEK293 cells. In this experiment, identifications for a protein were
30 considered as valid, if at least two fully tryptic peptides were assigned to a protein, with at least one peptide harboring a modification due to AHA or a deuterated leucine.

5 **Example 9 An Affinity Reagent with Two Affinity Groups**

To increase the specificity of the initial purification of tagged proteins by a two-step procedure. Applicants have developed an affinity reagent which allows for a two-step purification. One goal is to diminish contamination by proteins that adhere non-specifically to the matrices used for affinity-purification. In addition, a tag is introduced, which tag is suitable for detection by mass spectral analysis, and therefore can be used as an inner validation of the data.

The first alkyne tag used in the development of the technique, Biotin-PEO-Propargylamide, resulted in a gain of 578 mass units per reaction and was difficult to detect in the analysis because of its high molecular weight. To increase the purification of the modified proteins and facilitate detection by mass spectrometry, we have designed a new alkyne tag (Fig 2A) with the following properties: A biotin moiety (Biocytin, for avidin affinity purification) at the N-terminus followed by the FLAG-antibody epitope (DYKDDDDK, SEQ ID NO: 1), which is covalently linked to propargylglycine harboring the reactive alkyne group. A short spacer (GGA) was introduced between Biocytin and the FLAG epitope to ensure the steric accessibility of both biotin and FLAG during affinity purification. This tag can be also proteolytically cleaved by trypsin; following cleavage the resulting mass gain of tagged AHA is 107, which can be easily detected in the mass spectral analysis. Tandem purification is achieved by using first a monomeric avidin resin (Pierce) and then an anti-FLAG-M2 affinity gel (Sigma). In the first experiments, Applicants were able to demonstrate the successful use of this new tag. Fig 2B shows the subsequent purification of AHA-labeled proteins. Following tagging with the Biotin-FLAG-alkyne tag, AHA-labeled proteins from a HEK293 cell lysate were first subjected to an avidin column. After SDS elution from the matrix and removal of the majority of SDS, these proteins were further purified using an anti-FLAG-M2 affinity gel. Matrix-bound proteins were eluted by boiling in SDS sample buffer. As a control, a cell lysate from methionine-incubated HEK293 cells was treated in parallel. Note the absence of biotin signal in the methionine control. Both matrices allow for specific and mild elution of bound proteins by competition with biotin or FLAG-peptide, respectively, which will increase the specificity of the procedure even more. An establishment of these mild elution conditions is currently underway.

5 To assess the efficiency of the tandem purification procedure further, Applicants examine the composition of a tandem purified mixture of two proteins, GST and DHFR. One of them is expressed in *E. Coli* in the presence of AHA in minimal medium and the other is expressed in rich medium containing methionine. After [3+2]-cycloaddition with the tandem featured tag, varying ratios of the two
10 proteins are purified together and analyzed by mass spectrometry for contamination by the non-tagged protein. Additionally, different amounts of a single, tagged protein (*e.g.*, GST) are added to a non-tagged cell lysate. Tandem purification and mass spectrometry are performed to establish how much of a tagged protein is needed for successful identification. Finally, the levels of AHA labeling is varied
15 and determined in order to optimize and characterize the method.

 Further internal validation of the mass spectral data is achieved by using 10-fold deuterated L-leucine (D10-L-leucine) together with AHA during protein synthesis [35]. Peptides derived from proteins labeled with AHA + D10-L-leucine are expected to show an increase of 10 mass units per incorporated D10-L-leucine,
20 and as such can be distinguished from peptides derived from preexisting proteins.

References

1. Zhang *et al.*, Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes. *Curr Opin Chem Biol* **8(1)**: 66-75, 2004.
- 25 2. Kiick *et al.*, Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc Natl Acad Sci U S A* **99(1)**: 19-24, 2002.
3. Link and Tirrell, Cell Surface Labeling of *Escherichia coli* via Copper(I)-Catalyzed [3+1] Cycloaddition. *Journal of the American Chemical Society*
30 **125(37)**: 11164-11165, 2003.
4. Aakalu *et al.*, Dynamic visualization of local protein synthesis in hippocampal neurons. *Neuron* **30(2)**: 489-502, 2001.
5. Huber *et al.*, Role for rapid dendritic protein synthesis in hippocampal mGluR-dependent long-term depression. *Science* **288(5469)**: 1254-7, 2000.

- 5 6. Kang and Schuman, A requirement for local protein synthesis in
neurotrophin-induced hippocampal synaptic plasticity. *Science* **273(5280)**:
1402-6, 1996.
7. Kacharina *et al.*, Stimulation of glutamate receptor protein synthesis and
membrane insertion within isolated neuronal dendrites. *Proc Natl Acad Sci U*
10 *S A* **97(21)**: 11545-50, 2000.
8. Ouyang *et al.*, Visualization of the distribution of autophosphorylated
calcium / calmodulin-dependent protein kinase II after tetanic stimulation in
the CAI area of the hippocampus. *J Neurosci* **17(14)**: 541-627, 1997.
9. Ouyang *et al.*, Tetanic stimulation leads to increased accumulation of Ca^{2+} /
15 calmodulin-dependent protein kinase II via dendritic protein synthesis in
hippocampal neurons. *J Neurosci*. **19(18)**: 7823-33, 1999.
10. Steward and Levy, Preferential localization of polyribosomes under the base
of dendritic spines in granule cells of the dentate gyrus. *J Neurosci* **2(3)**: 284-
91, 1982.
- 20 11. Steward and Fass, Polyribosomes associated with dendritic spines in the
denervated dentate gyrus: evidence for local regulation of protein synthesis
during reinnervation. *Prog Brain Res* **58**: 131-6, 1983.
12. Steward and Falk, Protein-synthetic machinery at postsynaptic sites during
synaptogenesis: a quantitative study of the association between
25 polyribosomes and developing synapses. *J Neurosci* **6(2)**: 412-23, 1986.
13. Nguyen *et al.*, Requirement of a critical period of transcription for induction
of a late phase of LTP. *Science* **265(5175)**: 1104-7, 1994.
14. Martin *et al.*, Synapse-specific, long-term facilitation of aplysia sensory to
motor synapses: a function for local protein synthesis in memory storage.
30 *Cell* **91(7)**: 927-38, 1997.
15. Kang and Schuman, A requirement for local protein synthesis in
neurotrophin-induced hippocampal synaptic plasticity. *Science* **273(5280)**:
1402-6, 1996.

- 5 16. Casadio *et al.*, A transient, neuron-wide form of CREB-mediated long-term facilitation can be stabilized at specific synapses by local protein synthesis. *Cell* **99(2)**: 221-37, 1999.
17. Bito *et al.*, CREB phosphorylation and dephosphorylation: a Ca^{2+} - and stimulus duration-dependent switch for hippocampal gene expression. *Cell*
10 **87(7)**: 1203-14, 1996.
18. Steward and Schuman, Protein synthesis at synaptic sites on dendrites. *Annu Rev Neurosci* **24**: 299-325, 2001.
19. Steward and Schuman, Compartmentalized synthesis and degradation of proteins in neurons. *Neuron* **40(2)**: 347-59, 2003.
- 15 20. Wu *et al.*, CPEB-mediated cytoplasmic polyadenylation and the regulation of experience-dependent translation of alpha-CaMKII mRNA at synapses. *Neuron* **21(5)**: 1129-39, 1998.
21. Weiler *et al.*, Synapse-activated protein synthesis as a possible mechanism of plastic neural change. *Prog Brain Res* 100: 189-94, 1994.
- 20 22. Steward, mRNA localization in neurons: a multipurpose mechanism? *Neuron* **18(1)**: 9-12, 1997.
23. Torre and Steward, Demonstration of local protein synthesis within dendrites using a new cell culture system that permits the isolation of living axons and dendrites from their cell bodies. *J Neurosci* **12(3)**: 762-72, 1992.
- 25 24. Glanzer and Eberwine, Mechanisms of translational control in dendrites. *Neurobiol Aging* **24(8)**: 1105-11, 2003.
25. Crino and Eberwine, Molecular characterization of the dendritic growth cone: regulated mRNA transport and local protein synthesis. *Neuron* **17(6)**: 1173-87, 1996.
- 30 26. Crino *et al.*, Presence and phosphorylation of transcription factors in developing dendrites. *Proc Natl Acad Sci U S A* **95(5)**: 2313-8, 1998.
27. Job and Eberwine, Identification of sites for exponential translation in living dendrites. *Proc Natl Acad Sci U S A* 98(23): 13037-42, 2001.
28. Ju *et al.*, Activity-dependent regulation of dendritic synthesis and trafficking
35 of AMPA receptors. *Nat Neurosci* **7(3)**: 244-53, 2004.

- 5 29. Brendel *et al.*, Characterization of Staufen 1 ribonucleoprotein complexes. *Biochem J* **384**(Pt 2): 239-46, 2004.
30. Eberwine *et al.*, Local translation of classes of mRNAs that are targeted to neuronal dendrites. *Proc Natl Acad Sci U S A* **98**(13): 7080-5, 2001.
31. Gygi *et al.*, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**(10): 994-9, 1999.
- 10 32. Washburn *et al.*, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**(3): 242-7, 2001.
33. Tabb *et al.*, DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* **1**(1): p. 21-6, 2002.
- 15 34. Ducret *et al.*, High throughput protein characterization by automated reverse-phase chromatography / electrospray tandem mass spectrometry. *Protein Sci* **7**(3): 706-19, 1998.
- 20 35. Ong *et al.*, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**(5): 376-86, 2002.
36. Bassell *et al.*, Sorting of beta-actin mRNA and protein to neurites and growth cones in culture. *J Neurosci* **18**(1): 251-65, 1998.
- 25 37. Kiebler *et al.*, Purification and characterization of rat hippocampal CA3-dendritic spines associated with mossy fiber terminals. *FEBS Lett* **445**(1): 80-6, 1999.
38. Meffert *et al.*, Nitric oxide stimulates Ca^{2+} -independent synaptic vesicle release. *Neuron* **12**(6): 1235-44, 1994.
- 30 39. Torre and Steward, Protein synthesis within dendrites: glycosylation of newly synthesized proteins in dendrites of hippocampal neurons in culture. *J Neurosci* **16**(19): 5967-78, 1996.
40. Gerber *et al.*, Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* **100**(12): 6940-5, 2003.

- 5 41. Peng *et al.*, Semi-quantitative proteomic analysis of rat forebrain postsynaptic density fractions by mass spectrometry. *J Biol Chem* **279(20)**: 21003-11, 2004.
42. Yin *et al.*, The brain-derived neurotrophic factor enhances synthesis of Arc in synaptoneurosome. *Proc Natl Acad Sci U S A* **99(4)**: 2368-73, 2002.
- 10 43. Scbratt *et al.*, BDNF regulates the translation of a select group of mRNAs by a mammalian target of rapamycin-phosphatidylinositol 3-kinase-dependent pathway during neuronal development. *J Neurosci* **24(33)**: 7366-77, 2004.
44. Bagni *et al.*, Chemical stimulation of synaptosomes modulates alpha - Ca²⁺/calmodulin-dependent protein kinase II mRNA association to polysomes. *J Neurosci* **20(10)**: RC76, 2000.
- 15 45. Agard *et al.*, A strain promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J Am Chem Soc* **126(46)**: 15046-7, 2004.
46. Mangold *et al.*, Azidoalanine mutagenicity in Salmonella: effect of homologation and alpha-methyl substitution. *Mutat Res* **216(1)**: 27-33, 1989.
- 20 47. Wang *et al.*, Bioconjugation by copper(I) -catalyzed azide-alkyne [3 + 2] cycloaddition. *J Am Chem Soc* **125(11)**: 3192-3, 2003.
48. Malgaroli and Tsien, Glutamate-induced long-term potentiation of the frequency of miniature synaptic currents in cultured hippocampal neurons. *Nature* **357(6374)**: 134-9, 1992.
- 25 49. Graumann *et al.*, Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. *Mol Cell Proteomics* **3(3)**: 226-37, 2004.

The contents of all cited references (including literature references, issued
30 patents, published patent applications as cited throughout this application) are
hereby expressly incorporated by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific method and

- 5 reagents described herein, including alternatives, variants, additions, deletions, modifications and substitutions. Such equivalents are considered to be within the scope of this invention and are covered by the following claims.

5 **CLAIMS:**

1. A method for determining protein expression pattern in a cell or a tissue, comprising:
 - (1) contacting the cell or the tissue with a non-natural amino acid comprising a first reactive group, under a condition where the non-natural amino acid is incorporated into the newly synthesized proteins of the cell or the tissue;
 - (2) contacting proteins or fragments thereof from the cell or the tissue with an affinity reagent comprising a second reactive group and an affinity moiety, wherein the first and second reactive groups react to label the non-natural amino acid with the affinity reagent;
 - (3) isolating the newly synthesized proteins or fragments thereof comprising the non-natural amino acid labeled with the affinity reagent via the affinity moiety; and,
 - (4) determining protein expression pattern by identifying the isolated proteins or fragments thereof.
2. The method of claim 1, wherein step (4) further comprises quantitating each identified / isolated protein or fragments thereof.
3. The method of claim 1, wherein the first reactive group is an azido group.
4. The method of claim 3, wherein the non-natural amino acid is azidoalanine, azidohomoalanine (AHA), azidonorvaline, or azidonorleucine.
5. The method of claim 1, wherein the first reactive group is a ketone or aldehyde moiety.
6. The method of claim 1, wherein the first reactive group is a diboronic acid moiety.
7. The method of claim 1, wherein the first reactive group is a terminal alkyne moiety.

- 5 8. The method of claim 1, wherein in step (1), the non-natural amino acid is incorporated *in vivo* by endogenous protein synthesis machinery of the cell or the tissue.
9. The method of claim 1, wherein in step (1), the non-natural amino acid is site-specifically incorporated in place of a natural amino acid selected from methionine or phenylalanine.
- 10 10. The method of claim 1, wherein the affinity reagent further comprising an antigenic moiety that can be recognized by an antibody.
11. The method of claim 10, wherein the affinity moiety is biotin, and the antigenic moiety is an epitope tag.
- 15 12. The method of claim 11, wherein the epitope tag is FLAG tag.
13. The method of claim 12, wherein the FLAG tag comprises one or more cleavage sites for a sequence-specific protease.
14. The method of claim 13, wherein the sequence-specific protease is trypsin.
15. The method of claim 12, wherein the FLAG tag is situated between the affinity moiety and the second reactive group.
- 20 16. The method of claim 1, wherein the second reactive group and the affinity moiety are linked by one or more cleavage functional groups.
17. The method of claim 16, wherein the cleavable functional groups are photo-cleavable groups, chemically cleavable groups, or enzymatically cleavable groups.
- 25 18. The method of claim 1, wherein the second reactive group and the affinity moiety are linked by a photo-cleavable linker.
19. The method of claim 1, wherein the fragments of the proteins are generated by protease digestion or chemical fragmentation.
- 30 20. The method of claim 19, wherein the protease digestion is performed using a protease that generates a C-terminal Arginine or Lysine.
21. The method of claim 20, wherein the protease is trypsin or endoproteinase Lys C.

- 5 22. The method of claim 19, wherein the chemical fragmentation is performed by CNBr.
23. The method of claim 19, wherein the fragments of the proteins are generated either before or after the isolation step (2).
24. The method of claim 1, wherein identification of the target protein or
10 fragments thereof is performed by mass spectrometry.
25. The method of claim 1, wherein the fragments are generated either before step (2) or after step (3).
26. The method of claim 1, wherein in step (1), the condition is such that the
15 concentration of the non-natural amino acid is optimized to effect a pre-selected amount or percentage of incorporation.
27. The method of claim 1, wherein step (1) is carried out by contacting the cell cultured *in vitro*.
28. The method of claim 1, wherein step (1) is carried out by administering the non-natural amino acid to an animal.
- 20 29. The method of claim 1, wherein in step (1), the cell or the tissue is further contacted with a second non-natural amino acid.
30. The method of claim 29, wherein the second non-natural amino acid contains an isotope tag.
31. The method of claim 30, wherein the second non-natural amino acid is a
25 deuterated natural amino acid.
32. The method of claim 30, wherein the isotope tag is one or more of ^2H , ^3H , ^{13}C , ^{14}C , ^{15}N , ^{18}O , or ^{35}S .
33. A method of comparing protein expression profile of a first sample and a second sample, comprising:
- 30 (1) using the method of claim 1, determining the protein expression profiles of the first and the second samples,
- (2) comparing the protein expression profile of the first sample with that of the second sample.

- 5 34. The method of claim 33, wherein the first sample is a control sample, and the second sample is the control sample treated by an agent.
35. The method of claim 34, wherein the agent is a pharmaceutical agent.
36. A method for monitoring protein expression in a cell or a tissue, comprising:
- 10 (1) contacting the cell or the tissue with a non-natural amino acid comprising a first reactive group, under a condition where the non-natural amino acid is incorporated into the newly synthesized proteins of the cell or the tissue;
- 15 (2) contacting the cell or the tissue with a fluorescent reagent comprising a second reactive group and a fluorescent moiety or a reagent that can be subsequently coupled to a fluorescent reagent, wherein the first and second reactive groups react to label the non-natural amino acid with the fluorescent reagent, said fluorescent reagent is substantially more fluorescent after the first and second reactive groups react;
- 20 (3) monitoring the intensity of fluorescence in the cell or the tissue, thereby monitoring protein expression in said cell or said tissue.
37. The method of claim 36, wherein the cell or the tissue is in a live animal.
38. A functional reagent comprising:
- 25 (1) a reactive group for reacting with a non-natural amino acid incorporated into a protein;
- (2) an affinity moiety or a fluorescent moiety; and,
- (3) one or more cleavable groups situated between the reactive group and the affinity moiety or the fluorescent moiety.
39. The functional reagent of claim 38, wherein the reactive group is an azido group.
- 30 40. The functional reagent of claim 38, wherein the reactive group is a ketone or an aldehyde moiety.
41. The functional reagent of claim 38, wherein the reactive group is a terminal alkyne moiety.

- 5 42. The functional reagent of claim 38, which is sufficiently membrane permeable to diffuse into a live cell.
43. The functional reagent of claim 38, which is substantially non-toxic.
44. The functional reagent of claim 38, wherein the reactive group reacts with a non-natural amino acid incorporated in a protein.
- 10 45. The functional reagent of claim 38, further comprising an antigenic moiety that can be recognized by an antibody.
46. The functional reagent of claim 45, wherein the antigenic moiety is an epitope tag.
47. The functional reagent of claim 46, wherein the epitope tag is FLAG tag.
- 15 48. The functional reagent of claim 47, wherein the FLAG tag comprises one or more cleavage sites for a sequence-specific protease.
49. The functional reagent of claim 48, wherein the sequence-specific protease is trypsin.
50. The functional reagent of claim 47, wherein the FLAG tag is situated
- 20 between the affinity moiety and the second reactive group.
51. The functional reagent of claim 38, wherein the second reactive group and the affinity moiety are linked by one or more cleavage functional groups.
52. The functional reagent of claim 51, wherein the cleavable functional groups are photo-cleavable groups, chemically cleavable groups, or enzymatically
- 25 cleavable groups.
53. The functional reagent of claim 38, wherein the affinity moiety is biotin.
54. The functional reagent of claim 48, wherein the affinity moiety is biotin, wherein the FLAG tag comprises two cleavage sites for a sequence-specific protease, and wherein the FLAG tag is situated between the affinity moiety
- 30 and the second reactive group.
55. The functional reagent of claim 48, wherein the linker is a photo-cleavable linker.

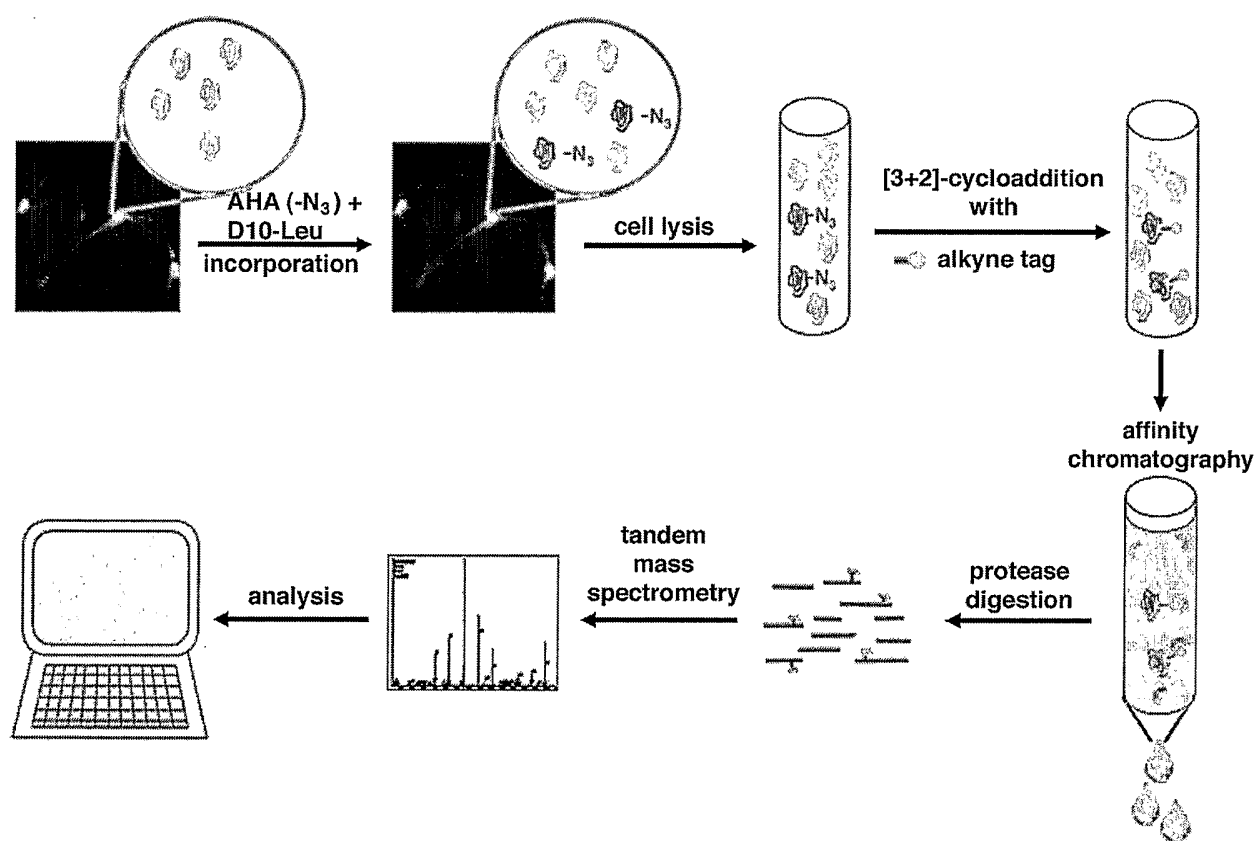


Figure 1

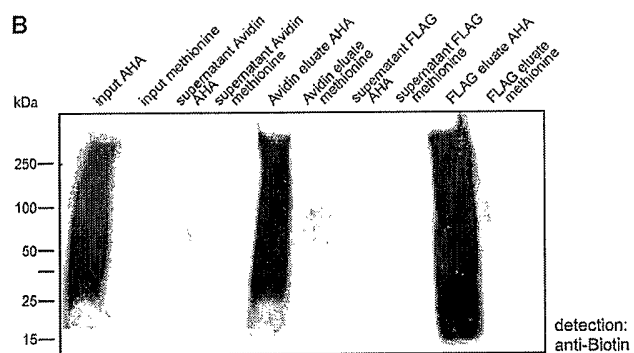
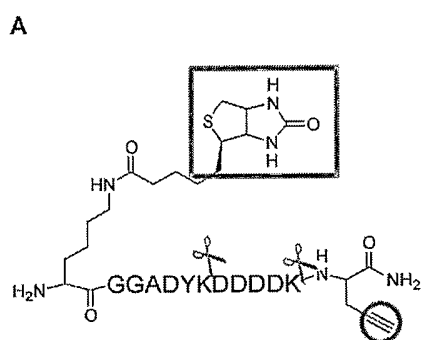


Figure 2