(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
08 May 2025 (08.05.2025)

WIPO | PCT

(10) International Publication Number
WO 2025/096924 A2

(51) International Patent Classification:
G01N 33/58 (2006.01) *G01N 33/68* (2006.01)

(21) International Application Number:
PCT/US2024/054095

(22) International Filing Date:
01 November 2024 (01.11.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/596,074 03 November 2023 (03.11.2023) US

(71) Applicant: WISCONSIN ALUMNI RESEARCH FOUNDATION [US/US]; 614 Walnut Street, 13th Floor, Madison, Wisconsin 53726 (US).

(72) Inventors: SUSSMAN, Michael; 8761 Hidden Valley Road, Cross Plains, Wisconsin 53528 (US). **WOLFER, Jamison**; 1710 Ludden Drive, Cross Plains, Wisconsin 53528 (US). **MINKOFF, Benjamin**; 4821 Mineral Point Road, Madison, Wisconsin 53705 (US).

(74) Agent: BLASIOLE, Daniel et al.; 25 W. Main Street, Suite 800, Madison, Wisconsin 53726 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: METHODS OF LABELING, PROCESSING, AND ANALYZING PROTEIN SAMPLES

(57) Abstract: Methods of labeling, processing and analyzing protein samples. Surface-accessible residues in a protein sample (such as a proteome) can be labeled using a carbodiimide coupling agent an amine comprising a click chemistry handle, such as glycine propargyl amide. The click chemistry handle can then be derivatized using high-yield "click" chemistry to form a stable linkage with an enrichable label. The proteome can then be digested, peptides that have been modified can be enriched, and the enriched peptides can be analyzed, for example, by mass spectrometry. The analyzed peptides can constitute a highly enriched subset of the original peptides while still representing many sites of modification on many proteins (~95%) originally present in the proteome. The methods can be used to identify solvent-accessible regions of proteins, conformational changes, and drug targets across a whole proteome.

WO 2025/096924 A2

METHODS OF LABELING, PROCESSING, AND ANALYZING PROTEIN SAMPLES

CROSS-REFERENCE TO RELATED APPLICATIONS

Priority is hereby claimed to US Provisional Application 63/596,074, filed November 3, 2023, which is incorporated herein by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

This invention was made with government support under 2010789 and 1943816 awarded by the National Science Foundation and under DE-FG02-88ER13938 awarded by the US Department of Energy. The government has certain rights in the invention.

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted in XML format and is hereby incorporated by reference in its entirety. The XML copy, created on October 25, 2024, is named PCT--09824578-P230380WO01--APP--SEQ_LIST.xml and is 8,522 bytes in size.

FIELD OF THE INVENTION

The invention is directed to methods of labeling, processing and analyzing protein samples for, for example, detecting proteome-wide conformational changes and drug targets in crude cell lysates.

BACKGROUND

Protein footprinting technologies have conventionally been used to map solvent accessibility and infer protein structure. Such technologies are currently experiencing a renaissance, as a growing body of reagents are being created to label different subsets of amino acids and mass spectrometers have become increasingly sensitive. Protein footprinting methods have wide utility, from basic academic research to early- and mid-stage drug discovery, due to the now routine use of protein footprinting as a means of mapping protein:drug interactions, the growing array of protein therapeutics themselves on the market, and the difficulty of analyzing a large number of protein interaction systems using classic techniques such as X-ray crystallography.

Most of the protein footprinting methods being developed are aimed at mapping amino acid side chains that mediate interactions between a single protein and its interacting partner at the protein's surface. A few have been developed with the goal of determining proteome-wide conformational analyses. These methods include chemical crosslinking, thermal proteome profiling, methionine oxidation, and limited proteolysis These proteome-wide conformational analyses, however, are incapable of dealing with the enormous complexity inherent in identifying modified peptides from an unfractionated proteomic digest of a whole cell.

Among these proteomic methods, for example, only chemical crosslinking has the ability to enrich for modified peptides, but identifying two distinct peptides chemically connected to each other after enrichment remains challenging, both from a technical and a bioinformatic standpoint. Thermal proteome profiling exploits shifts in protein melting temperatures upon binding a ligand to identify potential binding partners from a whole proteome, but the information this technique yields is limited to the protein level (e.g., protein X or Y is identified, but the location on protein X or Y is not identified). There is also no way to enrich for peptides of interest. Methionine oxidation, unsurprisingly, analyzes oxidation on methionine as a proxy for solvent accessibility, but there is no way to enrich for peptides of interest. The most promising technique to emerge, limited proteolysis, analyses areas of protease accessibility as a proxy for solvent accessibility, but this method still has significant limitations.

Methods of detecting proteome-wide conformational changes in complex protein samples, such as crude cell lysates, are needed.

SUMMARY OF THE INVENTION

One aspect of the invention is directed to covalent labeling systems for use in protein footprinting (see, *e.g.*, FIGS. 1 and 2). In at least some of the methods provided herein, surface-accessible aspartate and glutamate residues in a proteome can be labeled using a reagent combination of a carbodiimide coupling agent such as 1-ethyl-3-(3'-dimethylaminopropyl)carbodiimide (EDC) and dicyclohexylcarbodiimide (DCC) and an amine comprising a click chemistry handle, such as glycine propargyl amide (GPA). The click chemistry handle can then be derivatized using high-yield "click" chemistry to form a stable linkage with a label such as biotin. The proteome can then be digested, peptides that have been modified can be enriched using streptavidin:biotin interaction methods, and the enriched peptides can be analyzed, for example, by mass spectrometry. The analyzed peptides can thus constitute a highly enriched subset of the original peptides while still representing many sites of modification on many proteins (~95%) originally present in the proteome. In this way, one can identify solvent-accessible regions of proteins across a whole proteome.

The methods provided herein offer simple instrumental and bioinformatic analysis with the added benefit of enriching peptides of interest, to ultimately get more information about the conformational proteome (also known as the 3D proteome or the conformatiome) than before. The simplicity of the labeling, enrichment, and analyses make this a desirable technique for pharma, especially given the advantages over the current methods. The methods provided herein have utility not only for academia but also for biotech and pharma companies, specifically for drug discovery and also for critical ongoing efforts such as identifying the "off-target" proteins that degrade the utility of any new drug. As just one example, one could add a drug of interest to a mammalian cell lysate, label the proteome pre- and post-treatment, and ask which of the thousands of proteins have new regions exposed or blocked and are thus likely interesting with the drug of interest or have had their conformation changed.

In some embodiments, the invention is directed to methods of labeling and, optionally, processing and analyzing protein samples. In some versions, the methods comprise reacting a protein sample in a first labeling reaction with a carboxylic acid activating agent and a first labeling agent to label a first set of protein residues in the protein sample with the first labeling agent. In some versions, the first labeling agent has the structure of $Nu\text{-}L^1\text{-}Y^1$, wherein Nu is a nucleophilic group, L^1 is a first linker, and Y^1 is a first click chemistry handle.

In some versions, the protein sample during the first labeling reaction comprises a protein mixture comprising at least 100 different proteins. In some versions, the protein sample during the first labeling reaction comprises a full cellular proteome.

In some versions, the nucleophilic group is selected from the group consisting of substituted or unsubstituted amino, hydroxyl, and thiol. In some versions, the nucleophilic group is unsubstituted amino.

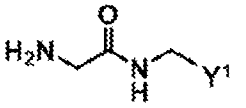
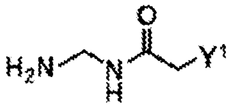
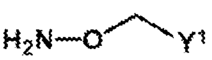
In some versions, the carboxylic acid activating agent is selected from the group consisting of a carbodiimide coupling agent and 2-ethyl-5-phenylisoxazolium-3'-sulfonate. In some versions, the carbodiimide coupling agent is selected from the group consisting of 1-ethyl-3-(3'-dimethylaminopropyl)carbodiimide (EDC), 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide methiodide (EDC), dicyclohexylcarbodiimide (DCC), 1-ethyl-3-(3'-dimethylaminopropyl)carbodiimide (DIC), and N-cyclohexyl-N'-(2-morpholinylethyl)carbodiimide methyl-p-toluenesulfonate (CMC).

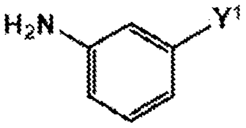
In some versions, the first linker comprises one or more instances of one or more first linking moieties selected from the group consisting of optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroaralkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted arylene, optionally substituted heteroarylene, oxygen, sulfur, phosphorus, silicon, secondary amine, tertiary amine, carbonyl, and sulfonyl. In some versions, the first linker comprises one or more instances of one or more first linking moieties selected from the group consisting of: $-(CH_2)_x-$; $-(CH_2)_x-$, $-(CH_2CH_2O)_x-$, $-(OCH_2CH_2)_x-$, $-(CH_2O)_x-$, and $-(OCH_2)_x-$; $-(CH_2)_x-$, $-C(O)NH-$, and $-NHC(O)-$; or *para*-phenylene, *meta*-phenylene, and *ortho*-phenylene,

wherein *x* in each instance is independently an integer from 1 to 6. In some versions, the first linker comprises a backbone chain that is devoid of an ester group.

In some versions, the first click chemistry handle is selected from the group consisting of a terminal alkyne, an azide, a strained alkyne, a diene, a dieneophile, an alkoxyamine, a phosphine, a hydrazide, a thiol, a tetrazine, and an alkene. In some versions, the first click chemistry handle is selected from the group consisting of a terminal alkyne and an azide.

In some versions, the first labeling agent is selected from the group consisting of:


;

;

;

; and

,

wherein Y^1 is a terminal alkyne.

Some versions further comprise reacting the protein sample in a second labeling reaction with a second labeling agent to label a second set of protein residues in the protein sample with the second labeling agent. In some versions, the second labeling agent has the structure of Y^2-L^2-Z, wherein Y^2 is a second click chemistry handle, L^2 is a second linker, Z is a label, and the second click chemistry handle is a partner click chemistry handle of the first click chemistry handle.

In some versions, the second click chemistry handle is selected from the group consisting of a terminal alkyne, an azide, a strained alkyne, a diene, a dieneophile, an alkoxyamine, a phosphine, a hydrazide, a thiol, a tetrazine, and an alkene.

In some versions, the second linker is a cleavable linker.

In some versions, the label comprises a ligand.

In some versions, the second labeling agent comprises PC biotin azide (1-[4-[[22-[(3aR,4R,6aS)-Hexahydro-2-oxo-1H-thieno[3,4-d]imidazol-4-yl]-4,18-dioxo-8,11,14-trioxa-5,17-diazadocos-1-yl]oxy]-5-methoxy-2-nitrophenyl]ethyl N-(3-azidopropyl)carbamate).

Some versions further comprise fragmenting the protein sample. In some versions, the fragmenting is performed after the first labeling reaction and prior to the second labeling reaction. In some versions, the fragmenting is performed after the first labeling reaction and after the second labeling reaction.

Some versions further comprise separating labeled polypeptides from unlabeled polypeptides, wherein the labeled polypeptides are polypeptides comprising at least one protein residue from the second set of protein residues, and wherein the unlabeled polypeptides are polypeptides lacking any of the second set of protein residues.

Some versions further comprise, after the separating, cleaving the second linkers to separate the labeled polypeptides from the label.

Some versions further comprise determining amino acid sequences of the labeled polypeptides.

In some versions, the protein sample comprises multiple protein samples and the method further comprises sample-labeling each of the multiple protein samples with detectably different sample labels and subsequently combining the multiple protein samples. In some versions, the combining the multiple protein samples is performed prior to determining the amino acid sequences of the labeled polypeptides. In some versions, the reacting the protein sample in the first labeling reaction comprises separately reacting each of the multiple protein samples in separate first labeling reactions and subsequently sample-labeling the multiple protein samples. In some versions, the reacting the protein sample in the second labeling reaction comprises separately reacting each of the multiple protein samples in separate second labeling reactions and subsequently sample-labeling the multiple protein samples. In some versions, the fragmenting the protein sample comprises separately fragmenting each of the multiple protein samples and subsequently sample-labeling the multiple fragmented protein samples. In some versions, the sample labeling and the combining occur prior to the reacting the protein sample in the second labeling reaction. In some versions, the sample labeling and the combining occur prior to the fragmenting the protein sample. In some versions, the detectably different sample labels are mass labels, such as tandem mass tag labels.

In some versions, the protein sample comprises a first protein sample and a second protein sample. In some versions, during the first labeling reaction, a proteome of the first protein sample is identical to a proteome of the second protein sample, and wherein, during the first labeling reaction, the first protein sample comprises at least one molecule not present in the second protein sample. In some versions, during the first labeling reaction, the first protein sample comprises a first sequence variant of a protein and the second sample comprises a second sequence variant of a protein. In some versions, the first sequence variant is a native form of the protein and the second sequence variant is a mutated form of the native form of the protein. In some versions, the method further comprises comparing the sequences of the labeled polypeptides from the first protein sample with the sequences of the labeled polypeptides from the second protein sample.

The objects and advantages of the invention will appear more fully from the following detailed description of the preferred embodiment of the invention made in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schema of exemplary aspects of the invention, including exemplary labeling steps.

FIG. 2 is a schema of exemplary aspects of the invention, including exemplary labeling steps, protein fragmentation (*e.g.*, digestion), peptide enrichment, linker cleavage, and sequence analysis (*e.g.*, via MS).

FIGS. 3A and B. Exemplary click chemistry handle pairs, reactions, and reaction products. In FIG. 3A, A and B represent exemplary click chemistry handle pairs. In FIG. 3B, R^1 and R^2 represent moieties to which the click chemistry handles are attached.

FIG. 4. Labeling of aspartic acid and glutamic acid side chains with glycine ethyl ester (GEE) or glycine propargyl amide (GPA) and the subsequent hydrolysis product of GEE. EDC: 1-ethyl-3-[3-(dimethylamino)propyl]carbodiimide.

FIG. 5. Clickable amines screened in the EDC-mediated coupling reaction with BSA.

FIG. 6. Comparison of the number of labeled BSA Asp and Glu sites using GEE and GPA.

FIG. 7. Coupling reagents screened in the footprinting of BSA with GPA.

FIG. 8. Differential modification observed on labeling sites between pre-digested BSA and intact BSA.

FIG. 9. Structures of the enrichment scaffolds tested with GPA-labeled *Arabidopsis* cytosol. The cleavage site is indicated.

FIG. 10. Streptavidin blotting of proteins from *Arabidopsis* cytosol. GPA-labeled proteins were clicked to a biotin-azide and then blotted with streptavidin.

FIG. 11. Degree of MS-measured labeling to a normalized SASA value for footprinted sites on ALFP2. Shown is a plot of 0-1 normalized MS abundance values against SASA values.

DETAILED DESCRIPTION OF THE INVENTION

One aspect of the invention is directed to methods of labeling, processing, and/or analyzing protein samples.

“Protein sample” as used herein refers to any sample comprising one or more proteins. In addition to the protein(s), the protein sample may comprise any of a number of other components, such as a liquid medium, salts, nucleic acids, lipids, biochemical reagents, etc.

The terms “protein,” “peptide” and “polypeptide” are used interchangeably herein, and refer to a polymer of amino acid residues linked together by peptide (amide) bonds. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide may refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or polypeptide may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, etc. A protein, peptide, or polypeptide may also be a single molecule or may be a multi-molecular complex. A protein, peptide, or polypeptide may be just a fragment of a naturally occurring protein or peptide. A protein, peptide, or polypeptide may be naturally occurring, recombinant, or synthetic, or any combination thereof.

The methods of the invention can comprise a first labeling reaction. The first labeling reaction can comprise reacting the protein sample in a first labeling reaction with a carboxylic activating agent and a first labeling agent to label a first set of protein residues in the protein sample with the first labeling agent. “First set of protein residues” in this context refers to the set of individual amino acid residues in the proteins in the protein sample that are labeled (chemically reacted) with the first labeling agent.

Carboxylic activating agents are compounds that activate carboxylic acid (carboxylate) groups to render them susceptible to nucleophilic attack. A number of carboxylic activating are known in the art. Exemplary carboxylic activating agents include carbodiimide coupling agents and Woodward’s Reagent K (2-ethyl-5-phenylisoxazolium-3'-sulfonate), among others.

Carbodiimide coupling agents are compounds containing carbodiimides that serve as dehydration agents in the conversion of carboxylic acids to amides or esters. A number carbodiimide coupling agents are known in the art. Examples include 1-ethyl-3-(3'-dimethylaminopropyl)carbodiimide (EDC), 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide methiodide (EDC), dicyclohexylcarbodiimide (DCC), 2-ethyl-5-phenylisoxazolium-3'-sulfonate (Woodward’s reagent K, WRK), 1-ethyl-3-(3'-

dimethylaminopropyl)carbodiimide (DIC), and N-cyclohexyl-N'-(2-morpholinylethyl)carbodiimide methyl-p-toluenesulfonate (CMC).

The first labeling agent can comprise at least a nucleophilic group and a click chemistry handle, the latter of which is referred to as a “first click chemistry handle” in the first labeling agent. The nucleophilic group can be connected to the first click chemistry handle via a linker, the latter referred to as a “first linker” in the first labeling agent. The first labeling agent can accordingly have the structure of Nu-L^1-Y^1, wherein Nu is a nucleophilic group, L^1 is a first linker and Y^1 is a first click chemistry handle.

Exemplary nucleophilic groups include substituted or unsubstituted amino, hydroxyl, and thiol, among others.

The term “linker” as used herein refers to one or more linking moieties covalently linking two or more linked moieties. The linked moieties can be any moieties of interest. Exemplary linked moieties include, for example, an amino group and a click chemistry handle or a click chemistry handle and a label (as discussed in further detail below). The linking moieties can comprise one or chemical moieties suitable for covalently linking the linked moieties. Exemplary linking moieties include optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroaralkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted arylene, optionally substituted heteroarylene, oxygen, sulfur, phosphorus, silicon, optionally substituted amine, carbonyl, sulfonyl, and cleavable linking moieties. The linkers of the invention can generally comprise one (or more, if branched) backbone chains. “Backbone chain” as used herein, refers to the chain of linking moieties covalently linking two linked moieties. In various embodiments, the linkers of the invention can have a length from 1 to 250 or more atoms. For the purposes herein, the length of a backbone chain of a linker is defined by the shortest contiguous chain of atom(s) linking the two linked moieties.

In various embodiments of the invention, the first linker can have a length from 1 to 50 atoms, such as from 1 to 25 atoms, from 1 to 20 atoms, from 1 to 15 atoms, from 1 to 10 atoms, from 1 to 9 atoms, from 1 to 8 atoms, from 1 to 7 atoms, from 1 to 6 atoms, from 1 to 5 atoms, from 1 to 4 atoms, from 1 to 3 atoms, from 1 to 2 atoms, or 1 atom.

In some versions of the invention, the first linker can comprise one or more instances of one or more first linking moieties. In some versions, the one or more first linking moieties comprise or are selected from optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroaralkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted arylene, optionally substituted heteroarylene, oxygen, sulfur, phosphorus, silicon, secondary amine, tertiary amine, carbonyl, and sulfonyl.

In some versions, the one or more first linking moieties comprise or are selected from optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroaralkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted arylene, optionally substituted heteroarylene, oxygen, sulfur, phosphorus, silicon, secondary amine, tertiary amine, carbonyl, and sulfonyl.

In some versions, the one or more first linking moieties comprise or are selected from alkylene, alkenylene, alkynylene, heteroaralkylene, heteroalkenylene, heteroalkynylene, arylene, heteroarylene, oxygen, sulfur, phosphorus, silicon, amine, carbonyl, and sulfonyl.

In some versions, the one or more first linking moieties comprise or are selected from optionally substituted alkylene, optionally substituted arylene (*e.g.*, *para*-phenylene, *meta*-phenylene, *ortho*-phenylene), oxygen, secondary amine, tertiary amine, and carbonyl.

In some versions, the one or more first linking moieties comprise or are selected from alkylene, arylene (*e.g.*, *para*-phenylene, *meta*-phenylene, *ortho*-phenylene), oxygen, secondary amine, and carbonyl.

In some versions, the one or more first linking moieties, comprise or are selected from $-(CH_2)_x-$, $-(CH_2CH_2O)_x-$, $-(OCH_2CH_2)_x-$, $-(CH_2O)_x-$, $-(OCH_2)_x-$, -C(O)NH-, -NHC(O)-, *para*-phenylene, *meta*-phenylene, and *ortho*-phenylene, wherein x in each instance is independently an integer from 1 to 6, such as 1, 2, 3, 4, 5, or 6.

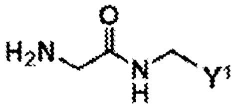
In some versions, the one or more first linking moieties, comprise or are selected from: $-(CH_2)_x-$; $-(CH_2)_x-$, $-(CH_2CH_2O)_x-$, $-(OCH_2CH_2)_x-$, $-(CH_2O)_x-$, and $-(OCH_2)_x-$; $-(CH_2)_x-$, -C(O)NH-, and -NHC(O)-; or *para*-phenylene, *meta*-phenylene, and *ortho*-phenylene, wherein x in each instance is independently an integer from 1 to 6, such as 1, 2, 3, 4, 5, or 6.

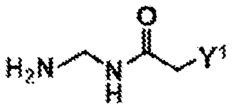
In some versions, the first linking moieties are devoid of an ester group (-C(O)O-). In some versions, the backbone chain of the first linker is devoid of an ester group. In some versions, the first linker is devoid of an ester group.

The term “click chemistry handle,” as used herein, refers to a reactant, or a reactive group, that can partake in a click chemistry reaction. The term “click chemistry” refers to a chemical philosophy introduced by K. Barry Sharpless of The Scripps Research Institute, describing chemistry tailored to generate covalent bonds quickly and reliably by joining small units comprising reactive groups together. In some embodiments, click chemistry reactions are modular, wide in scope, give high chemical yields, generate inoffensive byproducts, are stereospecific, exhibit a large thermodynamic driving force >84 kJ/mol to favor a reaction with a single reaction product, and/or can be carried out under physiological conditions. A distinct exothermic reaction makes a reactant “spring loaded”. In some embodiments, a click chemistry reaction exhibits high atom economy, can be carried out under simple reaction conditions, use readily available starting materials and reagents, uses no toxic solvents or use a solvent that is benign or easily removed (preferably water), and/or provides simple product isolation by non-chromatographic methods (crystallization or distillation). In general, click chemistry reactions require at least two molecules comprising click chemistry handles that can react with each other. Such click chemistry handle pairs that are reactive with each other are sometimes referred to herein as partner click chemistry handles. For example, an azide is a partner click chemistry handle to a cyclooctyne or any other alkyne. Exemplary click chemistry handles suitable for use here are shown in FIGS. 3A and 3B and Table 1.

In some versions, the first click chemistry handle is selected from the group consisting of a terminal alkyne, an azide, a strained alkyne, a diene, a dieneophile, an alkoxyamine, a phosphine, a hydrazide, a thiol, a tetrazine, and an alkene. In some versions, the first click chemistry handle is selected from the group consisting of a terminal alkyne and an azide.

Exemplary first labeling agents include:

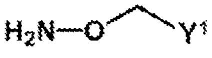


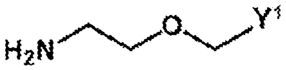
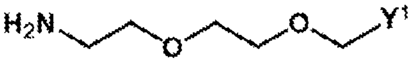
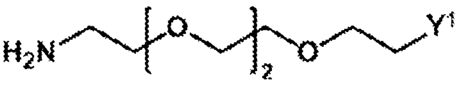
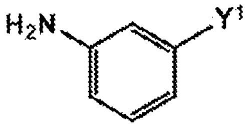
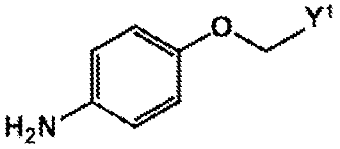
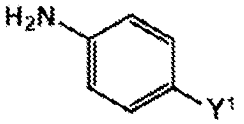
;


;


;


;


;

;

;

;

;

;

;

; and

,

wherein Y^1 is a click chemistry handle.

The methods of the invention can further comprise a second labeling reaction. The second labeling reaction can comprise reacting the protein sample in a second labeling reaction with a second labeling agent to label a second set of protein residues in the protein sample with the second labeling agent. "Second set of protein residues" in this context refers to the set of individual amino acid residues in the proteins in the protein sample that are labeled (chemically reacted) with the second labeling agent. The second labeling reagent can label amino acid residues in the proteins by reacting with the first label on the first set of protein residues. The second set of protein residues can thereby comprise or be a subset of the first set of protein residues. The term "subset" defines the second set of protein residues such that all the protein residues in the second set are protein residues in the first set. In such a case, the second set of protein residues can be equal to the first set of protein residues or can be a proper subset thereof (*i.e.*, wherein the second set of protein residues is equal to the first set of protein residues).

The second labeling agent can comprise a click chemistry handle, referred to as a "second click chemistry handle" in the second labeling agent. The second labeling agent can further comprise a label. The second click chemistry handle can be connected to the label via a linker, the latter referred to as a "second linker" in the second labeling agent. The first labeling agent can accordingly have the structure of Y^2-L^2-Z, wherein Y^2 is a second click chemistry handle, L^2 is a second linker, and Z is a label.

The second click chemistry handle is suitably a partner click chemistry handle of the first click chemistry handle. The second click chemistry handle in some versions is selected from the group consisting of a terminal alkyne, an azide, a strained alkyne, a diene, a dieneophile, an alkoxyamine, a phosphine, a hydrazide, a thiol, a tetrazine, and an alkene.

The second linker can be any linker and can have any characteristic described herein for linkers generally. The second linker can comprise any characteristic described herein for the first linker, except that the second linker preferably comprises a cleavable linking moiety as one of its linking moieties. The second linker in some versions is a cleavable linker, wherein a cleavable linker is a linker comprising a cleavable linker moiety.

In various embodiments of the invention, the second linker can have a length from 1 to 250 or more atoms, such as from 1-50 atoms.

In some versions of the invention, the second linker can comprise one or more instances of one or more second linking moieties. In some versions, the one or more second linking moieties comprise or are selected from optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroaralkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted arylene, optionally substituted heteroarylene, oxygen, sulfur, phosphorus, silicon, secondary amine, tertiary amine, carbonyl, sulfonyl, and a cleavable linking moiety.

In some versions, the one or more second linking moieties comprise or are selected from optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroaralkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted arylene, optionally substituted heteroarylene, oxygen, sulfur, phosphorus, silicon, secondary amine, tertiary amine, carbonyl, sulfonyl, and a cleavable linking moiety.

In some versions, the one or more second linking moieties comprise or are selected from alkylene, alkenylene, alkynylene, heteroaralkylene, heteroalkenylene, heteroalkynylene, arylene, heteroarylene, oxygen, sulfur, phosphorus, silicon, secondary amine, carbonyl, sulfonyl, and a cleavable linking moiety.

In some versions, the one or more second linking moieties comprise or are selected from optionally substituted alkylene, optionally substituted arylene (*e.g.*, *para*-phenylene, *meta*-phenylene, *ortho*-phenylene), oxygen, secondary amine, tertiary amine, carbonyl, and a cleavable linking moiety.

In some versions, the one or more second linking moieties comprise or are selected from alkylene, arylene (*e.g.*, *para*-phenylene, *meta*-phenylene, *ortho*-phenylene), oxygen, secondary amine, carbonyl, and a cleavable linking moiety.

In some versions, the one or more second linking moieties, comprise or are selected from $-(CH_2)_x-$, $-(CH_2CH_2O)_x-$, $-(OCH_2CH_2)_x-$, $-(CH_2O)_x-$, $-(OCH_2)_x-$, -C(O)NH-, -NHC(O)-, *para*-phenylene, *meta*-phenylene, *ortho*-phenylene, and a cleavable linking moiety, wherein x in each instance is independently an integer from 1 to 6, such as 1, 2, 3, 4, 5, or 6.

In some versions, the one or more second linking moieties, comprise or are selected from: $-(CH_2)_x-$; $-(CH_2)_x-$, $-(CH_2CH_2O)_x-$, $-(OCH_2CH_2)_x-$, $-(CH_2O)_x-$, and $-(OCH_2)_x-$; $-(CH_2)_x-$, -C(O)NH-, and -NHC(O)-; or *para*-phenylene, *meta*-phenylene, *ortho*-phenylene, and a cleavable linking moiety, wherein x in each instance is independently an integer from 1 to 6, such as 1, 2, 3, 4, 5, or 6.

"Label" as used herein is a chemical moiety that: contains an isotopic moiety, which may be radioactive or a heavy isotope, including, but not limited to, ^{2}H, ^{3}H, ^{13}C, ^{14}C, ^{14}N, ^{15}N, ^{18}F, ^{31}P, ^{32}P, ^{35}S, ^{67}Ga, ^{76}Br, ^{99m}Tc (Tc-99m), ^{111}In, ^{123}I, ^{125}I, ^{131}I, ^{153}Gd, ^{169}Yb, and ^{186}Re; contains an immune moiety, which may be an antibody or antigen, which may be bound to an enzyme (*e.g.*, such as horseradish peroxidase); comprises a colored, luminescent, phosphorescent, or fluorescent moiety (*e.g.*, such as the fluorescent label fluoresceinisothiocyanat (FITC); comprises one or more photo affinity moieties; and/or comprises a ligand for one or more known binding partners (e.g., biotin-streptavidin, FK506-FKBP). In certain embodiments, a label comprises a radioactive isotope, preferably an isotope which emits detectable particles, such as β particles. In certain embodiments, the label comprises a fluorescent moiety. In certain embodiments, the label is the fluorescent label fluoresceinisothiocyanat (FITC). In certain embodiments, the label comprises a ligand moiety with one or more known binding partners. In certain embodiments, the label comprises biotin. In some embodiments, a label is a fluorescent polypeptide (e.g., GFP or a derivative thereof such as enhanced GFP (EGFP)) or a luciferase (e.g., a firefly, *Renilla*, or *Gaussia* luciferase). It will be appreciated that, in certain embodiments, a label may react with a suitable substrate (*e.g.*, a luciferin) to generate a detectable signal. Non-limiting examples of fluorescent proteins include GFP and derivatives thereof, proteins comprising chromophores that emit light of different colors such as red, yellow, and cyan fluorescent proteins, *etc*. Exemplary fluorescent proteins include, e.g., Sirius, Azurite, EBFP2, TagBFP, mTurquoise, ECFP, Cerulean, TagCFP, mTFP1, mUkGl, mAG1, AcGFP1, TagGFP2, EGFP, mWasabi, EmGFP, TagYPF, EYFP, Topaz, SYFP2, Venus, Citrine, mKO, mKO2, mOrange, mOrange2, TagRFP, TagRFP-T, mStrawberry, mRuby, mCherry, mRaspberry, mKate2, mPlum, mNeptune, T-Sapphire, mAmetrine, mKeima. See, e.g., Chalfie, M. and Kain, SR (eds.) Green fluorescent protein: properties, applications, and protocols (Methods of biochemical analysis, v. 47). Wiley-Interscience, Hoboken, N.J., 2006, and/or Chudakov, D M, et al., Physiol Rev. 90(3):1103-63, 2010 for discussion of GFP and numerous other fluorescent or luminescent proteins.

In some versions, the label on the second labeling agent comprises a ligand, such as biotin.

Exemplary second labeling agents are provided in the following examples, and include PC biotin azide (1-[4-[[22-[(3aR,4R,6aS)-Hexahydro-2-oxo-1H-thieno[3,4-d]imidazol-4-yl]-4,18-dioxo-8,11,14-trioxa-5,17-diazadocos-1-yl]oxy]-5-methoxy-2-nitrophenyl]ethyl N-(3-azidopropyl)carbamate), among others (FIG. 16).

The methods of the invention can further comprise fragmenting the protein sample. “Fragmenting” a protein sample as used herein refers to breaking down proteins present in a given protein sample into smaller polypeptides. Any method of protein fragmentation can be used in the present methods. Exemplary protein fragmentation methods include enzymatic digestion (*e.g.*, protease digestion), chemical fragmentation (*e.g.*, CNBr), and physical fragmentation (*e.g.*, sonication), among others.

The fragmenting step is preferably performed after the first labeling reaction but can be performed either prior to or after the second labeling reaction. Accordingly, in some versions, the fragmenting is performed after the first labeling reaction and prior to the second labeling reaction. In some versions, the fragmenting is performed after the first labeling reaction and after the second labeling reaction.

The methods of the invention can further comprise separating labeled polypeptides from unlabeled polypeptides. The labeled polypeptides in this step can be defined as polypeptides comprising at least one protein residue from the second set of protein residues. The unlabeled polypeptides in this step can be defined as polypeptides lacking any of the second set of protein residues. Any method suitable for separating labeled polypeptides from unlabeled polypeptides can be used in the separating step. The particular method used in some cases will depend on the nature of the label employed in the second labeling agent. Examples include affinity chromatography, centrifugation, and antibody-mediated precipitation, among others. The separating step can be performed after the second labeling reaction. If a fragmenting step is performed, the separating step is preferably performed after the fragmenting step.

The methods of the invention can further comprise a step of cleaving the second linkers. The cleaving step can comprise cleaving a cleavable linking moiety in the second linker. The cleaving step can separate the labeled polypeptides from the label. The cleaving step can be performed after the separating step to thereby release the labeled polypeptides from any agents (*e.g.*, beads) or substrates (*e.g.*, solid substrates) employed in the separating step.

The methods of the invention can further comprise a step of determining the amino acid sequences of the labeled polypeptides. The determining step is preferably performed after any fragmenting, separating, and cleaving steps are performed. Any method of determining amino acid sequences of peptides can be used. Exemplary methods include mass spectrometry, Sanger sequencing, next-generation sequencing, among others.

The sequences obtained from the sequencing step can be aligned with the sequences of the proteins in the original protein sample to determine protein residues or domains susceptible to labeling via the first labeling reaction.

The methods of the invention can be used to infer changes in protein three-dimensional structure upon exposure to various conditions. Exemplary conditions can include differences in temperature, differences in pH, or differences in the presence or concentration of one or more molecules, such as drug candidates. In the case of differences in the presence or concentration of one or more molecules, the methods of the invention can be used to infer protein-drug interactions, particularly with specific proteins in a protein mixture.

Accordingly, in some versions of the invention, the protein sample can comprise a first protein sample and a second protein sample. The proteome of the first protein sample can be identical to the proteome of the second protein sample during one or more steps of the methods of the invention, such as during the first labeling reaction and also preferably during the second labeling reaction. “Proteome” as used herein with respect to a sample refers to the set of proteins present in a given sample. In some versions, the first protein sample can comprise at least one difference with respect to the second protein sample. Such differences can comprise a difference in temperature, a difference in pH, a difference in the presence or concentration of one or more molecules, a difference in a sequence of one or more proteins in the protein samples (wherein the first protein sample comprises a first sequence of a given protein and the second protein sample comprises a second sequence, such as a sequence variant or mutant, of the given protein), or any other condition of interest for testing differences in protein labeling characteristics (due to, for example, changes in protein conformational structure). In some versions, after determining the sequences of the labeled peptides derived from the first and second protein samples, the sequences of the labeled polypeptides from the first protein sample can be compared with the sequences of the labeled polypeptides from the second protein sample. As outlined above as well as below in the examples, the presence of differently labeled proteins can reveal changes in the conformation of such proteins when exposed to the tested condition.

In various versions of the invention, the protein sample used in the methods of the invention can be in the form of an initial protein sample and a processed protein sample. The initial protein sample refers to the form of the protein sample prior to and leading up to the point of initiation of any of the processing steps of the invention (*e.g.*, reacting steps, fragmenting steps, separating steps, amino acid determining steps, *etc.*, as discussed below). The processed protein sample refers to the form of the protein sample after any one or more of the processing steps of the invention have been initiated, are in process, or have been completed.

In some versions of the invention, the initial and/or processed protein sample can be in the form of a protein mixture.

In various versions of the invention, the protein mixture in the initial and/or in the processed protein sample during or after any given processing step can comprise at least 10, at least 25, at least 50, at least 75, at least 100, at least 500, at least 1,000, at least 2,500, or at least 5,000 different proteins. In various versions of the invention, the protein mixture in the initial and/or in the processed protein sample during or after any given processing step can comprise up to 1,000, up to 2,500, up to 5,000, up to 7,500, up to 10,000, up to 15,000, up to 20,000, or up to 50,000 or more different proteins. "Different proteins" in this context refers to proteins having sequences with at least one substitution, insertion, or deletion with respect to each other.

In some versions, the proteins in the initial protein sample and/or processed protein sample during or after at least some of the processing steps (*e.g.*, the first and/or second labeling reactions) can be in a folded state.

In some versions, the initial protein sample and/or processed protein sample during or after at least some of the processing steps (*e.g.*, the first and/or second labeling reactions) comprises a full cellular proteome. "Full cellular proteome" refers to the full set of proteins present in a cell prior to cell lysis. In some versions, the initial protein sample and/or processed protein sample during or after at least some of the processing steps (*e.g.*, the first and/or second labeling reactions) comprises a full tissue proteome. "Full tissue proteome" refers to the full set of proteins present in a tissue prior to processing of the tissue. In some versions, the initial protein sample and/or processed protein sample during or after at least some of the processing steps (*e.g.*, the first and/or second labeling reactions) comprises a full organ proteome. "Full organ proteome" refers to the full set of proteins present in an organ prior to processing of the tissue. Exemplary tissues, include muscle, skin, brain, bone, lungs, liver, bladder, kidneys, heart, stomach, intestines pancreas, kidney, pharynx, larynx, arteries, liver gallbladder, bladder veins, bone marrow, spleen, lymph nodes. For the purposes herein, blood is considered an organ. In some versions, the initial protein sample and/or processed protein sample during or after at least some of the processing steps (*e.g.*, the first and/or second labeling reactions) comprises a full lesion proteome. "Full organ proteome" refers to the full set of proteins present in a bodily lesion prior to processing of the tissue. "Lesion" as used herein refers to an abnormal or diseased bodily cell, tissue, or organ. In some versions, the lesion comprises a solid tumor.

In some versions, the protein sample can be split into two or more portions at any point during the process, and different first labeling reagents and/or second reagents can be used for the two or more portions in the first labeling reaction and/or the second labeling reaction. In some versions, the protein sample is split into two or more portions prior to the first labeling reaction, and the first labeling reaction is performed on each of the two or portions with two or more different first labeling reagents. The two or more different first labeling agents can comprise at least one difference between them, such as having different nucleophiles, different linkers, and/or click chemistry handles. In some versions, the protein sample is split into two or more portions prior to the second labeling reaction, and the second labeling reaction is performed on the two or more portions with two or more different second labeling reagents. The two or more second labeling agents can comprise at least one difference between them, such as having different click chemistry handles, different likers, and/or different labels. In some versions, the protein sample is split into two or more portions after the first labeling reaction but prior to the second labeling reaction, and the second labeling reaction is performed with two or more different second labeling reagents. The two or more second labeling agents can comprise at least one difference between them, such having different click chemistry handles, different likers, and/or different labels. In some versions, two or more different protein samples are processed in parallel, and different first labeling reagents and/or second reagents can be used for the two or more different protein samples in the first labeling reaction and/or the second labeling reaction.

Similarly, the protein sample can comprise multiple protein samples (*e.g.*, by being split as described above) and each of the multiple protein samples can be sample-labeled with a with detectably different sample labels. The multiple protein samples can subsequently be combined for downstream processing and/or analysis under a number of different workflow protocols as described and exemplified elsewhere herein. The detectably different sample labels can include any type of label described herein. In some versions, the detectably different sample labels are mass labels. Detectably different mass labels are labels that have or can be processed to have different masses. Exemplary mass labels are tandem mass tag labels as described and exemplified elsewhere herein. Tandem mass tags are described in Thompson et al. 2003 (Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Johnstone R, Mohammed AK, Hamon C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem. 2003 Apr 15;75(8):1895-904), Zecha et al. 2019, (Zecha J, Satpathy S, Kanashova T, Avanessian SC, Kane MH, Clauser KR, Mertins P, Carr SA, Kuster B. TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. Mol Cell Proteomics. 2019 Jul;18(7):1468-1478), and elsewhere in the art.

Some embodiments can comprise "multiple labeling" of a protein sample in either the first labeling reaction and/or the second labeling reaction. For example, a protein sample can be labeled in a first labeling reaction with an initial first labeling agent. A buffer exchange can be performed to remove excess reagent while maintaining the protein. A stimulus of effector can then be added to the protein sample, and the protein sample can then be again labeled in a first labeling reaction with a first labeling agent that is different from the initial first labeling agent. This process can be iterated several times. The various different first labeling agents can comprise at least one difference between them, such as having different nucleophiles, different linkers, and/or click chemistry handles. In some embodiments, a protein sample can be labeled in a second labeling reaction with an initial second labeling agent. A buffer exchange can be performed to remove excess reagent while maintaining the protein. A stimulus of effector can then be added to the protein sample, and the protein sample can then be again labeled in a second labeling reaction with a second labeling agent that is different from the initial second labeling agent. This process can be iterated several times. The various different second labeling agents can comprise at least one difference between them, such having different click chemistry handles, different likers, and/or different labels.

Definitions of specific functional groups and chemical terms are described in more detail below. For purposes of this invention, the chemical elements are identified in accordance with the Periodic Table of the Elements, CAS version, Handbook of Chemistry and Physics, 75th Ed., inside cover, and specific functional groups are generally defined as described therein. Additionally, general principles of organic chemistry, as well as specific functional moieties and reactivity, are described in Organic Chemistry, Thomas Sorrell, University Science Books, Sausalito, 1999; Smith and March March's Advanced Organic Chemistry, 5th Edition, John Wiley & Sons, Inc., New York, 2001; Larock, Comprehensive Organic Transformations, VCH Publishers, Inc., New York, 1989; Carruthers, Some Modern Methods of Organic Synthesis, 3rd Edition, Cambridge University Press, Cambridge, 1987.

The term "aliphatic," as used herein, includes both saturated and unsaturated, nonaromatic, straight chain (i.e., unbranched), branched, acyclic, and cyclic (i.e., carbocyclic) hydrocarbons, which are optionally substituted with one or more functional groups. As will be appreciated by one of ordinary skill in the art, “aliphatic” is intended herein to include, but is not limited to, alkyl, alkenyl, alkynyl, cycloalkyl, cycloalkenyl, and cycloalkynyl moieties. Thus, as used herein, the term “alkyl” includes straight, branched and cyclic alkyl groups. An analogous convention applies to other generic terms such as “alkenyl,” “alkynyl,” and the like. Furthermore, as used herein, the terms “alkyl,” “alkenyl,” “alkynyl,” and the like encompass both substituted and unsubstituted groups. In certain embodiments, as used herein, “aliphatic” is used to indicate those aliphatic groups (cyclic, acyclic, substituted, unsubstituted, branched or unbranched) having 1-20 carbon atoms (C_{1-20} aliphatic). In certain embodiments, the aliphatic group has 1-10 carbon atoms (C_{1-10} aliphatic). In certain embodiments, the aliphatic group has 1-6 carbon atoms (C_{1-6} aliphatic). In certain embodiments, the aliphatic group has 1-5 carbon atoms (C_{1-5} aliphatic). In certain embodiments, the aliphatic group has 1-4 carbon atoms (C_{1-4} aliphatic). In certain embodiments, the aliphatic group has 1-3 carbon atoms (C_{1-3} aliphatic). In certain embodiments, the aliphatic group has 1-2 carbon atoms (C_{1-2} aliphatic). Aliphatic group substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety.

The term “alkyl,” as used herein, refers to saturated, straight- or branched-chain hydrocarbon radicals derived from a hydrocarbon moiety containing between one and twenty carbon atoms by removal of a single hydrogen atom. In some embodiments, the alkyl group employed in the invention contains 1-20 carbon atoms (C_{1-20}alkyl). In another embodiment, the alkyl group employed contains 1-15 carbon atoms (C_{1-15}alkyl). In another embodiment, the alkyl group employed contains 1-10 carbon atoms (C_{1-10}alkyl). In another embodiment, the alkyl group employed contains 1-8 carbon atoms (C_{1-8}alkyl). In another embodiment, the alkyl group employed contains 1-6 carbon atoms (C_{1-6}alkyl). In another embodiment, the alkyl group employed contains 1-5 carbon atoms (C_{1-5}alkyl). In another embodiment, the alkyl group employed contains 1-4 carbon atoms (C_{1-4}alkyl). In another embodiment, the alkyl group employed contains 1-3 carbon atoms (C_{1-3}alkyl). In another embodiment, the alkyl group employed contains 1-2 carbon atoms (C_{1-2}alkyl). Examples of alkyl radicals include, but are not limited to, methyl, ethyl, n-propyl, isopropyl, n-butyl, iso-butyl, sec-butyl, sec-pentyl, iso-pentyl, tert-butyl, n-pentyl, neopentyl, n-hexyl, sec-hexyl, n-heptyl, n-octyl, n-decyl, n-undecyl, dodecyl, and the like, which may bear one or more substituents. Alkyl group substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety. The term “alkylene,” as used herein, refers to a biradical derived from an alkyl group, as defined herein, by removal of two hydrogen atoms. Alkylene groups may be cyclic or acyclic, branched or unbranched, substituted or unsubstituted. Alkylene group substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety.

The term “alkenyl,” as used herein, denotes a monovalent group derived from a straight- or branched-chain hydrocarbon moiety having at least one carbon-carbon double bond by the removal of a single hydrogen atom. In certain embodiments, the alkenyl group employed in the invention contains 2-20 carbon atoms (C_{2-20}alkenyl). In some embodiments, the alkenyl group employed in the invention contains 2-15 carbon atoms (C_{2-15}alkenyl). In another embodiment, the alkenyl group employed contains 2-10 carbon atoms (C_{2-10}alkenyl). In still other embodiments, the alkenyl group contains 2-8 carbon atoms (C_{2-8}alkenyl). In yet other embodiments, the alkenyl group contains 2-6 carbons (C_{2-6}alkenyl). In yet other embodiments, the alkenyl group contains 2-5 carbons (C_{2-5}alkenyl). In yet other embodiments, the alkenyl group contains 2-4 carbons (C_{2-4}alkenyl). In yet other embodiments, the alkenyl group contains 2-3 carbons (C_{2-3}alkenyl). In yet other embodiments, the alkenyl group contains 2 carbons (C_2alkenyl). Alkenyl groups include, for example, ethenyl, propenyl, butenyl, 1-methyl-2-buten-1-yl, and the like, which may bear one or more substituents. Alkenyl group substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety. The term "alkenylene," as used herein, refers to a biradical derived from an alkenyl group, as defined herein, by removal of two hydrogen atoms. Alkenylene groups may be cyclic or acyclic, branched or unbranched, substituted or unsubstituted. Alkenylene group substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety.

The term "alkynyl," as used herein, refers to a monovalent group derived from a straight- or branched-chain hydrocarbon having at least one carbon-carbon triple bond by the removal of a single hydrogen atom. In certain embodiments, the alkynyl group employed in the invention contains 2-20 carbon atoms ($C_{2\text{-}20}$alkynyl). In some embodiments, the alkynyl group employed in the invention contains 2-15 carbon atoms ($C_{2\text{-}15}$alkynyl). In another embodiment, the alkynyl group employed contains 2-10 carbon atoms ($C_{2\text{-}10}$alkynyl). In still other embodiments, the alkynyl group contains 2-8 carbon atoms ($C_{2\text{-}8}$alkynyl). In still other embodiments, the alkynyl group contains 2-6 carbon atoms ($C_{2\text{-}6}$alkynyl). In still other embodiments, the alkynyl group contains 2-5 carbon atoms ($C_{2\text{-}5}$alkynyl). In still other embodiments, the alkynyl group contains 2-4 carbon atoms ($C_{2\text{-}4}$alkynyl). In still other embodiments, the alkynyl group contains 2-3 carbon atoms ($C_{2\text{-}3}$alkynyl). In still other embodiments, the alkynyl group contains 2 carbon atoms (C_2alkynyl). Representative alkynyl groups include, but are not limited to, ethynyl, 2-propynyl (propargyl), 1-propynyl, and the like, which may bear one or more substituents. Alkynyl group substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety. The term "alkynylene," as used herein, refers to a biradical derived from an alkynylene group, as defined herein, by removal of two hydrogen atoms. Alkynylene groups may be cyclic or acyclic, branched or unbranched, substituted or unsubstituted. Alkynylene group substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety.

The term "carbocyclic" or "carbocyclyl" as used herein, refers to an as used herein, refers to a cyclic aliphatic group containing 3-10 carbon ring atoms ($C_{3\text{-}10}$carbocyclic). Carbocyclic group substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety.

The term "heteroaliphatic," as used herein, refers to an aliphatic moiety, as defined herein, which includes both saturated and unsaturated, nonaromatic, straight chain (i.e., unbranched), branched, acyclic, cyclic (i.e., heterocyclic), or polycyclic hydrocarbons, which are optionally substituted with one or more functional groups, and that further contains one or more heteroatoms (e.g., oxygen, sulfur, nitrogen, phosphorus, or silicon atoms) between carbon atoms. In certain embodiments, heteroaliphatic moieties are substituted by independent replacement of one or more of the hydrogen atoms thereon with one or more substituents. As will be appreciated by one of ordinary skill in the art, "heteroaliphatic" is intended herein to include, but is not limited to, heteroalkyl, heteroalkenyl, heteroalkynyl, heterocycloalkyl, heterocycloalkenyl, and heterocycloalkynyl moieties. Thus, the term "heteroaliphatic" includes the terms "heteroalkyl," "heteroalkenyl," "heteroalkynyl," and the like. Furthermore, as used herein, the terms "heteroalkyl," "heteroalkenyl," "heteroalkynyl," and the like encompass both substituted and unsubstituted groups. In certain embodiments, as used herein, "heteroaliphatic" is used to indicate those heteroaliphatic groups (cyclic, acyclic, substituted, unsubstituted, branched or unbranched) having 1-20 carbon atoms and 1-6 heteroatoms ($C_{1\text{-}20}$heteroaliphatic). In certain embodiments, the heteroaliphatic group contains 1-10 carbon atoms and 1-4 heteroatoms (C_{1-10}heteroaliphatic). In certain embodiments, the heteroaliphatic group contains 1-6 carbon atoms and 1-3 heteroatoms (C_{1-6}heteroaliphatic). In certain embodiments, the heteroaliphatic group contains 1-5 carbon atoms and 1-3 heteroatoms (C_{1-5}heteroaliphatic). In certain embodiments, the heteroaliphatic group contains 1-4 carbon atoms and 1-2 heteroatoms (C_{1-4}heteroaliphatic). In certain embodiments, the heteroaliphatic group contains 1-3 carbon atoms and 1 heteroatom (C_{1-3}heteroaliphatic). In certain embodiments, the heteroaliphatic group contains 1-2 carbon atoms and 1 heteroatom (C_{1-2}heteroaliphatic). Heteroaliphatic group substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety.

The term “heteroalkyl,” as used herein, refers to an alkyl moiety, as defined herein, which contain one or more heteroatoms (e.g., oxygen, sulfur, nitrogen, phosphorus, or silicon atoms) in between carbon atoms. In certain embodiments, the heteroalkyl group contains 1-20 carbon atoms and 1-6 heteroatoms (C_{1-20}heteroalkyl). In certain embodiments, the heteroalkyl group contains 1-10 carbon atoms and 1-4 heteroatoms (C_{1-10}heteroalkyl). In certain embodiments, the heteroalkyl group contains 1-6 carbon atoms and 1-3 heteroatoms (C_{1-6}heteroalkyl). In certain embodiments, the heteroalkyl group contains 1-5 carbon atoms and 1-3 heteroatoms (C_{1-5}heteroalkyl). In certain embodiments, the heteroalkyl group contains 1-4 carbon atoms and 1-2 heteroatoms (C_{1-4}heteroalkyl). In certain embodiments, the heteroalkyl group contains 1-3 carbon atoms and 1 heteroatom (C_{1-3}heteroalkyl). In certain embodiments, the heteroalkyl group contains 1-2 carbon atoms and 1 heteroatom (C_{1-2}heteroalkyl). The term “heteroalkylene,” as used herein, refers to a biradical derived from an heteroalkyl group, as defined herein, by removal of two hydrogen atoms. Heteroalkylene groups may be cyclic or acyclic, branched or unbranched, substituted or unsubstituted. Heteroalkylene group substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety.

The term “heteroalkenyl,” as used herein, refers to an alkenyl moiety, as defined herein, which further contains one or more heteroatoms (e.g., oxygen, sulfur, nitrogen, phosphorus, or silicon atoms) in between carbon atoms. In certain embodiments, the heteroalkenyl group contains 2-20 carbon atoms and 1-6 heteroatoms (C_{2-20}heteroalkenyl). In certain embodiments, the heteroalkenyl group contains 2-10 carbon atoms and 1-4 heteroatoms (C_{2-10}heteroalkenyl). In certain embodiments, the heteroalkenyl group contains 2-6 carbon atoms and 1-3 heteroatoms (C_{2-6}heteroalkenyl). In certain embodiments, the heteroalkenyl group contains 2-5 carbon atoms and 1-3 heteroatoms (C_{2-5}heteroalkenyl). In certain embodiments, the heteroalkenyl group contains 2-4 carbon atoms and 1-2 heteroatoms (C_{2-4}heteroalkenyl). In certain embodiments, the heteroalkenyl group contains 2-3 carbon atoms and 1 heteroatom (C_{2-3}heteroalkenyl). The term “heteroalkenylene,” as used herein, refers to a biradical derived from an heteroalkenyl group, as defined herein, by removal of two hydrogen atoms. Heteroalkenylene groups may be cyclic or acyclic, branched or unbranched, substituted or unsubstituted.

The term “heteroalkynyl,” as used herein, refers to an alkynyl moiety, as defined herein, which further contains one or more heteroatoms (e.g., oxygen, sulfur, nitrogen, phosphorus, or silicon atoms) in between carbon atoms. In certain embodiments, the heteroalkynyl group contains 2-20 carbon atoms and 1-6 heteroatoms (C_{2-20}heteroalkynyl). In certain embodiments, the heteroalkynyl group contains 2-10 carbon atoms and 1-4 heteroatoms (C_{2-10}heteroalkynyl). In certain embodiments, the heteroalkynyl group contains 2-6 carbon atoms and 1-3 heteroatoms (C_{2-6}heteroalkynyl). In certain embodiments, the heteroalkynyl group contains 2-5 carbon atoms and 1-3 heteroatoms (C_{2-5}heteroalkynyl). In certain embodiments, the heteroalkynyl group contains 2-4 carbon atoms and 1-2 heteroatoms (C_{2-4}heteroalkynyl). In certain embodiments, the heteroalkynyl group contains 2-3 carbon atoms and 1 heteroatom (C_{2-3}heteroalkynyl). The term "heteroalkynylene," as used herein, refers to a biradical derived from an heteroalkynyl group, as defined herein, by removal of two hydrogen atoms. Heteroalkynylene groups may be cyclic or acyclic, branched or unbranched, substituted or unsubstituted.

The term "heterocyclic," "heterocycles," or "heterocyclyl," as used herein, refers to a cyclic heteroaliphatic group. A heterocyclic group refers to a non-aromatic, partially unsaturated or fully saturated, 3- to 10-membered ring system, which includes single rings of 3 to 8 atoms in size, and bi- and tri-cyclic ring systems which may include aromatic five- or six-membered aryl or heteroaryl groups fused to a non-aromatic ring. These heterocyclic rings include those having from one to three heteroatoms independently selected from oxygen, sulfur, and nitrogen, in which the nitrogen and sulfur heteroatoms may optionally be oxidized and the nitrogen heteroatom may optionally be quaternized. In certain embodiments, the term heterocyclic refers to a non-aromatic 5-, 6-, or 7-membered ring or polycyclic group wherein at least one ring atom is a heteroatom selected from O, S, and N (wherein the nitrogen and sulfur heteroatoms may be optionally oxidized), and the remaining ring atoms are carbon, the radical being joined to the rest of the molecule via any of the ring atoms. Heterocycyl groups include, but are not limited to, a bi- or tri-cyclic group, comprising fused five, six, or seven-membered rings having between one and three heteroatoms independently selected from the oxygen, sulfur, and nitrogen, wherein (i) each 5-membered ring has 0 to 2 double bonds, each 6-membered ring has 0 to 2 double bonds, and each 7-membered ring has 0 to 3 double bonds, (ii) the nitrogen and sulfur heteroatoms may be optionally oxidized, (iii) the nitrogen heteroatom may optionally be quaternized, and (iv) any of the above heterocyclic rings may be fused to an aryl or heteroaryl ring. Exemplary heterocycles include azacyclopropanyl, azacyclobutanyl, 1,3-diazatidinyl, piperidinyl, piperazinyl, azocanyl, thiaranyl, thietanyl, tetrahydrothiophenyl, dithiolanyl, thiacyclohexanyl, oxiranyl, oxetanyl, tetrahydrofuranyl, tetrahydropuranyl, dioxanyl, oxathiolanyl, morpholinyl, thioxanyl, tetrahydronaphthyl, and the like, which may bear one or more substituents. Substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety.

The term "aryl," as used herein, refers to an aromatic mono- or polycyclic ring system having 3-20 ring atoms, of which all the ring atoms are carbon, and which may be substituted or unsubstituted. In certain embodiments of the present invention, "aryl" refers to a mono, bi, or tricyclic C_4-C_{20} aromatic ring system having one, two, or three aromatic rings which include, but are not limited to, phenyl, biphenyl, naphthyl, and the like, which may bear one or more substituents. Aryl substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety. The term "arylene," as used herein refers to an aryl biradical derived from an aryl group, as defined herein, by removal of two hydrogen atoms. Arylene groups may be substituted or unsubstituted. Arylene group substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety. Additionally, arylene groups may be incorporated as a linker group into an alkylene, alkenylene, alkynylene, heteroalkylene, heteroalkenylene, or heteroalkynylene group, as defined herein.

The term "heteroaryl," as used herein, refers to an aromatic mono- or polycyclic ring system having 3-20 ring atoms, of which one ring atom is selected from S, O, and N; zero, one, or two ring atoms are additional heteroatoms independently selected from S, O, and N; and the remaining ring atoms are carbon, the radical being joined to the rest of the molecule via any of the ring atoms. Exemplary heteroaryls include, but are not limited to pyrrolyl, pyrazolyl, imidazolyl, pyridinyl, pyrimidinyl, pyrazinyl, pyridazinyl, triazinyl, tetrazinyl, pyyrolizinyl, indolyl, quinolinyl, isoquinolinyl, benzoimidazolyl, indazolyl, quinolinyl, isoquinolinyl, quinolizinyl, cinnolinyl, quinazolynyl, phthalazinyl, naphthridinyl, quinoxalinyl, thiophenyl, thianaphthenyl, furanyl, benzofuranyl, benzothiazolyl, thiazolynyl, isothiazolyl, thiadiazolynyl, oxazolyl, isoxazolyl, oxadiaziolyl, oxadiaziolyl, and the like, which may bear one or more substituents. Heteroaryl substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety. The term "heteroarylene," as used herein, refers to a biradical derived from an heteroaryl group, as defined herein, by removal of two hydrogen atoms. Heteroarylene groups may be substituted or unsubstituted. Additionally, heteroarylene groups may be incorporated as a linker group into an alkylene, alkenylene, alkynylene, heteroalkylene, heteroalkenylene, or heteroalkynylene group, as defined herein. Heteroarylene group substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety.

The term "acyl," as used herein, is a subset of a substituted alkyl group, and refers to a group having the general formula $-C({=}O)R^A$, $-C({=}O)OR^A$, $-C({=}O)-O-C({=}O)R^A$, $-C({=}O)SR^A$, $-C({=}O)N(R^A)_2$, $-C({=}S)R^A$, $-C({=}S)N(R^A)_2$, and $-C({=}S)S(R^A)$, $-C({=}NR^A)R^A$, $-C({=}NR^A)OR^A$, $-C({=}NR^A)SR^A$, and $-C({=}NR^A)N(R^A)_2$, wherein R^A is hydrogen; halogen; substituted or unsubstituted hydroxyl; substituted or unsubstituted thiol; substituted or unsubstituted amino; acyl; optionally substituted aliphatic; optionally substituted heteroaliphatic; optionally substituted alkyl; optionally substituted alkenyl; optionally substituted alkynyl; optionally substituted aryl, optionally substituted heteroaryl, aliphaticoxy, heteroaliphaticoxy, alkyloxy, heteroalkyloxy, aryloxy, heteroaryloxy, aliphaticthioxy, heteroaliphaticthioxy, alkylthioxy, heteroalkylthioxy, arylthioxy, heteroarylthioxy, mono- or di-aliphaticamino, mono- or di-heteroaliphaticamino, mono- or di-alkylamino, mono- or di-heteroalkylamino, mono- or di-arylamino, or mono- or di heteroarylamino; or two R^A groups taken together form a 5- to 6-membered heterocyclic ring. Exemplary acyl groups include aldehydes (-CHO), carboxylic acids ($-CO_2H$), ketones, acyl halides, esters, amides, imines, carbonates, carbamates, and ureas. Acyl substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety.

The term "acylene," as used herein, is a subset of a substituted alkylene, substituted alkenylene, substituted alkynylene, substituted heteroalkylene, substituted heteroalkenylene, or substituted heteroalkynylene group, and refers to an acyl group having the general formulae: $-R^0-(C{=}X^1)-R^0-$, $-R-X^2(C{=}X^1)-R^0-$, or $-R^0-X^2(C{=}X^1)X^3-R^0-$, where X^1, X^2, and X^3 is, independently, oxygen, sulfur, or NR^r, wherein R^r is hydrogen or optionally substituted aliphatic, and R^0 is an optionally substituted alkylene, alkenylene, alkynylene, heteroalkylene, heteroalkenylene, or heteroalkynylene group, as defined herein. Exemplary acylene groups wherein R^0 is alkylene includes $-(CH_2)_T-O(C{=}O)-(CH_2)_T-$; $-(CH_2)_T-NR^r(C{=}O)-(CH_2)_T-$; $-(CH_2)_T-O(C{=}NR^r)-(CH_2)_T-$; $-(CH_2)_T-NR^r(C{=}NR^r)-(CH_2)_T-$; $-(CH_2)_T-(C{=}O)-(CH_2)_T-$; $-(CH_2)_T-(C{=}NR^r)-(CH_2)_T-$; $-(CH_2)_T-S(C{=}S)-(CH_2)_T-$; $-(CH_2)_T-NR^r(C{=}S)-(CH_2)-$; $-(CH_2)_T-S(C{=}NR^r)-(CH_2)_T-$; $-(CH_2)_T-O(C{=}S)-(CH_2)_T-$; $-(CH_2)_T-(C{=}S)-(CH_2)_T-$; or $-(CH_2)_T-S(C{=}O)-(CH_2)_T-$, and the like, which may bear one or more substituents; and wherein each instance of T is, independently, an integer between 0 to 20. Acylene substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety.

The term “amino” or “unsubstituted amino” as used herein, refers to a group of the formula (-NH_2). A “substituted amino” refers either to a mono-substituted amine (-NHR^h) or a disubstituted amine (-NR^h_2), wherein the R^h substituent is any substituent as described herein that results in the formation of a stable moiety (e.g., an amino protecting group; aliphatic, alkyl, alkenyl, alkynyl, heteroaliphatic, heterocyclic, aryl, heteroaryl, acyl, amino, nitro, hydroxyl, thiol, halo, aliphaticamino, heteroaliphaticamino, alkylamino, heteroalkylamino, arylamino, heteroarylamino, alkylaryl, arylalkyl, aliphaticoxy, heteroaliphaticoxy, alkyloxy, heteroalkyloxy, aryloxy, heteroaryloxy, aliphaticthioxy, heteroaliphaticthioxy, alkylthioxy, heteroalkylthioxy, arylthioxy, heteroarylthioxy, acyloxy, and the like, each of which may or may not be further substituted). In certain embodiments, the R^h substituents of the di-substituted amino group (-NR^h_2) form a 5- to 6-membered heterocyclic ring.

The term “secondary amine” refers to a group of the formula (-NH-). The term “tertiary amine” refers to a group of the formula (-NR^h-), wherein the R^h substituent is any substituent as described herein that results in the formation of a stable moiety (e.g., an amino protecting group; aliphatic, alkyl, alkenyl, alkynyl, heteroaliphatic, heterocyclic, aryl, heteroaryl, acyl, amino, nitro, hydroxyl, thiol, halo, aliphaticamino, heteroaliphaticamino, alkylamino, heteroalkylamino, arylamino, heteroarylamino, alkylaryl, arylalkyl, aliphaticoxy, heteroaliphaticoxy, alkyloxy, heteroalkyloxy, aryloxy, heteroaryloxy, aliphaticthioxy, heteroaliphaticthioxy, alkylthioxy, heteroalkylthioxy, arylthioxy, heteroarylthioxy, acyloxy, and the like, each of which may or may not be further substituted).

The term “hydroxy” or “hydroxyl,” as used herein, refers to a group of the formula (-OH). A “substituted hydroxyl” refers to a group of the formula (-OR^i), wherein R^i can be any substituent which results in a stable moiety (e.g., a hydroxyl protecting group; aliphatic, alkyl, alkenyl, alkynyl, heteroaliphatic, heterocyclic, aryl, heteroaryl, acyl, nitro, alkylaryl, arylalkyl, and the like, each of which may or may not be further substituted).

The term “thio” or “thiol,” as used herein, refers to a group of the formula (-SH). A “substituted thiol” refers to a group of the formula (-SR^r), wherein R^r can be any substituent that results in the formation of a stable moiety (e.g., a thiol protecting group; aliphatic, alkyl, alkenyl, alkynyl, heteroaliphatic, heterocyclic, aryl, heteroaryl, acyl, sulfinyl, sulfonyl, cyano, nitro, alkylaryl, arylalkyl, and the like, each of which may or may not be further substituted).

The term “imino,” as used herein, refers to a group of the formula (=NR^r), wherein R^r corresponds to hydrogen or any substituent as described herein, that results in the formation of a stable moiety (for example, an amino protecting group; aliphatic, alkyl, alkenyl, alkynyl, heteroaliphatic, heterocyclic, aryl, heteroaryl, acyl, amino, hydroxyl, alkylaryl, arylalkyl, and the like, each of which may or may not be further substituted).

The term “azide” or “azido,” as used herein, refers to a group of the formula (-N_3).

The terms “halo” and “halogen,” as used herein, refer to an atom selected from fluorine (fluoro, -F), chlorine (chloro, -Cl), bromine (bromo, -Br), and iodine (iodo, -I).

Exemplary substituents include aliphatic, alkyl, alkenyl, alkynyl, heteroaliphatic, heterocyclic, aryl, heteroaryl, acyl, amino, nitro, hydroxyl, thiol, halo, aliphaticamino, heteroaliphaticamino, alkylamino, heteroalkylamino, arylamino, heteroarylamino, alkylaryl, arylalkyl, aliphaticoxy, heteroaliphaticoxy, alkyloxy, heteroalkyloxy, aryloxy, heteroaryloxy, aliphaticthioxy, heteroaliphaticthioxy, alkylthioxy, heteroalkylthioxy, arylthioxy, heteroarylthioxy, and acyloxy.

"Cleavable linking moiety" as used herein refers to a moiety that is cleaved upon exposure to a chemical reagent, an enzyme, photo-irradiation, or other stimulus. Cleavable linking moieties are well known in the art. See, *e.g.*, Beard et al. 2021 (Beard HA, Korovesis D, Chen S, Verhelst SHL. Cleavable linkers and their application in MS-based target identification. Mol Omics. 2021 Apr 1;17(2):197-20), Leriche et al. 2012 (Leriche G, Chisholm L, Wagner A. Cleavable linkers in chemical biology. Bioorg Med Chem. 2012 Jan 15;20(2):571-82), Yang et al. 2017 (Yang Y, Fonović M, Verhelst SH. Cleavable Linkers in Chemical Proteomics Applications. Methods Mol Biol. 2017;1491:185-203), Sinz et al. 2017 (Sinz A. Divide and conquer: cleavable cross-linkers to study protein conformation and protein-protein interactions. Anal Bioanal Chem. 2017 Jan;409(1):33-44). Exemplary chemically labile linking moieties include acid-cleavable linking moieties (*e.g.*, hydrazone), and reduction-sensitive linking moieties (*e.g.*, disulfides and azo groups), among others (Beard HA, Korovesis D, Chen S, Verhelst SHL. Cleavable linkers and their application in MS-based target identification. Mol Omics. 2021 Apr 1;17(2):197-20) (Leriche G, Chisholm L, Wagner A. Cleavable linkers in chemical biology. Bioorg Med Chem. 2012 Jan 15;20(2):571-82) (Yang Y, Fonović M, Verhelst SH. Cleavable Linkers in Chemical Proteomics Applications. Methods Mol Biol. 2017;1491:185-203) (Sinz A. Divide and conquer: cleavable cross-linkers to study protein conformation and protein-protein interactions. Anal Bioanal Chem. 2017 Jan;409(1):33-44). Exemplary enzyme-cleavable linking moieties include valine-citrulline (vc) dipeptide (cleavable by cathepsin B), among others (Beard HA, Korovesis D, Chen S, Verhelst SHL. Cleavable linkers and their application in MS-based target identification. Mol Omics. 2021 Apr 1;17(2):197-20) (Leriche G, Chisholm L, Wagner A. Cleavable linkers in chemical biology. Bioorg Med Chem. 2012 Jan 15;20(2):571-82) (Yang Y, Fonović M, Verhelst SH. Cleavable Linkers in Chemical Proteomics Applications. Methods Mol Biol. 2017;1491:185-203) (Sinz A. Divide and conquer: cleavable cross-linkers to study protein conformation and protein-protein interactions. Anal Bioanal Chem. 2017 Jan;409(1):33-44). Exemplary photocleavable linking moieties include *ortho*-nitrobenzyloxy groups, *ortho*-nitrobenzylamino groups, α-substituted *ortho*-nitrobenzyl groups, *ortho*-nitroveratryl groups, phenacyl groups, *para*-methoxyphenacyl groups, and benzoin groups, among others (Christian Bochet and Sébastien Mercier. Author(s) of Chapter. "Photolabile Linker Units" Linker Strategies in Solid-Phase Organic Synthesis, edited by Peter J.H. Scott, John Wiley & Sons, Ltd., 23 October 2009, pp. 151-193).

Two moieties or agents comprising a click chemistry handle each (e.g., a first protein comprising a click chemistry handle providing a nucleophilic (Nu) group and a second protein comprising an electrophilic (E) group that can react with the Nu group of the first click chemistry handle) can be covalently conjugated under click chemistry reaction conditions. Click chemistry is a chemical philosophy introduced by Sharpless in 2001 and describes chemistry tailored to generate substances quickly and reliably by joining small units together (see, e.g., Kolb, Finn and Sharpless *Angewandte Chemie International Edition* (2001) 40: 2004-2021; Evans, *Australian Journal of Chemistry* (2007) 60: 384-395). Additional exemplary click chemistry handles, reaction conditions, and associated methods useful according to aspects of this invention are described in Joerg Lahann, *Click Chemistry for Biotechnology and Materials Science,* 2009, John Wiley & Sons Ltd, ISBN 978-0-470-69970-6, the entire contents of which are incorporated herein by reference.

Click chemistry can be modular, wide in scope, give high chemical yields, generate inoffensive byproducts, be stereospecific, be physiologically stable, exhibit a large thermodynamic driving force (e.g., >84 kJ/mol to favor a reaction with a single reaction product), and/or have high atom economy. Several reactions have been identified which fit this concept.

One reaction is the Huisgen 1,3-dipolar cycloaddition (e.g., the Cu(I)-catalyzed stepwise variant, often referred to simply as the "click reaction"; see, e.g., Tornoe et al., *Journal of Organic Chemistry* (2002) 67: 3057-3064). Copper and ruthenium are the commonly used catalysts in the reaction. The use of copper as a catalyst results in the formation of 1,4-regioisomer whereas ruthenium results in formation of the 1,5-regioisomer. Other click chemistry reactions include other cycloaddition reactions, such as the Diels-Alder reaction. Other click chemistry reactions include nucleophilic addition to small strained rings like epoxides and aziridines. Other click chemistry reactions include nucleophilic addition to activated carbonyl groups. Other click chemistry reactions include addition reactions to carbon-carbon double or triple bonds.

For two moieties or agents to be conjugated via click chemistry, the click chemistry handles of the moieties or agents have to be reactive with each other, for example, in that the reactive moiety of one of the click chemistry handles can react with the reactive moiety of the second click chemistry handle to form a covalent bond. Such reactive pairs of click chemistry handles are well known to those of skill in the art and include but are not limited to those described in FIGS. 3A and 3B and Table 1.

Table 1. Exemplary click chemistry handles and reactions.*

Handle 1	Handle 2	Mechanism	Notes on reaction
Azide	Alkyne	Cu-catalyzed [3 + 2] azide-akyne cycloaddition (CuAAC)	2 h at 60 °C in H_2O
Azide	Cyclooctyne	Strain promoted [3 + 2] azide-alkyne cycloaddition (SPAAC)	1 h at RT
Azide	Activated alkyne	[3 + 2] Huisgen cycloaddition	4 h at 50 °C
Azide	Electron-deficient alkyne	[3 + 2] cycloaddition	12 h at RT in H_2O
Azide	Aryne	[3 + 2] cycloaddition	4 h at RT in THF with crown ether or 24 h at RT in CH_3CN
Tetrazine	Alkene	Diels-Alder retro-[4 + 2] cycloaddition	40 min at 25 °C (N_2 is by-product)
Tetrazole	Alkene	1,3-Dipolar cycloaddition (photoclick)	Few min UV irradiation then overnight at 4 °C
Dithioester	Diene	Hetero-Diels-Alder reaction	10 min at RT
Anthracene	Madeimide	[4 + 2] Diels-Alder reaction	2 days reflux in toluene
Thiol	Alkene	Radical addition (thio click)	30 min UV (quantitative conv.) or 24 h UV irridation (>96%)
Thiol	Enone	Michael addition	24 h at RT in CH_3CN
Thiol	Maleimide	Michael addition	1 h at 40 °C in THF or 16 h

			at RT in dioxane
Thiol	*para*-Fluoro	Nucleophilic substitution	Overnight at RT in DMF or 60 min at 40°C in DMF
Amine	*para*-Fluoro	Nucleophilic substitution	20 min MW at 95 °C in NMP as solvent

* From Becer CR, Hoogenboom R, Schubert US. Click chemistry beyond metal-catalyzed cycloaddition. Angew Chem Int Ed Engl. 2009;48(27):4900-8.

RT = Room temperature

DMF = N,N-dimethylformamide

NMP = N-methylpyrolidone

THF = tetrahydrofuran

CH_3CN = acetonitrile

In some embodiments, click chemistry handles are used that can react to form covalent bonds in the absence of a metal catalyst. Such click chemistry handles are well known to those of skill in the art and include the click chemistry handles described in Becer, Hoogenboom, and Schubert, *click Chemistry beyond Metal-Catalyzed Cycloaddition*, Angewandte Chemie International Edition (2009) 48: 4900-4908.

Additional click chemistry handles suitable for use in the methods of the invention are well known to those of skill in the art, and such click chemistry handles include, but are not limited to, the click chemistry reaction partners, groups, and handles described in the art (H. C. Kolb, M. G. Finn, K. B. Sharpless, Angew. Chem. 2001, 113, 2056-2075) (C. J. Hawker, K. L. Wooley, Science 2005, 309, 1200-1205) (D. Fournier, R. Hoogenboom, U. S. Schubert, Chem. Soc. Rev. 2007, 36, 1369-1380) (W. H. Binder, R. Sachsenhofer, Macromol. Rapid Commun. 2007, 28, 15-54) (H. C. Kolb, K. B. Sharpless, Drug Discovery Today 2003, 8, 1128-1137) (V. D. Bock, H. Hiemstra, J. H. van Maarseveen, Eur. J. Org. Chem. 2006, 51-68) (V. O. Rodionov, V. V. Fokin, M. G. Finn, Angew. Chem. 2005, 117, 2250-2255) (P. L. Golas, N. V. Tsarevsky, B. S. Sumerlin, K. Matyjaszewski, Macromolecules 2006, 39, 6451-6457) (C. N. Urbani, C. A. Bell, M. R. Whittaker, M. J. Monteiro, Macromolecules 2008, 41, 1057-1060) (Chassaing, A. S. S. Sido, A. Alix, M. Kumarraja, P. Pale, J. Sommer, Chem. Eur. J. 2008, 14, 6713-6721) (B. C. Boren, S. Narayan, L. K. Rasmussen, L. Zhang, H. Zhao, Z. Lin, G. Jia, V. V. Fokin, J. Am. Chem. Soc. 2008, 130, 8923-8930) (B. Saba, S. Sharma, D. Sawant, B. Kundu, Synlett 2007, 1591-1594) (J. F. Lutz, Angew. Chem. 2008, 120, 2212-2214) (Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless, M. G. Finn, J. Am. Chem. Soc. 2003, 125, 3192-3193) (Gierlich, G. A. Burley, P. M. E. Gramlich, D. M. Hammond, T. Carell, Org. Lett. 2006, 8, 3639-3642) (J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli, C. R. Bertozzi, Proc. Natl. Acad. Sci. USA 2007, 104, 16793-16797) (S. T. Laughlin, J. M. Baskin, S. L. Amacher, C. R. Bertozzi, Science 2008, 320, 664-667) (J. A. Johnson, J. M. Baskin, C. R. Bertozzi, J. F. Koberstein, N. J. Turro, Chem. Commun. 2008, 3064-3066) (J. A. Codelli, J. M. Baskin, N. J. Agard, C. R. Bertozzi, J. Am. Chem. Soc. 2008, 130, 11486-11493) (E. M. Sletten, C. R. Bertozzi, Org. Lett. 2008, 10, 3097-3099) (J. M. Baskin, C. R. Bertozzi, QSAR Comb. Sci. 2007, 26, 1211-1219) (D. H. Ess, G. O. Jones, K. N. Houk, Org. Lett. 2008, 10, 1633-1636), Sharpless et al. 2005 (W. D. Sharpless, P. Wu, T. V. Hansen, J. G. Lindberg,

J. Chem. Educ. 2005, 82, 1833-1836) (Y. Zou, J. Yin, Bioorg. Med. Chem. Lett. 2008, 18, 5664-5667) (X. Ning, J. Guo, M. A. Wolfert, G. J. Boons, Angew. Chem. 2008, 120, 2285-2287) (S. Sawoo, P. Dutta, A. Chakraborty, R. Mukhopadhyay, O. Bouloussa, A. Sarkar, Chem. Commun. 2008, 5957-5959) (Z. Li, T. S. Seo, J. Ju, Tetrahedron Lett. 2004, 45, 3143-3146) (S. S. van Berkel, A. J. Dirkes, M. F. Debets, F. L. van Delft, J. J. L. Cornelissen, R. J. M. Nolte, F. P. J. Rutjes, ChemBioChem 2007, 8, 1504-1508) (S. S. van Berkel, A. J. Dirks, S. A. Meeuwissen, D. L. L. Pingen, O. C. Boerman, P. Laverman, F. L. van Delft, J. J. L. Cornelissen, F. P. J. Rutjes, ChemBio-Chem 2008, 9, 1805-1815) (F. Shi, J. P. Waldo, Y. Chen, R. C. Larock, Org. Lett. 2008, 10, 2409-2412) (L. Campbell-Verduyn, P. H. Elsinga, L. Mirfeizi, R. A. Dierckx, B. L. Feringa, Org. Biomol. Chem. 2008, 6, 3461-3463) (The Chemistry of the Thiol Group (Ed.: S. Patai), Wiley, New York, 1974) (A. F. Jacobine, In Radiation Curing in Polymer Science and Technology III (Eds.: J. D. Fouassier, J. F. Rabek), Elsevier, London, 1993, Chap. 7, pp. 219-268) (C. E. Hoyle, T. Y. Lee, T. Roper, J. Polym. Sci. Part A 2008, 42, 5301-5338) (L. M. Campos, K. L. Killops, R. Sakai, J. M. J. Paulusse, D. Damiron, E. Drockenmuller, B. W. Messmore, C. J. Hawker, Macromolecules 2008, 41, 7063-7070) (R. L. A. David, J. A. Kornfield, Macromolecules 2008, 41, 1151-1161) (C. Nilsson, N. Simpson, M. Malkoch, M. Johansson, E. Malmstrom, J. Polym. Sci. Part A 2008, 46, 1339-1348) (A. Dondoni, Angew. Chem. 2008, 120, 9133-9135; Angew. Chem. Int. Ed. 2008, 47, 8995-8997) (J. F. Lutz, H. Schlaad, Polymer 2008, 49, 817-824) (A. Gress, A. Voelkel, H. Schlaad, Macromolecules 2007, 40, 7928-7933) (N. ten Brummelhuis, C. Diehl, H. Schlaad, Macromolecules 2008, 41, 9946-9947) (K. L. Killops, L. M. Campos, C. J. Hawker, J. Am. Chem. Soc. 2008, 130, 5062-5064) (J. W. Chan, B. Yu, C. E. Hoyle, A. B. Lowe, Chem. Commun. 2008, 4959-4961) (G. Moad, E. Rizzardo, S. H. Thang, Acc. Chem. Res. 2008, 41, 1133-1142) (C. Barner-Kowollik, M. Buback, B. Charleux, M. L. Coote, M. Drache, T. Fukuda, A. Goto, B. Klumperman, A. B. Lowe, J. B. McLeary, G. Moad, M. J. Monterio, R. D. Sanderson, M. P. Tonge, P. Vana, J. Polym. Sci. Part A 2006, 44, 5809-5831) (R. J. Pounder, M. J. Stanford, P. Brooks, S. P. Richards, A. P. Dove, Chem. Commun. 2008, 5158-5160) (M. J. Stanford, A. P. Dove, Macromolecules 2009, 42, 141-147) (M. Li, P. De, S. R. Gondi, B. S. Sumerlin, J. Polym. Sci. Part A 2008, 46, 5093-5100) (Z. J. Witczak, D. Lorchak, N. Nguyen, Carbohydr. Res. 2007, 342, 1929-1933) (D. Samaroo, M. Vinodu, X. Chen, C. M. Drain, J. Comb. Chem. 2007, 9, 998-1011) (X. Chen, D. A. Foster, C. M. Drain, Biochemistry 2004, 43, 10918-10929; c) D. Samaroo, C. E. Soll, L. J. Todaro, C. M. Drain, Org. Lett. 2006, 8, 4985-4988) (P. Battioni, O. Brigaud, H. Desvaux, D. Mansuy, T. G. Traylor, Tetrahedron Lett. 1991, 32, 2893-2896) (C. Ott, R. Hoogenboom, U. S. Schubert, Chem. Commun. 2008, 3516-3518) (V. Ladmiral, G. Mantovani, G. J. Clarkson, S. Cauet, J. L. Irwin, D. M. Haddleton, J. Am. Chem. Soc. 2006, 128, 4823-4830) (S. G. Spain, M. I. Gibson, N. R. Cameron, J. Polym. Sci. Part A 2007, 45, 2059-2072) (C. R. Becer, K. Babiuch, K. Pilz, S. Hornig, T. Heinze, M. Gottschaldt, U. S. Schubert, Macromolecules 2009, 42, 2387-2394) (H. L. Holmes, R. M. Husband, C. C. Lee, P. Kawulka, J. Am. Chem. Soc. 1948, 70, 141-142; b) M. Lautens, W. Klute, W. Tam, Chem. Rev. 1996, 96, 49-92) (K. C. Nicolaou, S. A. Snyder, T. Montagnon, G. Vassilikogiannakis, Angew. Chem. 2002, 114, 1742-1773) (E. J. Corey, Angew. Chem. 2002, 114, 1724-1741) (H. Durmaz, A. Dag, O. Altintas, T. Erdogan, G. Hizal, U. Tunca, Macromolecules 2007, 40, 191-198) (H. Durmaz, A. Dag, A. Hizal, G. Hizal, U. Tunca, J. Polym. Sci. Part A 2008, 46, 7091-7100) (A. Dag, H. Durmaz, E. Demir, G. Hizal, U. Tunca, J. Polym. Sci. Part A 2008, 46, 6969-6977) (B. Gacal, H. Akat, D. K. Balta, N. Arsu, Y. Yagci, Macromolecules 2008, 41, 2401-2405) (A. Dag, H. Durmaz, U. Tunca, G. Hizal, J. Polym. Sci. Part A 2009, 47, 178-187) (M. L. Blackman, M. Royzen, J. M. Fox, J. Am. Chem. Soc. 2008, 130, 13518-13519) (N. K. Devaraj, R. Weissleder, S. A. Hilderbrand, Bioconjugate Chem. 2008, 19, 2297-2299) (Song, Y. Wang, J. Qu, Q. Lin, J. Am. Chem. Soc. 2008, 130, 9654-9655) (W. Song, Y. Wang, J. Qu, M. M. Madden, Q. Lin, Angew. Chem. 2008, 120, 2874-2877; Angew. Chem. Int. Ed. 2008, 47, 2832-2835) (A. Dag, H. Durmaz, G. Hizal, U. Tunca, J. Polym. Sci. Part A 2008, 46, 302-313) (A. J. Inglis,

S. Sinnwell, T. P. Davis, C. Barner-Kowollik, M. H. Stenzel, Macromolecules 2008, 41, 4120-4126) (S. Sinnwell, A. J. Inglis, T. P. Davis, M. H. Stenzel, C. Barner-Kowollik, Chem. Commun. 2008, 2052-2054) (A. J. Inglis, S. Sinwell, M. H. Stenzel, C. Barner-Kowollik, Angew. Chem. 2009, 121, 2447-2450). All references cited above are incorporated herein by reference for disclosure of click chemistry handles suitable for installation on proteins according to inventive concepts and methods provided herein.

The elements and method steps described herein can be used in any combination whether explicitly described or not.

All combinations of method steps as used herein can be performed in any order, unless otherwise specified or clearly implied to the contrary by the context in which the referenced combination is made.

As used herein, the singular forms "a," "an," and "the" include plural referents unless the content clearly dictates otherwise.

Numerical ranges as used herein are intended to include every number and subset of numbers contained within that range, whether specifically disclosed or not. Further, these numerical ranges should be construed as providing support for a claim directed to any number or subset of numbers in that range. For example, a disclosure of from 1 to 10 should be construed as supporting a range of from 2 to 8, from 3 to 7, from 5 to 6, from 1 to 9, from 3.6 to 4.6, from 3.5 to 9.9, and so forth.

All patents, patent publications, and peer-reviewed publications (*i.e.*, "references") cited herein are expressly incorporated by reference to the same extent as if each individual reference were specifically and individually indicated as being incorporated by reference. In case of conflict between the present disclosure and the incorporated references, the present disclosure controls.

It is understood that the invention is not confined to the particular construction and arrangement of parts herein illustrated and described, but embraces such modified forms thereof as come within the scope of the claims.

EXAMPLES

ENRICHABLE PROTEIN FOOTPRINTING FOR WHOLE PROTEOME 3D STRUCTURAL ANALYSIS

Summary

Protein footprinting via covalent labeling is a useful method for studying conformational changes induced by chemical and/or genetic perturbations. Compared to other methods that provide single amino acid-level structural resolution, such as cryo-EM, X-ray diffraction, and NMR, mass spectrometry (MS)-based methods can be advantageous, as they require lower protein amounts and purity. As with other MS-based proteomic methods, such as post-translational modification analysis, enrichment techniques have proven necessary for both optimal sensitivity and sequence coverage when analyzing highly complex proteomes. Currently used reagents for footprinting via covalent labeling, such as hydroxyl radicals and carbodiimide-based methods, do not yet have a suitable enrichable method, limiting their applicability to whole proteome analysis. Here, we report a method for enrichable covalent labeling to covalently label aspartic acid and glutamic acid residues. Novel labeling reagents containing alkynyl functionality can be 'clicked' to any azido-containing molecule with copper-catalyzed azide-alkyne cycloaddition (CuAAC), allowing for enrichment or further labeling. Multiple azide- and alkyne-containing GEE-like molecules were tested, and the most efficient method was determined to be EDC-facilitated coupling of glycine propargyl amide (GPA) to proteins. The pipeline we report includes clicking via CuAAC to a commercially available

biotin-azide containing a photocleavable linker followed by enrichment using a streptavidin resin and subsequent cleavage under ultraviolet light. The enrichment process was optimized through the screening of clickable amines, coupling reagents, and enrichment scaffolds and methods with pure model proteins, and has also been applied to complex mixtures of proteins isolated from the model plant, *Arabidopsis thaliana*, suggesting that our method may ultimately be used to measure protein conformation on a proteomic scale.

Introduction

Covalent labeling, also called "footprinting," of proteins has proven to be a valuable method for determining structural changes in pure proteins (Liu et al. 2020). These techniques include hydroxyl radical footprinting (HRF) (McKenzie-Coe et al. 2022) methods, such as fast photochemical oxidation of proteins (FPOP) (Johnson et al. 2019, Kiselar et al. 2002), stability of proteins from rates of oxidation (SPROX) (West et al. 2008), and plasma induced modification of biomolecules (PLIMB) (Minkoff et al. 2017), and chemical labeling methods, such as glycine ethyl ester (GEE) (Sanderson et al. 1996) and diethylpyrocarbonate (DEPC) (Limpikirati et al. 2018) labeling. By mapping amino acid residues within a protein that react with a labeling reagent, it is possible to infer the relative degree of solvent accessible surface area (SASA) of protein regions and to monitor how these surfaces change with perturbation. Covalent labeling methods target reactive amino acid side chains within a protein including carboxylic acid, amine, and thiol functional groups. These mass spectrometry (MS)-based methods have advantages over methods such as NMR, cryo-EM, and deuterium isotope incorporation in requiring lower amounts of protein and sample purity.

Carbodiimide chemistry for coupling carboxylic acids to amines is a well-established reaction in both organic and peptide synthesis (Valeur et al. 2009). The carbodiimide coupling method was first successfully applied to proteins in 1966 to quantify aspartic (Asp) and glutamic acid (Glu) residues in lysozyme, trypsin, and chymotrypsin using glycine methyl ester and N-benzyl-N-3'-dimethylamino propylcarbodiimide (BDC) (Hoare et al. 1966). It was later applied to footprinting with glycinamide (GA) (Akashi et al. 1993), glycine ethyl ester (GEE) (Zhang et al. 2012), and heavy isotope-containing GEE (Zhang et al. 2016) as the reactive amine labeling compounds and 1-ethyl-3-(3'-dimethylaminopropyl)carbodiimide (EDC) as coupling reagent. GEE labeling has been used to map the SASA of a membrane bound receptor-like protein kinase from the model plant *Arabidopsis thaliana* and monitor changes in labeling caused by the addition of its peptide hormone ligand (Haruta et al. 2014).

GEE footprinting has only been applied to pure proteins in solution and not to vastly more complex mixtures of proteins such as cell lysates due largely to fact that labeled peptides are more difficult to analyze by MS when they are in a mixture of far more abundant unlabeled peptides. One way to overcome this problem is to enrich for labeled peptides, pulling them out of the more complex mixture for easier MS analysis. To address this technical gap in the field, we reasoned that a chemically reactive clickable "handle" could be added to the covalent label to enable modification via click chemistry and enrichment with standard methods. Click reactions are defined by high yields under a wide range of conditions. First described in 2002 by Sharpless (Rostovstev et al. 2002) and Meldal (Tornøe et al. 2002), copper-catalyzed azide-alkyne cycloaddition (CuAAC) was selected as an enrichment method due to its biological orthogonality, short reaction time, tolerance to buffered aqueous solutions, and high yields. The chemical "handle" would thus be an alkynyl or azido functional group.

GEE footprinting is further complicated by the ethyl ester moiety's lability under standard conditions of the proteomic pipeline (FIG. 4). The GEE-labeled and hydrolyzed-GEE-labeled peptides resolve chromatographically and present multiple versions of a labelled peptide to quantify, complicating quantitative analysis. While this issue may be acceptable for pure, single protein experiments, extending the analysis to complex protein samples would be facilitated by mitigating GEE hydrolysis. For example, glycine propargyl amide (GPA) described in this work does not hydrolyze under normal conditions (FIG. 4).

To find reaction conditions for labeling exposed protein carboxylic acids with an enrichable amine, we screened various compounds containing either alkyne or azide functionality and structures that impart an array of solubility, hydrophobicity, and steric properties. Similarly, a set of carboxylic acid activating reagents consisting of five carbodiimides and Woodward's reagent K was screened. Five commercially available copper click-based enrichment methods were tested with alkyne-labeled peptides. Overall, we report that GPA is a promising covalent label for protein footprinting and ultimately proteome-wide conformational analysis.

Experimental

Materials

All chemical compounds used in this study are commercially available. 2-amino-N-(prop-2-yn-1-yl)acetamide, propargyl amine, but-3-yn-1-amine, pent-4-yn-1-amine, O-(prop-2-yn-1-yl)hydroxylamine, 2-(2-propynyloxy)ethylamine, 2-[2-(2-propynyloxy)ethoxy]ethylamine, 4-(2-propynyloxy)aniline, and 3-ethynylaniline, and Glycine propargyl ester (2-amino-N-(prop-2-yn-1-yl)acetamide hydrochloride, GPA), and O-[2-(prop-2-yn-1-yloxy)ethyl]hydroxylamine were purchased from Ambeed. 2-(4-((Bis((1-(tert-butyl)-1H-1,2,3-triazol-4-yl)methyl)amino)methyl)-1H-1,2,3-triazol-1-yl)acetic acid, 1-[4-[[22-[(3aR,4R,6aS)-hexahydro-2-oxo-1H-thieno[3,4-d]imidazol-4-yl]-4,18-dioxo-8,11,14-trioxa-5,17-diazadocos-1-yl]oxy]-5-methoxy-2-nitrophenyl]ethyl N-(3-azidopropyl)carbamate (BTTAA), AzDye 680 Azide Plus, PC Azide Agarose, PC-biotin azide, DADPS Biotin Azide, Agarose-GLAGLLAR-Azide (SKU CCT-1581-2, Vector Laboratories, Inc., Newark, CA), and Dde Azide Agarose were purchased from Vector Laboratories Inc. (Newark, CA). 2-amino-N-ethylacetamide hydrochloride (Glycine ethyl amide, GEA) was obtained from Matrix Scientific. Glycine ethyl ester (GEE) hydrochloride, bovine serum albumin – lyophilized powder, copper sulfate pentahydrate, and sodium ascorbate were purchased from Sigma-Aldrich. Trypsin/LysC mix was purchased from Promega. Lysyl endopeptidase was purchased from Wako. LC/MS grade formic acid and acetonitrile were Fisher Optima®. Streptavidin Sepharose High Performance resin was purchased from Cytiva. All water was Milli-Q purified. Photocleavage reactions were conducted in a Stratagene UV Stratalinker 1800 UV microwave.

Preparation of Arabidopsis thaliana Cytosol and Microsomes

Arabidopsis seeds (12 mg) were surface-sterilized with a sterilizing solution containing 75% (v/v) ethanol, 0.1% (v/v) Triton X-100, and 2% bleach (v/v) for 5 min and rinsed with 180-proof ethanol twice. Seedlings were grown in a magenta box containing 75 ml 1/2x Murashige and Skoog (1/2x MS) media plus 1% sucrose and 75 µl 1000x (100 mg/ml) ampicillin. Seeds were stratified in the dark for 48 hours at 4°C and then plants were grown on a shaker (~30 rpm) under continuous light at room temperature. After 10 days, plant tissue was removed, blotted to remove excess media, and weighed.

Tissue was ground in 2x volume homogenization buffer (HB) with inhibitors (230 mM sorbitol, 50 mM Tris-HCl pH 7.5, 10 mM KCl, 3 mM EGTA, 0.5 µg/ml Leupeptin, 0.7 µg pepstatin, Sigma EDTA-free protease inhibitor tablet, 1 mM potassium metabisulfite, 1 mM phenylmethylsulfonyl fluoride (PMSF)). Using Pro250 homogenizer, tissue was ground on ice for 60-90 seconds, at 11,000 rpm and then filtered through four layers of Miracloth. The homogenate was centrifuged for 10 min at 6000 x g, at 4°C. The supernatant was removed and then centrifuged again for 45 min at 65,000 x g at 4°C. The supernatant was then decanted. The microsomal pellet was resuspended in resuspension buffer (RB) (330 mM sorbitol, 5 mM KCl, 5 mM K_2HPO_4 pH 7.8) and homogenized in a Teflon potter before being aliquoted and frozen.

General Covalent Labeling Procedure

Footprinting buffer (FB) was prepared just prior to use (10 mM KH_2PO_4, 50 mM KCl, and 12 mM MgCl2). Reactions were carried out in 1.5 ml low-binding tubes. Stock solutions of the carbodiimide (1 M) and the amine (3 M) were prepared in FB. A final reaction volume of 200 µl was used for 50-100 µg of protein. Amine stock was added to the protein solution to a final concentration of 450 mM, followed by the carbodiimide stock solution to a final concentration of 60 mM, the reaction vortexed for 5 s and then allowed to sit at room temperature for 4 minutes before being quenched with 2 M NH_4OAc to a final concentration of 400 mM.

Cold trichloroacetic acid (TCA) was then added to 10% (v/v) and the reaction mixture placed on ice for 15 min to precipitate protein. The resulting suspension was centrifuged at 16,000 x g for 10 minutes at 4°C to pellet the protein. The supernatant was discarded, and 1 ml of cold acetone was added. At this point a pipet tip was used to gently break up the protein pellet and the suspension was centrifuged for 10 min at 16,000 x g at 4°C. The supernatant was removed, and the pellet air-dried.

General CuAAC Procedure

The protein pellet was resolubilized in minimal 8 M urea / 25 mM NH_4HCO_3 with vortexing or probe sonication if needed. The solution was then diluted to 4 M urea with 25 mM NH_4HCO_3. The appropriate azide was added at a final concentration of 100 mM, followed by sodium ascorbate to a final concentration of 5 mM from a 50 mM stock solution. The solution was briefly vortexed to mix. A pre-mixed copper/ligand stock solution (10 mM $CuSO_4$ and 50 mM BTTAA) was then added to a final concentration of 1 mM copper and 5 mM ligand. The reaction was gently mixed by pipetting up and down to avoid the incorporation of air. The reaction was kept at room temperature for 1 hr with gentle end-over-end mixing. At the end of 1 hr, the protein was precipitated with TCA as described above for the covalent labeling step.

General Procedure for Reduction, Alkylation, and Digestion of Proteins

The protein was resolubilized in minimal 8 M urea / 25 mM NH_4HCO_3 with probe sonication and then diluted to 4 M urea with 25 mM NH_4HCO_3. Dithiothreitol (DTT) was added to a final concentration of 2 mM, the reaction incubated at 42°C for 20 min and then cooled to room temperature. For procedures using a hydrolyzable amine (GEE or GPE) a 500 mM solution of iodoacetic acid in 0.5 M NaOH was added to a final concentration of 20 mM. For all other, nonhydrolyzable amines iodoacetamide (IAA) was added to a final concentration of 20 mM. This alkylation reaction was allowed to sit at room temperature, in the dark for 30 min and then quenched by adding DTT to a final concentration of 2 mM. The solution was diluted to 1 M urea with 25 mM NH_4HCO_3 and the protease trypsin/LysC mix added at a ratio of 50:1 protein to protease. The digestion was incubated for 12 hr at 37°C.

Streptavidin Enrichment of PC-Biotin Clicked Peptides

The tryptic digestion was inhibited with phenylmethylsulfonyl fluoride (PMSF) to a final concentration of 1 mM and put on ice. Streptavidin Sepharose was added to a new 1.5 ml tube using a wide-bore pipet tip and washed with 1 ml of 50 mM NH_4HCO_3, followed by centrifuging at 1000 x g for 2 min. The supernatant was removed, and wash steps repeated. The resin was then washed twice with 1 M urea / 50 mM NH_4HCO_3, centrifuged at 1000 x g for 2 min and with the supernatant being discarded in between steps. The peptide solution was then added to the streptavidin Sepharose, and the mixture incubated for 2 hr at 4°C with gentle end-over-end rotation. The resin was pelleted by centrifuging at 1000 x g for 3 min and the supernatant removed. The resin was washed twice with 1 ml of 1 M urea / 50 mM NH_4HCO_3 followed by centrifuging at 1000 x g for 2 min and the removal of the supernatant. A second wash was carried out twice with 1 ml of 50 mM NH_4HCO_3 followed by centrifuging at 1000 x g for 2 min, and removal of the supernatant.

Peptide Elution from PC-Biotin-Streptavidin

The pelleted resin was suspended in 150 µl of 50 mM NH_4HCO_3 and transferred to a thin-walled PCR tube or glass vial to allow for efficient irradiation with minimal reflection or absorption of light. The resin was irradiated with 365 nm light for 1 hr at 4°C. The suspension was then transferred back to a clean 1.5 ml low-binding tube, centrifuged for 3 min at 1000 x g and the eluent collected. The resin was washed with 100 µl of 50 mM NH_4HCO_3 and this wash combined with the previous eluent.

General Peptide Purification

Peptide samples were acidified with formic acid to a final 1% (v/v) concentration and then purified with Omix C18 stage tips according to the manufacturer’s protocol. The eluted peptide solution was dried in a vacuum centrifuge and resolubilized in 1% formic acid for MS analysis.

General Mass Spectrometry Methods

Samples were injected onto a 75 µm x 50 cm Thermo Fisher Scientific C18 Easy Spray Column with 2 µm particles and 100 Å pore size. Mobile phases used for chromatographic separation were LC/MS-grade 0.1% formic acid in water (A) and LC/MS grade 0.1% formic acid in 80% acetonitrile (B). Peptides were separated using a gradient from 2% to 37.5% B over 25 minutes (BSA samples) or 33 minutes (complex samples), after which the column was flushed with 95% B for 5 minutes and re-equilibrated to 2% A for 10 minutes.

Peptides eluting from the column were sprayed at 1900 V into a Thermo Scientific Orbitrap Fusion Lumos Tribrid Mass Spectrometer. Data-dependent MS acquisition parameters were as follows: MS1 spectra were acquired in the Orbitrap in profile mode with a resolution of 120 K, quadrupole isolation activated, a scan range of 375-1800 m/z, a radio frequency (RF) lens % of 30, normalized automatic gain control (AGC) target of 250%, max injection time of 50 ms. For selecting ions for fragmentation and MS2 acquisition, monoisotopic peak selection was set to peptide, charge states other than 2-7 were rejected, and dynamic exclusion was set to n=1, a duration of 10 s, and a mass tolerance of +/-10 ppm. MS spectra were acquired in the ion trap using higher-energy collisional dissociation (HCD) fragmentation and a fixed collision energy of 32%. Quadrupole isolation was used with the isolation window set to 0.7 m/z. A scan rate of turbo, mass range of normal, and scan range mode of auto were selected for MS acquisition, and an AGC target of 200% and max inject time of 50 ms was used. MS2 spectra were acquired in centroid mode. A cycle time of 1 s between MS1 spectra was used for data dependent acquisition.

For downstream FragPipe analysis, the same chromatographic, electrospray, and MS1 acquisition parameters were used. MS2 were acquired in the Orbitrap with the following parameters: Quadrupole isolation was used with the isolation window set to 0.7 m/z, scan range mode was set to auto, an AGC target of 300% and max inject time of 54 ms was used, and HCD fragmentation with a fixed collision energy of 30% was used.

Proteomic Data Analysis

Data were searched using the Sequest node of Proteome Discoverer v2.4. Depending on the experiment either the BSA sequence (SEQ ID NO:1) or the *Arabidopsis thaliana* Uniprot data-base with added contaminant proteins was searched, specifying tryptic cleavage with up to 2 missed cleavages and a precursor mass tolerance of 10 ppm, fragment mass tolerance of 0.6 Da. Variable oxidation (M), deamidation (N/Q), phosphorylation (S/T/Y), were allowed. Dynamic modification for labeling or the appropriate clicked label were searched (D/E). Carboxymethylation (C) for GEE and GEA and carbamidomethylation (C) for all other non-hydrolyzable amines was set as static. The percolator node was used to filter resulting data with a false discovery rate (FDR) of 0.05.

MSFragger version 4.0 in FragPipe version 20.0 was used to conduct open searches of MS data with precursor mass tolerance set to -150 to 500, fragment mass tolerance of 20, enzymatic cleavage set to strict trypsin, BSA and *Arabidopsis thaliana* databases were obtained from UniProt with decoys and contaminants added, peptide length range of 7 to 50, 2 missed cleavages, peptide mass range of 500 to 5000, FDR of 1%, allowing for variable modifications of methionine oxidation (15.9949), and fixed modification of cysteine carboxyamidomethylation (57.02146). Open searches were used to determine the number of labeled PSMs and the percent labeled of all PSMs.

Calculation of MS Abundance of Individual Labeled Protein Sites

Numbered protein sequences were used to assign a position number to each labeled site of all labeled peptides in Excel. The MS abundance values for each peptide containing an individual labeled site were summed. This value was then normalized between 0 and 1 using the formula $a = [x - \min(x)] / [\max(x) - \min(x)]$ where “a” is the normalized abundance value, “x” is an abundance value in the dataset, “min(x)” is the minimum value in the dataset and “max(x)” is the maximum value in the dataset.

Fluorescent Labeling of Proteins

The GPA labeling and protein precipitation steps were carried out as described above in the general procedures. The resulting protein pellet was resolubilized in minimal PBS using probe sonication if necessary. Then the fluorescent AzDye-azide was added to a final concentration of 10 mM followed by sodium ascorbate to a final concentration of 5 mM. The solution was briefly vortexed to mix. Then 20 µl pre-mixed copper/ligand solution was added to a final concentration of 1 mM copper and 5 mM ligand. The reaction mixture was gently mixed by pipetting up and down to avoid the incorporation of air which is detrimental to the CuAAC reaction. The reaction was kept at room temperature for 1 hour in the dark with gentle end-over-end mixing.

Gel Electrophoresis

Fluorescent labeled samples for SDS-PAGE analysis were run using a Novex Bolt mini gel tank. Gels were imaged using a Li-Cor Odyssey scanner and Li-Cor Image Studio v3.1, and Coomassie gels were imaged using an Epson V850 Pro scanner and Silverfast 8 software.

Streptavidin Blotting and Membrane Imaging

Gels run as above were blotted according to manufacturer’s protocol. Briefly, they were incubated in Bolt transfer buffer for 20 minutes and transferred to blotting membranes for 10 minutes with the Bolt transfer system. They were then blocked for 1 hour with Licor PBS blocking buffer and incubated overnight in blocking buffer with 0.2% TWEEN-20, 0.1% SDS, and 1uL of LiCor’s fluorescently labeled streptavidin stock. After washing 3x for 5 min with 0.2% TWEEN-20 in PBS, and 1x in PBS, membrane was dried and imaged with a Li-Cor Odyssey scanner and Li-Cor Image Studio v3.1.

GPA Labeling and CuAAC Reaction with DADPS Biotin Azide

The GPA labeling and protein precipitation were carried out as described above. The protein pellet was resolubilized in minimal PBS and DADPS biotin azide was added to a final concentration of 100 mM, followed by sodium ascorbate to a final concentration of 5 mM. The solution was briefly vortexed to mix. Then pre-mixed copper/ligand solution was added to a final concentration of 1 mM copper and 5 mM ligand. The reaction mixture was gently mixed by pipetting up and down to avoid the incorporation of air. The reaction was kept at room temperature for 2 hours in the dark with gentle end-over-end mixing. The protein was then precipitated by one of the methods described above (TCA or methanol/chloroform).

Enrichment of Labeled DADPS-Biotin Peptides

Streptavidin resin (100 µl) was added to a 1.5 ml tube and washed twice with 1 ml 100 mM NH_4HCO_3, centrifuging and removing the wash solution after each wash. The peptide mixture was then transferred to the tube containing the streptavidin resin. The mixture was incubated for 2 hours at room temperature with end-over-end rotation. The resin was then pelleted by centrifuging at 1000 x g for 3 min. The supernatant was removed, and the resin was washed with 1 ml PBS, centrifuged at 1000 x g for 2 min, and the wash solution removed. This wash was repeated. The resin was then washed twice with 1 ml PBS / 2 M NaCl and then twice with water, spinning at 1000 x g for 2 min and discarding the wash solution between each cycle.

Elution of DADPS-Biotin Peptides from Streptavidin Resin

The beads were resuspended in 800 µl 2% formic acid and incubated at room temperature for 2 hr with end-over-end rotation. The resin was then pelleted by centrifuging at 1000 x g for 4 min and the eluent moved to a clean tube.

GPA Labeling and CuAAC Reaction with DDE Agarose Azide

The GPA labeling and protein precipitation were carried out as described above. The DDE resin (200 µl) was added to a 1.5 ml tube, washed with 1 ml 8 M urea / 25 mM NH_4HCO_3 and centrifuged at 1000 x g for 2 minutes. The precipitated protein was re-solubilized in a minimal amount of 4% CHAPS / 1 M NaCl / 8 M urea with vortexing or probe sonication and added to the tube containing the DDE resin. Sodium ascorbate was added to a final concentration of 5 mM and the mixture briefly vortexed. Then pre-mixed copper/BTTAA stock solution was added to a final concentration of 1 mM copper and 5 mM ligand. The resulting reaction mixture was capped to exclude air and left for 20 hours in the dark at room temperature with gentle shaking. The click reaction mixture was centrifuged at 1000 x g for 3 minutes and the supernatant was removed. The resin was then washed with 1 ml water, centrifuged for 2 minutes at 1000 x g, and the supernatant discarded. A second wash was done with 1 ml SDS agarose wash buffer (100 mM Tris, 1% SDS, 250 mM NaCl, 5 mM EDTA, pH 8.0), centrifuged at 1000 x g for 2 minutes to pellet and then the supernatant was removed. The resin was then washed with 1 ml 8 M urea / 25 mM NH_4HCO_3, centrifuged at 1000 x g for 2 minutes and the solvent removed. This wash step was repeated for a total of five washes. The resin was then washed with 1 ml 20% acetonitrile, centrifuged at 1000 x g for 2 minutes, and the solvent discarded. This wash step was repeated four more times for a total of five washes.

On-Bead Reduction, Alkylation and Digestion with DDE Agarose

The resin was resuspended in 50 µl of 8 M urea / 25 mM NH_4HCO_3 and then diluted to 4 M urea with 25 mM NH_4HCO_3. Dithiothreitol (DTT) was added to a final concentration of 2 mM and the reduction mixture was incubated at 42°C for 20 minutes. The tube was then placed on ice briefly to cool to room temperature and iodoacetamide (IAA) added to a final concentration of 20 mM. The alkylation reaction was allowed to sit at room temperature, in the dark for 30 min and then quenched by the addition of 1.5 µl DTT. This was diluted to 1 M urea by adding 300 µl of 25 mM NH_4HCO_3. Trypsin/LysC mixture was then added at a ratio of 50:1 protein to protease, and the digestion was incubated 12 hr at 37°C with shaking. The digestion mixture was then centrifuged at 1000 x g for 3 minutes and the supernatant removed. The resin was washed with 1 ml 1 M urea / 25 mM NH_4HCO_3 and centrifuged at 1000 x g for 2 min.

Peptide Elution from DDE-Agarose

The resin was resuspended in 200 µl DDE elution buffer (2% (v/v) hydrazine in 100 mM sodium phosphate solution) and incubated at room temperature with gentle shaking for 1 hr. The reaction mixture was then centrifuged at 1000 x g for 3 minutes to pellet the resin, and the eluent was collected and transferred to a new tube. The resin was then washed 2 times with 100 µl PBS with 2 min centrifugation at 1000 x g and washes combined with original eluent. The combined eluent was then dried with vacuum centrifugation and the peptides purified and analyzed by MS as described above in the general procedures.

GPA Labeling and CuAAC Reaction with Agarose-GLAGLLAR-Azide

The GPA labeling and TCA protein precipitation steps were carried out as described above. The protein was resolubilized in 4% CHAPS / 1 M NaCl / 8 M urea. The tryptic agarose resin (300 µl) was washed with 1 ml water, and centrifuged at 1000 x g for 2 min, and the solvent discarded. The protein solution was then added to the resin. Sodium ascorbate was added to a final concentration of 5 mM and the mixture briefly vortexed. Then pre-mixed copper/ligand solution was added to a final concentration of 1 mM copper and 5 mM ligand. The resulting reaction mixture was capped to exclude air and left for 20 hours in the dark at room temperature with gentle shaking.

Reduction, Alkylation, and LysC Digestion of Agarose-GLAGLLAR

The click reaction mixture was centrifuged at 1000 x g for 3 min to pellet the resin and the supernatant removed. The resin was washed with 1 ml water, centrifuged at 1000 x g for 2 min and the wash discarded. The resin was resuspended in 1 ml SDS agarose wash buffer and DTT added to a final concentration of 2 mM. The bead-protein mixture was reduced at 42°C for 30 minutes and then centrifuged at 1000 x g for 2 min and the supernatant discarded. The protein was alkylated by adding 1 ml of 2 mM iodoacetamide in SDS agarose wash buffer (100 mM Tris, 1% SDS, 250 mM NaCl, 5 mM EDTA, pH 8.0) and incubating at room temperature, in the dark for 30 minutes. After centrifuging for 2 minutes at 1000 x g and removing the supernatant, the resin was washed with 1 ml water, centrifuged again for 2 minutes at 1000 x g and the wash discarded. The resin was then washed a total of 5 times with 1 ml SDS agarose wash buffer, spinning at 1000 x g and discarding the buffer between each washing step. The resin was then washed a total of 3 times with 1 ml 8 M urea / 25 mM NH_4HCO_3, centrifuging for 2 min at 1000 x g and discarding the supernatant between each wash. The resin was then washed with 1 ml 20% MeCN, centrifuged for 2 min at 1000 x g and the wash discarded. The resin was washed with 1 ml digestion buffer (100 mM Tris pH 8.0, 2 mM CaCl2, 10% MeCN), centrifuged for 2 min at 1000 x g, supernatant removed, and this wash step repeated for a total of 2 washes. The resin was then suspended in 400 µl digestion buffer and lysyl endopeptidase added at a 50:1 protein: protease ratio. The digest was incubated at 37°C for 20 hr with shaking.

GLAGLLAR-Agarose Resin Wash

The digest was centrifuged for 3 min at 1000 x g and the supernatant removed. The resin was washed with 1 ml water, centrifuged 2 min at 1000 x g and the wash discarded. The resin was then washed six times with 1 ml SDS agarose wash buffer, centrifuging to pellet at 1000 x g for 2 min and discarding the wash solution between each step. The same process was then repeated with 8 M urea / 25 mM NH_4HCO_3 and then 20% MeCN for 6 wash steps each.

Trypsin Elution of GLAGLLAR-Agarose Bound Peptides

The resin was resuspended in 200 µl 1 M urea / 25 mM NH_4HCO_3 and trypsin added at a 50:1 protein: protease ratio. The digestion was incubated at 45°C for 4 hours with shaking and then centrifuged at 1000 x g for 5 minutes. The eluent is then moved to a new tube and the beads washed twice with 100 µl water and the washes added to the eluent.

GPA-Labeling of Microsomes. To 198 µl of a microsome sample prepared as described above was added 0.5 M GPA in 10 mM K2HPO4 to a final concentration of 45 mM, followed by 0.1 M EDC in 10 mM K2HPO4 to a final concentration of 6 mM. The reaction mixture was left on a shaker for 3 minutes at room temperature. A 100 µl aliquot was removed and added to 100 µl of 10% SDS. Samples were then cleaned and trypsin-digested using S-Trap™ spin columns (Protifi) according to the manufacturer's procedure. The resulting peptide samples were either enriched using the PC-biotin-azide method described above or acidified with formic acid to a final 1% (v/v) concentration and then purified with Omix C18 stage tips according to the manufacturer's protocol. The eluted peptide solutions were dried in a vacuum centrifuge and resolubilized in 1% formic acid for MS analysis, which was performed as described above.

Results and Discussion

We explored the development of a covalent labeling method that enables proteome-wide labeling, followed by enrichment and subsequent MS analysis of those labeled sites. We reasoned that by screening azide- or alkyne-containing amines and coupling compounds we could identify a covalent labeling pair that would augment standard labeling methods with enrichability. Thus, we chose a set of readily available amines with either an alkyne or azide functionality.

Glycine Propargyl Amide is a Clickable Labeling Agent and Labels the Model Protein BSA

Initially, several amine compounds containing either an alkyne or azide functionality were screened in an EDC coupling reaction with the pure protein bovine serum albumin (BSA) (SEQ ID NO:1). EDC was utilized as a zero-length crosslinking reagent (Grabarek et al. 1990). For this reason, this pathway linking lysine residues to glutamic acid and aspartic acid is minimized by adding the amine first and in excess when used as a reagent in covalent footprinting reactions (Mintseris et al. 2020).

Glycine propargyl ester (GPE, FIG. 5, b) and glycine propargyl amide (GPA, FIG. 5, c) were chosen as alternatives to GEE (FIG. 5, a), with the hypothesis that they would have similar solubility and steric properties. Three aliphatic alkynyl amines (propargyl amine, (FIG. 5, d); but-3-yn-1-amine, (FIG. 5, e); pent-4-yn-1-amine, (FIG. 5, f) were screened. The hydroxylamines, O-(prop-2-yn-1-yl)hydroxylamine (FIG. 5, g), O-[2-(prop-2-yn-1-yloxy)ethyl]hydroxylamine (FIG. 5, h) and the ethylene glycol containing compounds 2-(2-propynyloxy)ethylamine (FIG. 5, i) 2-[2-(2-propynyloxy)ethoxy]ethylamine (FIG. 5, j) were selected for potentially improved solubility in aqueous buffers. Two representative aromatic amines, 3-ethynylaniline (FIG. 5, k), 4-(2-propynyloxy)aniline (FIG. 5, l) were also selected. In addition to the alkyne-containing amines above, three azide-containing amines were also screened: the aliphatic (3-azidopropylamine, FIG. 5, m), PEG-containing 11-azido-3,6,9-trioxaundecan-1-amine (FIG. 5, n), and aromatic compound 4-azidoaniline (FIG. 5, o).

As a quick assessment of labeling itself, GPA-labeled and PC-biotin-azide-derived cytosol was blotted with fluorescent streptavidin, which confirmed labeling (FIG. 10).

For a basic comparison of the extent of labeling with the selected amines an open search in FragPipe (Yu et al. 2020) was used to generate the number of peptide spectral matches (PSMs) for peptides containing the corresponding label as well as the percent of total PSMs of labeled peptides (Table 2).

Table 2. Extent of BSA labeling in the EDC-mediated reaction with various clickable amine compounds and the pure protein BSA.

Compound	Amine	Mass Shift	PSMs	% PSMs
a	glycine ethyl ester (GEE)	85.0528	235	5.852
	hydrolyzed GEE	57.0215	86	2.142
b	glycine propargyl ester (GPE)	95.0376	41	1.08
	hydrolyzed GPE	57.0215	346	9.13
c	glycine propargyl amide (GPA)	94.0536	1459	14.15
d	propargylamine	37.0316	235	4.58
e	but-3-yn-1-amine	51.0472	390	8.09
f	pent-4-yn-1-amine	65.0628	253	4.98
g	O-(prop-2-yn-1-yl)hydroxylamine	53.0268	1347	30.30
h	O-[2-(prop-2-yn-1-yloxy)ethyl]hydroxylamine	97.0527	594	9.19
i	2-(2-propynyloxy)ethylamine	81.0580	105	2.28
j	2-[2-(2-propynyloxy)ethoxy]ethylamine	125.0840	74	1.52
k	3-ethynylaniline	99.0478	731	22.72
l	4-(2-propynyloxy)aniline	-	0	0.00
m	3-azidopropylamine	82.0644	158	2.57
n	11-azido-3,6,9-trioxaundecan-1-amine	200.1272	108	1.70
o	4-azidoaniline	116.0486	753	12.98

GPE, which is most structurally like GEE, reacted well with BSA, but, like GEE, the ester is readily hydrolysable, leading to the loss of the click functionality. The hydrolysis product of GPE was more prevalent than the non-hydrolyzed (~9% of all PSMs vs. 1%). Glycine propargyl amide (GPA, FIG. 5, c) was likewise chosen for its structural similarity. The stable amide bond prevents hydrolysis byproducts under normal conditions (FIG. 4). GPA proved to have higher reactivity than GEE in the labeling of BSA-several sites were labeled to a greater extent with GPA and 6 additional sites were labeled with GPA, but not in the GEE reaction (Table 3).

Table 3. Comparison of GEE and GPA labeling of BSA.*

Residue	Amino Acid	%GPA label	% GEE label	% HydroGEE
30	Glu	100.00	0.00	0.00
37	Asp	0.28	0.00	0.30

40	Glu	3.97	0.77	2.07
41	Glu	8.63	2.55	0.97
61	Asp	2.87	0.42	1.91
62	Glu	8.24	0.47	1.45
69	Glu	4.52	0.10	2.43
72	Glu	12.03	5.44	0.00
80	Asp	2.05	0.73	0.00
81	Glu	2.39	0.68	0.00
87	Glu	3.37	0.72	0.01
96	Asp	0.44	0.02	0.01
97	Glu	0.70	0.11	0.93
106	Glu	52.16	54.09	0.00
110	Asp	13.39	100.00	4.62
113	Asp	37.39	31.02	1.35
116	Glu	66.87	81.07	0.00
119	Glu	20.83	0.00	0.00
121	Glu	1.83	0.00	0.00
124	Glu	48.87	45.74	10.97
131	Asp	39.80	39.84	4.94
132	Asp	41.74	0.52	8.11
135	Asp	55.52	4.85	0.30
142	Asp	4.21	0.42	1.21
148	Asp	19.84	3.77	1.35
149	Glu	3.36	3.72	0.97
153	Asp	0.14	2.41	3.00
154	Glu	10.60	6.39	0.04
176	Glu	1.07	0.01	1.04
190	Glu	28.69	4.39	1.85
195	Glu	21.64	29.88	1.74
196	Asp	24.99	24.60	10.30
250	Glu	35.15	16.64	2.60
253	Glu	37.01	1.02	0.00
267	Glu	0.53	0.00	0.00
272	Asp	2.08	0.03	0.04

275	Glu	0.84	0.78	0.25
278	Asp	2.76	5.10	0.89
279	Asp	6.57	10.34	1.22
282	Asp	4.58	14.26	0.41
289	Asp	16.07	2.42	2.35
292	Asp	35.43	17.09	0.47
300	Glu	6.67	1.50	1.54
303	Asp	9.22	3.57	0.00
308	Glu	3.25	0.07	0.51
315	Glu	4.35	2.70	0.42
317	Glu	0.26	3.65	0.27
319	Asp	3.32	4.95	2.36
323	Glu	13.73	17.26	10.03
331	Asp	41.68	18.21	9.79
334	Glu	25.48	14.69	7.21
335	Asp	41.89	12.87	3.28
337	Asp	13.37	7.05	0.00
344	Glu	2.78	16.19	0.56
347	Asp	4.87	0.31	0.27
356	Glu	5.19	0.03	0.42
363	Glu	13.06	0.14	1.44
375	Glu	4.74	1.69	0.63
377	Glu	3.79	2.48	2.83
381	Glu	24.58	22.05	2.43
382	Glu	8.79	28.78	11.70
387	Asp	24.16	23.01	5.40
388	Asp	18.90	10.58	0.23
398	Asp	4.02	0.15	0.52
405	Asp	4.29	0.43	0.43
406	Glu	10.11	0.36	0.71
416	Asp	10.82	1.02	2.42
419	Glu	28.89	24.18	3.75
423	Glu	18.78	0.15	0.02
465	Glu	1.25	0.00	0.06

467	Glu	1.01	0.02	0.22
473	Glu	20.19	0.59	0.52
474	Asp	11.91	0.08	3.75
488	Glu	37.05	0.00	1.38
494	Glu	8.83	1.92	0.00
502	Glu	0.91	1.87	1.68
517	Asp	13.29	0.55	2.29
518	Glu	17.92	1.83	0.00
527	Glu	0.00	0.00	0.35
535	Asp	5.50	1.95	0.28
541	Asp	12.65	0.35	0.30
543	Glu	17.87	1.47	0.02
554	Glu	3.51	0.04	0.00
564	Glu	42.80	0.00	0.00
565	Glu	57.58	0.00	3.94
572	Glu	9.24	4.72	1.21
579	Asp	13.91	4.99	4.29
585	Asp	15.21	16.16	3.30
586	Asp	36.29	16.05	2.11
588	Glu	22.05	6.70	0.06
594	Glu	2.46	0.05	28.84
607	Ala	77.92	71.16	28.84

* Percent labeling of all glutamic and aspartic acid residues of BSA with EDC as activator. Asp and Glu residues that were not covered by MS are excluded from this list.

Labeling of BSA with alkynyl amine compounds d-l is shown in Table 4. The three aliphatic compounds (d-f) did not label BSA to the same extent as GPA, possibly due to reduced solubility in the aqueous buffered reaction mixture or hydrophobic interactions with the protein. The hydroxyl amine compounds g and h performed well and labeled Glu and Asp to a greater degree than GEE. This increased reactivity could be due to the fact that the hydroxylamine functionality has increased nucleophilicity due to the alpha effect (Edwards et al. 1962) of the neighboring oxygen lone pair. A hydroxylamine compound has been used previously for footprinting calmodulin (Guo et al. 2019). The PEG-containing alkynyl amines (i and j) afforded the lowest level of labeling of all the alkyne-containing compounds tested in this study. Among the aromatic compounds 3-ethynylaniline (k) and 4-azidoaniline (o) (see Table 5 for labeling with (o)), both provided relatively high levels of labeling. The aniline compound (l) was not soluble in aqueous buffer and was first dissolved in DMSO and then added to the buffered BSA solution at which point it appeared to precipitate out of solution. Thus, it was not surprising that no labeling of BSA was observed with this compound. Overall GPA generally labeled the same Glu and Asp residues of BSA as GEE, suggesting that it is a reasonable agent for mapping in-solution protein structure (FIG. 6).

Table 4. Labeling of BSA with alkynyl amine compounds d-l.*

Residue	% d Label	% e Label	% f Label	% g Label	% h Label	% i Label	% j Label	% k Label	% l Label
40	2.26	0.00	0.18	27.81	15.03	0.00	0.00	0.00	0.00
41	6.90	0.00	16.14	0.30	2.68	3.23	0.63	0.63	0.00
61	0.00	10.65	1.04	39.67	0.01	0.10	0.00	0.00	0.00
62	0.00	0.00	2.92	23.20	1.00	0.00	0.00	0.00	0.00
69	0.27	0.00	0.10	1.23	0.02	0.00	0.00	0.00	0.00
72	0.53	0.00	0.07	3.87	0.01	0.52	0.27	0.27	0.00
80	0.12	0.46	0.21	43.97	1.94	0.16	0.00	0.00	0.00
81	0.32	1.74	0.00	38.49	0.00	0.00	0.12	0.12	0.00
87	0.00	1.23	0.00	100.00	0.00	0.00	0.00	0.00	0.00
96	0.01	0.00	0.11	17.63	0.08	0.00	0.00	0.00	0.00
97	0.05	0.34	0.15	1.35	0.08	0.05	0.00	0.00	0.00
106	0.00	0.00	94.48	0.00	0.00	0.00	0.00	0.00	0.00
110	11.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
113	0.00	0.00	0.00	38.96	0.00	0.00	0.00	0.00	0.00
116	15.95	0.00	44.82	27.68	0.00	0.00	0.00	0.00	0.00
119	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
121	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
124	0.85	0.00	18.76	45.84	69.46	0.00	0.00	0.00	0.00
131	2.88	0.00	0.00	1.01	0.00	0.00	0.00	0.00	0.00
132	0.00	0.00	0.00	100.00	11.11	0.00	0.00	0.00	0.00
135	1.05	0.00	34.59	79.09	21.75	0.00	0.00	0.00	0.00
142	0.13	0.31	0.00	61.61	0.84	0.00	0.00	0.00	0.00
148	0.86	8.32	3.50	49.91	8.31	0.52	0.00	0.00	0.00
149	0.08	2.50	1.90	100.00	0.00	0.00	0.00	0.00	0.00
154	0.21	0.00	1.20	0.56	0.00	0.25	0.00	0.00	0.00
176	0.06	0.40	0.01	21.63	0.53	0.26	0.14	0.14	0.00
190	0.28	3.67	0.74	67.48	4.57	0.13	0.13	0.13	0.00
195	1.71	0.00	6.07	29.59	5.75	0.43	0.34	0.34	0.00

196	0.00	0.00	2.31	100.00	0.00	0.25	0.00	0.00	0.00
250	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
253	0.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
267	1.54	2.11	1.03	5.95	0.00	0.82	0.00	0.00	0.00
272	0.11	0.00	0.00	17.06	1.35	0.00	0.00	0.00	0.00
275	0.12	0.27	0.00	100.00	10.50	0.00	0.13	0.13	0.00
278	0.17	0.00	0.00	28.74	0.24	0.00	0.00	0.00	0.00
279	1.22	0.00	0.66	16.32	0.07	1.03	0.07	0.07	0.00
282	0.26	5.75	0.76	11.79	0.29	0.00	0.00	0.00	0.00
289	0.13	19.36	2.00	15.10	2.54	0.12	0.00	0.00	0.00
292	0.57	31.98	12.08	57.16	11.14	0.38	0.00	0.00	0.00
300	6.09	7.98	45.29	0.00	0.00	0.00	0.00	0.00	0.00
303	0.09	8.35	45.29	6.26	1.67	0.00	0.00	0.00	0.00
308	0.00	0.00	0.00	28.21	0.00	0.00	0.00	0.00	0.00
315	1.57	0.00	0.00	8.94	0.04	0.00	0.00	0.00	0.00
317	0.00	9.32	19.06	3.08	0.80	0.00	0.31	0.31	0.00
319	0.05	2.20	1.23	19.53	0.54	2.55	0.05	0.05	0.00
323	0.13	0.00	6.05	45.01	5.70	0.17	0.00	0.00	0.00
331	4.00	2.20	13.94	81.03	27.88	1.39	0.21	0.21	0.00
334	0.13	0.00	1.88	39.78	5.12	0.00	0.24	0.24	0.00
335	0.15	11.04	1.47	53.77	2.87	0.00	0.78	0.78	0.00
337	1.18	0.78	0.00	8.70	0.00	0.00	0.00	0.00	0.00
344	0.00	0.00	0.00	42.51	0.00	0.00	0.00	0.00	0.00
347	1.82	6.63	0.93	1.92	0.36	1.04	0.86	0.86	0.00
356	0.42	0.23	0.36	0.65	0.01	0.12	0.00	0.00	0.00
363	0.17	2.14	1.04	33.40	0.67	0.70	0.12	0.12	0.00
375	1.53	12.12	0.00	0.00	0.00	1.54	0.00	0.00	0.00
377	0.12	0.00	0.00	1.46	4.46	0.00	0.00	0.00	0.00
381	0.12	0.00	0.00	19.16	0.00	0.00	0.00	0.00	0.00
382	0.81	5.55	21.12	99.68	13.22	0.00	0.48	0.48	0.00
387	1.31	17.87	13.24	57.17	3.05	0.15	0.00	0.00	0.00
388	0.00	0.00	0.41	20.37	0.87	0.00	0.00	0.00	0.00
398	0.00	0.00	0.00	100.00	0.58	0.00	0.00	0.00	0.00
405	0.16	2.60	0.34	38.35	1.82	0.00	0.00	0.00	0.00

406	0.09	2.40	0.20	5.67	3.72	0.10	0.00	0.00	0.00
416	4.53	44.62	0.00	65.96	0.00	0.00	0.00	0.00	0.00
419	7.40	38.64	59.23	40.13	16.49	11.14	0.00	0.00	0.00
423	0.18	1.44	0.39	28.31	1.41	0.26	0.00	0.00	0.00
465	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
467	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
473	0.20	7.73	0.20	33.10	0.85	0.00	0.11	0.11	0.00
474	0.89	5.93	2.35	33.89	33.51	1.20	1.24	1.24	0.00
475	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
476	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
488	1.82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
494	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
517	0.16	4.70	0.19	33.66	1.12	0.02	0.00	0.00	0.00
518	0.81	4.55	1.11	48.10	6.69	0.52	0.16	0.16	0.00
526	0.00	0.00	0.00	84.93	0.00	0.00	0.00	0.00	0.00
527	0.00	0.00	0.00	8.15	0.00	0.00	0.00	0.00	0.00
535	0.50	0.74	0.34	42.99	1.24	0.12	0.15	0.15	0.00
541	0.00	0.31	0.01	34.23	0.71	0.00	0.08	0.08	0.00
543	0.28	0.93	0.80	100.00	1.48	0.33	0.00	0.00	0.00
554	0.04	0.66	0.10	19.28	0.10	0.05	0.00	0.00	0.00
564	9.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
565	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
572	1.00	8.83	1.73	44.47	9.59	0.45	0.25	0.25	0.00
579	0.29	4.27	1.71	30.73	1.59	0.12	0.00	0.00	0.00
585	0.39	0.80	4.00	48.17	5.64	0.38	0.18	0.18	0.00
586	0.00	10.82	0.83	12.57	0.32	0.00	0.00	0.00	0.00
588	0.24	12.61	1.24	100.00	0.69	0.31	0.00	0.00	0.00
594	0.36	8.73	0.00	8.52	0.64	0.13	0.04	0.04	0.00
607	2.46	0.00	100.00	100.00	100.00	100.00	0.00	0.00	0.00

* Sites (Glu and Asp) that were not labeled in any samples or were not covered by the mass spec are left out.

The other azido compounds azidopropylamine (m) and, azido-PEG-amine (n) labeled BSA to a lesser extent compared to many of the alkyne reagents (Table 5).

Table 5. Labeling of BSA with EDC as activator with azidoamine compounds m,n,o.*

Residue	% m Label	% n Label	% o Label
40	0.41	0.00	0.00
41	1.95	0.01	0.00
61	0.04	0.00	0.00
62	0.00	0.00	0.00
69	0.31	0.00	0.03
72	0.56	0.00	0.00
80	0.08	0.00	51.45
81	0.08	0.00	45.59
96	0.00	0.00	0.03
97	0.01	0.00	0.01
106	13.68	0.00	100.00
110	1.17	0.00	0.00
113	1.07	0.00	0.00
124	0.00	0.00	28.81
142	0.08	0.00	4.02
148	0.22	0.00	0.00
149	0.06	0.00	0.00
176	0.07	0.00	4.52
190	0.10	0.00	2.55
195	0.41	0.01	0.00
196	0.07	0.01	3.60
253	20.85	0.00	0.00
267	0.30	0.00	15.90
272	0.07	0.00	0.00
275	0.26	0.00	7.26
278	0.12	0.00	0.00
279	0.12	0.00	5.91
282	0.12	0.00	0.00
289	0.25	0.00	0.00
292	1.12	0.00	0.00
300	0.03	0.00	0.00
315	0.02	0.00	0.00

317	0.03	0.00	1.81
319	0.42	0.00	8.46
323	0.92	0.00	18.32
331	0.82	0.03	1.75
334	0.12	0.00	1.73
335	0.12	0.00	1.05
337	0.03	0.00	1.03
344	0.00	0.00	0.00
347	0.03	0.03	0.00
356	0.03	0.00	0.00
363	0.06	0.00	31.19
375	0.98	0.01	15.55
381	0.41	0.02	0.00
382	0.41	0.01	34.20
387	0.00	0.02	80.43
388	0.00	0.00	34.38
405	0.04	0.00	0.15
406	0.02	0.00	0.00
419	9.85	0.01	0.00
423	0.06	0.00	2.65
473	1.15	0.00	13.20
474	1.41	0.01	23.33
517	0.02	0.00	27.15
518	0.19	0.01	36.48
535	0.09	0.00	17.43
541	0.00	0.01	0.66
543	0.27	0.01	2.09
554	0.01	0.00	0.00
564	20.02	0.00	0.00
565	28.02	0.00	0.00
572	0.70	0.00	5.15
579	0.23	0.00	0.00
585	0.34	0.01	36.52
586	0.05	0.00	4.82

588	0.13	0.00	14.70
607	97.82	0.00	0.00

* Residues that were not labeled in any of the samples are excluded.

N-Hydroxysuccinimide (NHS) was tested as an additive in the labeling reactions with the three azido amine compounds. NHS is a standard additive in carbodiimide coupling reactions that forms a highly reactive ester intermediate which subsequently reacts with amine nucleophiles and has been used previously to enhance EDC crosslinking reactions in proteins (Mintseris et al. 2020). In the EDC facilitated labeling of BSA, NHS generally had no effect on the number of reacted sites (Table 6).

Table 6. Effect of N-hydroxysuccinimide on labeling of BSA with the azidoamine compounds.

Entry	Amine	Mass Shift (Expected)	Mass Shift (Observed)	PSMs	%PSMs
1	11-azido-3,6,9-trioxaundecan-1-amine (n)	200.1274	200.1272	112	1.736
2	3-azidopropylamine (m)	82.0644	82.0642	148	2.339
3	4-azidoaniline (o)	116.0487	-	-	-

Moving forward, GPA (c), with 14% of all PSMs labeled, hydroxyl amine compounds g, with 30% of all PSMs labeled, and h, with 9%, were chosen along with ethynylaniline k, with 23% of all PSMs labeled, as the best candidates for further click enrichment experiments with more complex samples of proteins.

Coupling Agents for GPA Labeling

We evaluated coupling agents with varying hydrophobic and steric properties. The coupling reagents tested were 1-ethyl-3-(3'-dimethylaminopropyl)carbodiimide (EDC, FIG. 7, p), 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide methiodide (EDC methiodide, FIG. 7, q), dicyclohexylcarbodiimide (DCC, FIG. 7, r), 2-ethyl-5-phenylisoxazolium-3'-sulfonate (Woodward's reagent K, WRK, FIG. 7, s), 1-ethyl-3-(3'-dimethylaminopropyl)carbodiimide (DIC, FIG. 7, t), N-cyclohexyl-N'-(2-morpholinylethyl)carbodiimide methyl-p-toluenesulfonate (CMC, FIG. 7, u), and 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride (DMTMM, v). The coupling agents were all screened in the covalent labeling of BSA with GPA as the labeling reagent.

EDC yielded the greatest extent of labeling of BSA of all reagents tested (Tables 7 and 8). Overall, the same Glu and Asp sites on BSA were labeled with GPA in the presence of all the coupling reagents. We hypothesized that the more hydrophobic carboxyl-activating reagents such as DCC and DIC might allow for increased accessibility to hydrophobic sites on the protein leading to different labeling patterns, but this was not the case (Table 8).

Table 7. Results of GPA labeling of BSA with various coupling reagents determined by an open search in FragPipe.

Compound	Coupling agent	PSMs	% PSMs
p	EDC	2143	18.072
q	EDC methiodide	1027	7.708
r	DCC	736	5.574
s	WRK	152	1.089
t	DIC	1096	8.75
u	CMC	761	6.018
v	DMTMM	220	2.203

Table 8. GPA labeling of BSA with activator compounds p-u.*

Residue	% Label EDC (p)	% Label EAC (q)	% Label DCC (r)	% Label WRK (s)	% Label DIC (t)	% Label CMC (u)	% Label DMTMM (v)
37	0.00	0.00	0.01	0.00	0.00	0.00	0.00
40	2.95	1.10	0.00	0.25	0.79	1.57	0.00
41	19.62	4.16	5.18	0.50	100.00	5.41	0.44
61	1.29	0.00	0.15	0.01	0.49	0.00	0.01
62	1.85	0.14	0.00	0.00	99.96	0.00	0.00
69	4.74	0.82	3.31	0.10	4.43	1.56	0.16
72	22.20	0.27	0.14	0.02	99.99	0.19	0.06
80	0.00	1.01	0.00	0.02	0.00	0.00	0.00
81	5.67	0.78	0.31	0.06	0.88	0.59	0.00
82	2.07	0.00	0.58	0.00	99.93	0.37	0.08
87	1.05	0.54	0.50	0.01	99.92	0.29	0.22
96	0.56	0.16	0.30	0.00	1.06	0.21	0.01
97	0.28	0.13	0.96	0.00	99.96	0.07	0.17
106	59.37	1.91	0.00	0.00	4.30	0.00	0.00
110	11.60	0.72	6.27	0.00	99.37	0.00	0.00
113	50.19	8.47	3.69	0.00	99.84	2.99	0.10
116	79.11	15.19	0.00	0.00	99.78	7.19	0.02
119	0.42	0.00	0.00	0.00	0.00	0.00	0.00
124	97.12	25.50	1.41	0.00	20.80	7.29	0.60
131	49.09	7.18	18.43	0.37	58.67	3.47	0.12
132	20.94	2.30	7.25	0.00	0.91	4.04	0.13

135	37.59	2.73	0.60	0.00	99.84	0.50	0.03
142	6.36	2.22	0.11	0.00	2.77	0.13	0.02
148	32.78	2.29	2.78	0.21	99.99	1.11	0.28
149	2.32	0.06	0.11	0.00	99.77	0.02	0.12
153	0.00	0.00	0.01	0.00	0.00	0.12	0.00
154	12.67	0.39	2.59	0.01	99.83	0.88	0.10
176	2.26	0.17	1.99	0.00	1.65	0.22	0.04
190	24.30	1.05	0.61	0.01	1.05	1.43	0.12
195	16.57	0.87	0.13	0.00	99.55	0.27	0.20
196	13.29	0.96	0.06	0.01	99.72	0.24	0.15
250	25.34	1.66	1.44	0.00	14.88	0.14	0.38
253	45.02	5.54	72.13	0.27	99.94	8.22	0.38
267	1.67	1.53	0.14	0.03	0.62	0.93	0.04
272	0.31	0.84	0.39	0.00	99.44	0.03	0.13
275	4.46	2.62	2.19	0.06	12.88	0.29	0.12
278	3.07	0.39	0.16	0.00	0.05	1.18	0.00
279	9.66	0.39	0.37	0.02	1.53	0.44	0.17
282	1.82	0.14	0.07	0.01	0.28	0.12	0.00
289	9.18	0.54	0.00	0.00	0.29	0.02	0.01
292	38.53	8.14	1.01	0.02	99.99	6.06	0.12
300	0.33	4.88	0.00	0.00	0.00	0.00	0.06
303	7.91	0.06	0.00	0.00	0.35	0.03	0.04
308	2.11	0.17	0.04	0.00	99.78	0.11	0.03
315	0.98	0.06	0.00	0.00	0.01	0.01	0.03
317	0.17	1.05	1.33	0.04	98.76	0.97	0.15
319	6.79	5.23	5.50	0.00	5.00	0.73	0.26
323	17.73	2.40	1.13	0.01	99.85	1.61	0.12
331	35.68	4.25	0.76	0.00	2.18	1.62	0.28
334	31.60	5.93	1.34	0.08	2.76	5.39	0.20
335	35.12	0.51	0.14	0.00	0.83	0.61	0.01
337	10.28	1.14	0.46	0.03	2.05	0.48	0.03
344	7.77	0.00	0.00	0.00	0.00	0.00	0.00
347	13.98	1.07	12.11	0.08	4.37	2.47	0.00
356	9.68	1.92	7.36	0.08	99.99	1.82	0.00

363	10.37	1.20	5.43	0.08	2.19	0.93	0.11
375	6.15	1.30	0.05	0.07	2.27	1.45	0.49
377	4.38	0.05	0.38	0.77	99.91	1.41	0.00
381	11.95	0.63	0.08	0.00	99.38	1.25	0.00
382	19.78	0.31	0.18	0.05	99.64	0.00	0.96
387	71.47	3.67	0.22	0.00	4.45	3.42	0.29
388	14.86	0.03	0.00	0.00	1.03	0.00	0.00
398	8.31	0.00	0.00	0.00	0.74	0.00	0.06
405	5.59	0.91	0.67	0.05	0.93	0.81	0.15
406	15.58	1.74	2.32	0.15	99.92	0.95	0.34
416	17.81	8.95	0.00	0.04	0.00	4.42	0.00
419	58.55	6.16	0.01	0.33	5.96	4.44	0.52
423	25.21	2.47	1.92	0.06	3.81	0.79	0.06
473	6.92	1.35	0.82	0.11	2.73	0.67	0.33
474	13.08	2.25	13.35	0.13	99.94	1.88	0.06
488	0.00	0.00	0.11	0.00	87.53	0.00	0.00
517	9.08	1.46	2.52	0.00	2.37	0.80	0.06
518	13.35	0.66	3.03	0.04	99.83	0.44	0.09
535	4.54	0.71	5.21	0.14	0.59	0.53	0.07
541	12.06	1.76	0.99	0.09	99.94	1.53	0.20
543	35.15	4.61	2.12	0.08	99.98	2.21	0.18
554	4.24	0.33	51.43	0.03	0.74	0.16	0.17
564	100.00	0.00	0.00	0.00	0.00	0.00	0.00
572	19.28	1.03	0.16	0.09	1.11	0.79	0.90
579	5.85	2.43	1.63	0.09	99.93	2.39	0.04
585	26.29	1.89	0.03	0.00	1.17	1.19	0.20
586	5.81	0.15	9.64	0.00	96.96	1.08	0.13
588	21.90	1.48	0.23	0.00	23.08	1.18	0.18
594	6.29	1.25	0.08	0.00	12.77	0.02	0.05
607	88.49	14.20	0.79	100.00	30.26	14.17	1.73

* Residues that were not labeled in any samples were excluded.

In the GPA labeling experiments with Woodward's reagent K (WRK), unannotated mass shifts of 69 and 253 m/z were observed in the open search. It was determined that these modifications were the result of a reaction between WRK and Glu and Asp residues as well as other nucleophilic side chains in BSA. In these cases, WRK acted as a covalent label itself instead of an activator. Clickable derivatives of WRK have been used to label and image proteins and were determined to react more readily with histidine and cysteine (Qian et al. 2016).

Labeling Does Not Perturb Protein Structure

To confirm that native solvent accessibility is indeed being monitored by GPA/EDC, GPA labeling was carried out on BSA tryptic peptides and compared to GPA labeling of intact (undigested) BSA. For all sites that were GPA labeled in both samples, the degree of labeling was greater in the pre-digested peptide samples as compared to that of whole protein BSA (FIG. 8). These data suggest that, at least for the conditions used herein, three-dimensional structure is maintained throughout labeling, as destruction of the structure and amelioration of native state shielding yields significantly increased labeling.

GPA and Photocleavable-biotin-azide are the Optimal Reagents for Labeling and Enrichment of Complex Samples

We next tested labeling, clicking, and enrichment on a complex protein sample, cytosol isolated from *Arabidopsis thaliana*. The enrichment scaffolds consisted of a clickable azide group that reacts with the alkyne-containing covalent label on the peptide, a cleavable linker to allow for subsequent elution of the enriched peptides, and a “handle” by which the enrichment is carried out. The “handle” consisted of either agarose beads which allow for unbound peptides to be washed away following the click reaction and biotin which can be bound to streptavidin beads prior to washing steps. Three agarose azide resins with linkers that are photocleavable (FIG. 9, a), hydrazine labile (FIG. 9, b) or cleaved via trypsin digestion (FIG. 9, c) were tested as were two biotin azides with either a photocleavable linker (FIG. 9, d) or acid-labile linker (FIG. 9, e) for subsequent release of the peptide following enrichment with streptavidin agarose.

None of the agarose-based enrichment platforms performed well following the copper click reaction with labeled peptides, vigorous washing steps and cleavage. For the GLAGLLAR-agarose method GPA labeled proteins were clicked to the resin followed by an on-bead digestion with LysC resin which was then washed thoroughly, and peptides eluted by tryptic digestion. Although many labeled peptides were observed with the tryptic agarose as the scaffold, only around 30% of the peptides were labeled with GPA despite extensive washing of the peptide-bound samples (Table 9, entry 3). A smaller number of peptides were identified when DDE agarose was used (Table 9, entry 2). No peptides containing the expected label were observed by MS following the PC-agarose-azide enrichment protocol (Table 9, entry 1). The DADPS-biotin-azide yielded around 400 peptide identifications, but only 44% carried the GPA label (Table 9, entry 7). The PC-biotin-azide method outperformed the others with around 4,000 labeled peptides identified and 98.5% of all peptides carrying the GPA label (Table 9, entry 4).

Table 9. Comparison of the click enrichment methods with GPA-labeled *Arabidopsis* cytosol.

Entry	Scaffold	Scaffold Name	Mass Shift	PSMs	% PSMs	Labeled Peptide Groups	% Labeled
1	a	PC-agarose-azide	-	-	-	-	-
2	b	DDE-agarose-azide	236.1772	43	43.434	29	73
3	c	Tryptic-agarose-azide	266.1494	2738	33.028	4845	30

4	d	PC-biotin-azide (37°C)	194.1236	3809	69.432	1398	98.9
5	d	PC-biotin-azide (20°C)	194.1284	3793	59.998	1442	94.6
6	d	PC-biotin (supernatant)	94.0534	262	3.455	1	1.2
7	e	DADPS-biotin-azide	237.1592	516	38.886	423	44.4

Elevated incubation temperature of the peptide/PC-biotin-azide mixture (Table 9, entries 4,5) had very little effect on the enrichment. To determine the extent of the click reaction between GPA-labeled peptides and the PC-biotin-azide, the supernatant was collected and analyzed following the streptavidin enrichment. Unclicked GPA labeled peptides accounted for only 3.5% of all PSMs (Table 9, entry 6) confirming that the copper-catalyzed click reaction was proceeding efficiently.

Given that PC-biotin-azide was the optimal enrichment method for *Arabidopsis* cytosol, we revisited a handful of alkynylamines that showed promise with BSA, shown in FIG. 5 and Table 2. Of these, GPA yielded the highest number of labeled peptide groups, and highest percentage of labeled peptides following enrichment (Table 10). In terms of total number of labeled PSMs, the ethylene glycol containing hydroxyl amine h gave similar results to GPA, but fewer unique peptides were identified.

Table 10. PC-biotin-azide / streptavidin agarose enrichment of clicked alkynylamine labels in *Arabidopsis* cytosol.

Compound	Proteins IDed	Labeled Peptide Groups	% Labeled	PSMs	% PSMs
c	707	2897	98.9	6367	79.16
g	378	1184	96.1	4698	84
h	470	1567	97.9	5897	80.97
k	311	902	96.6	1985	82.78

Accessing structural characteristics of membrane proteomes is another area to which enrichable labels would be a useful tool. We thus also compared GPA labeling alone with GPA labeling followed by enrichment in a complex sample of microsomes isolated from *Arabidopsis thaliana*. In the unenriched method there were 1195 labeled peptide groups from 614 proteins compared to 4976 labeled peptide groups from 1790 labeled proteins in the GPA-enrichment method. Using the enrichment method allowed for the identification of 770 unique proteins and 2016 unique peptide groups that were not identified in the unenriched sample, demonstrating the advantage of enrichment using this type of footprinting method.

In summary, we found that for complex proteomes, the optimal enrichable footprinting method is EDC-mediated coupling of the labeling agent GPA, PC-biotin-azide addition via click chemistry, streptavidin enrichment, and photocleavage as the elution mechanism for peptide release.

GPA Labelling Compared with Computed SASA for Labeled Proteins

Several well-labeled proteins in the *Arabidopsis* cytosol sample were chosen to compare degree of MS-measured labeling to a normalized SASA value for footprinted sites. The MS-based value was calculated in Excel by searching for peptides with each specific labeled site and summing the abundance of each peptide with that labeled site. The SASA values were calculated using the Accessible Surface Area and Accessibility Calculation for Protein version 1.2 offered by the Center for Informational Biology, Ochanomizu University (cib.cf.ocha.ac.jp) with input files obtained from the AlphaFold Protein Structure Database (Jumper et al. 2021, Varadi et al. 2024). The MS-based abundance values were normalized by scaling between 0 and 1 and compared to the generated relative (0-1) SASA value. In general, correlation between the labeling of a complex sample versus SASA values generated using a predicted “pure” protein structure was low. This was not surprising, given we were probing predicted static structures derived from homology modeling of other crystallized, solid-state structures, and protein footprinting itself accesses in solution, dynamic structures.

There were, however, examples of proteins that still showed some correlation. Fructose-bisphosphate aldolase 2 (ALFP2) from *Arabidopsis* cytosol is one example of such a protein (FIG. 11). Mapping GPA labeling on fructose-bisphosphate aldolase 2 (ALFP2) to a predicted three-dimensional structure revealed that the most heavily labeled sites are all in accessible regions of the protein, while those that were labeled at a much lower level are relatively shielded from solvent.

Tandem Mass Tag (TMT) Labeling

Experiments were performed to show the efficacy of Tandem mass tag (TMT) labeling in the workflow of the present invention for processing multiple samples in parallel. TMT can make the method quantitative and increase throughput for data acquisition and bioinformatics.

An exemplary workflow for processing a single sample without TMT labeling can comprise the following steps, performed in the following order: Lyse cells/tissue; label (*e.g.*, GPA) (increased concentrations for lysate); buffer exchange; click chemistry; buffer exchange; digest; enrich; cleanup; MS analysis. Possible workflow protocols for implementing TMT labeling into the workflow for processing multiple samples in parallel include the following protocols. Protocol 1: Lyse multiple cells/tissue samples; label (*e.g.*, GPA) (increased concentrations for lysate); buffer exchange; click chemistry; buffer exchange; digest; cleanup; TMT labeling; combine samples/cleanup; enrich; cleanup; MS analysis. Protocol 2: Lyse multiple cells/tissue samples; label (*e.g.*, GPA) (increased concentrations for lysate); buffer exchange; digest; cleanup; click chemistry; cleanup; TMT labeling; combine samples; cleanup; enrich; cleanup; MS analysis. Protocol 3: Lyse multiple cells/tissue samples; label (*e.g.*, GPA) (increased concentrations for lysate); buffer exchange; digest; cleanup; TMT labeling; combine samples; cleanup; click; cleanup; enrich; cleanup; MS analysis. All of the protocols outlined above work. This means that the order of TMT labeling versus clicking is flexible.

Protocol 2 (digestion; click chemistry; TMT labeling) and Protocol 3 (digestion; TMT labeling; click chemistry) were tested in *Arabidopsis*. *Arabidopsis* protein was labelled with 45mM GPA and 30mM EDC for 5 min at room temperature, then quenched with 200mM ammonium acetate for 3 min at room temp. Samples were reduced with DTT, alkylated with iodoacetamide, then digested with Trypsin/LysC overnight. Peptides were desalted and concentrated with C18-based solid phase extraction. From here, samples was solubilized into 50mM TEAB and split into 2 (protocol 2 and 3 as outlined above) for further processing.

For protocol 2, samples were clicked for 1 hour at room temperature with 175μM PC biotin azide, 3.3mM sodium ascorbate, and 33μM Cu(II) complexed with the ligand 2-(4-((Bis((1-(tert-butyl)-1H-1,2,3-triazol-4-yl)methyl)amino)methyl)-1H-

1,2,3-triazol-1-yl)acetic acid (BTTAA) at 165μM. Samples were then desalted and concentrated with C18-based solid phase extraction, resolubilized into 50mM TEAB and labeled with 600μg of the TMT reagent (TMT10plex™ Isobaric Label Reagent Set, Product No. 90110, ThermoFisher Scientific, Waltham, MA) for 1 hour at room temp, quenched with 5% hydroxylamine for 5 min at room temp, then again desalted and concentrated with C18-based solid phase extraction.

For protocol 3, samples were labeled with 600μg of the TMT reagent (TMT10plex™ Isobaric Label Reagent Set, Product No. 90110, ThermoFisher Scientific, Waltham, MA) for 1 hour at room temp, quenched with 5% hydroxylamine for 5 min at room temp, then desalted and concentrated with C18-based solid phase extraction. Samples then were resolubilized into resolubilized into 50mM TEAB and clicked for 1 hour at room temperature with 175μM PC biotin azide, 3.3mM sodium ascorbate, and 33μM Cu(II) complexed with the ligand BTTAA at 165μM, then again desalted and concentrated with C18-based solid phase extraction.

Both samples were resolubilized into PBS for streptavidin enrichment and were enriched as previously described.

In Protocol 2, 7,955/8,028 peptides were GPA labeled, 2,985/8,028 peptides were TMT labeled, and 2,960/8,028 peptides were dual labeled. In Protocol 3, 9,277/9,382 peptides were GPA labeled, 6,926/9,382 peptides were TMT labeled, and 6,861/9,382 peptides dual labeled.

Protocol 1 was tested in HeLa cells treated with rapamycin. HeLa cells were treated *in vivo* with (vs. empty buffer) to quantify GPA labelling differences using the technique and show that the method corroborates known biology. Separate samples included 100 μg input. Cells were treated with 2 μM rapamycin or DMSO control (n=3 for each condition). The labeling was performed with 45 mM GPA and 50 mM EDC for 5 minutes at room temperature with gentle vortexing. TMT quantification was performed with Proteome Discoverer 2.4. Among 4,145 peptides total, 4,118 had a TMT label, 3,652 had a cleaved GPA tag (87%), and 2,686 were dual labeled/quantified. Of the 2,686 dual labeled/quantified peptides, 255 were statistically significantly different in rapamycin-treated versus control samples (Table 11). The different peptides in rapamycin-treated versus control samples were from FK506-binding protein (FKBP) rapamycin-binding family members. This experiment shows that of the thousands of quantified GPA-labelled peptides, peptides from four different members of the family of proteins that bind rapamycin to ultimately effect cellular change were among the statistically significant differences between treated and control samples. This demonstrates the ability of method of the invention to detect *in vivo* conformational changes in mammalian cell systems in response to drug treatments.

Table 11. Differently labeled peptides from members of FKBP rapamycin-binding family members.

Identified Sequence	Modifications	Uniprot Accessions	Positions in Proteins	Protein Descriptions	Abundance Ratio: Rapamycin/ Control	Abundance Ratio P-Value
IEVEKPFAIAKE (SEQ ID NO:2)	2xTMT6plex [K5; K11]; 1xTMT6plex	P23284	P23284 [205-216]	Peptidyl-prolyl cis-trans isomerase B	1.356	4.92E-02

	[N-Term]; 1xpcGPA194 [E]			OS=Homo sapiens OX=9606 GN=PPIB PE=1 SV=2		
ITIADCGQLE (SEQ ID NO:3)	1xCarbamidomethyl [C6]; 1xTMT6plex [N-Term]; 1xpcGPA194 [E10]	P62937	P62937 [156-165]	Peptidyl-prolyl cis-trans isomerase A OS=Homo sapiens OX=9606 GN=PPIA PE=1 SV=2	1.418	2.91E-02
FEDENFILK (SEQ ID NO:4)	1xTMT6plex [K9]; 1xTMT6plex [N-Term]; 1xpcGPA194 [E2]	P62937	P62937 [83-91]	Peptidyl-prolyl cis-trans isomerase A OS=Homo sapiens OX=9606 GN=PPIA PE=1 SV=2	1.249	5.42E-03
SEETLDEGPPK (SEQ ID NO:5)	1xTMT6plex [K11]; 1xTMT6plex [N-Term]; 1xpcGPA194 [E/D]	Q00688	Q00688 [100-110]	Peptidyl-prolyl cis-trans isomerase FKBP3 OS=Homo sapiens OX=9606 GN=FKBP3 PE=1 SV=1	1.379	4.69E-02
LTFEVELVDID (SEQ ID NO:6)	1xTMT6plex [N-Term]; 1xpcGPA194 [D11]	Q00688	Q00688 [214-224]	Peptidyl-prolyl cis-trans isomerase FKBP3 OS=Homo	1.181	2.24E-02

				sapiens OX=9606 GN=FKBP3 PE=1 SV=1		
FEIGEGENLDLPYGLER (SEQ ID NO:7)	1xTMT6plex [N-Term]; 1xpcGPA194 [E2]	Q02790	Q02790 [190-206]	Peptidyl-prolyl cis-trans isomerase FKBP4 OS=Homo sapiens OX=9606 GN=FKBP4 PE=1 SV=3	1.219	1.55E-02

Conclusions

In conclusion, we sought to design a workflow that would enable in-solution solvent-accessibility measurements on proteome-wide scales. After testing many reagents, both with the model protein BSA and complex proteomic extracts, we report herein that the molecule glycine propargyl amide (GPA), as an alternative to the classic labeling reagent glycine ethyl ester, is advantageous for this purpose. Not only can GPA-labeled peptides be further derivatized with click chemistry and enriched using standard enrichment techniques, but the GPA adduct is not subject to hydrolytic degradation. Thus, many lower abundance labeling events can easily be enriched and identified from complex proteomes. We hope the method re-ported herein becomes widely used and built upon to measure proteome-wide conformational changes.

References

Akashi, S.; Niitsu, U.; Yuji, R.; Ide, H.; Hirayama, K. Investigation of the Interaction Between Enzyme and Inhibitor by the Combination of Chemical Modification, Electrospray Ionization Mass Spectrometry and Frit-Fast Atom Bombardment Liquid Chromatography/Mass Spectrometry. Biol. Mass Spectrom. 1993, 22 (2), 124-132. DOI: 10.1002/bms.1200220205

Edwards, J.O.; Pearson, R.G. The Factors Determining Nucleophilic Reactivities. J. Am. Chem. Soc. 1962, 84 (1), 16-24. DOI: 10.1021/ja00860a005

Gilles, M. A. H., A.Q.; Borders, Jr., C.L. Stability of Water-Soluble Carbodiimides in Aqueous Solution. Anal. Biochem. 1990, 184, 244-248. DOI: 10.1016/0003-2697(90)90675-y

Grabarek, Z. G., J. Zero-Length Crosslinking Procedure with the Use of Active Esters. Anal. Biochem. 1990, 185, 131-135. DOI: 10.1016/0003-2697(90)90267-D

Guo, C.; Cheng, M.; Gross, M.L. Protein-Metal-Ion Interactions Studied by Mass Spectrometry-Based Footprinting with Isotope-Encoded Benzhydrazide. Anal. Chem. 2019, 91 (2), 1416-1423. DOI: 10.1021/acs.analchem.8b04088

Haruta, M.; Sabat, G.; Stecker, K.; Minkoff, B. B.; Sussman, M. R. A Peptide Hormone and its Receptor Protein Kinase Regulate Plant Cell Expansion. Science 2014, 343 (6169), 408-411. DOI: 10.1126/science.1244454

Hoare, D.G.; Koshland, Jr., D.E. A Procedure for the Selective Modification of Carboxyl Groups in Proteins. J. Am. Chem. Soc. 1966, 88 (9), 2057-2058. DOI: 10.1021/ja00961a045

Hu, Z.; Li, H.; Wang, X.; Ullah, K.; Xu, G. Proteomic Approaches for the Profiling of Ubiquitylation Events and Their Applications in Drug Discovery. J. Proteomics 2021, 231, 103996. DOI: 10.1016/j.jprot.2020.103996

Johnson, D.T.; Di Stefano, L.H.; Jones, L.M. Fast Photochemical Oxidation of Proteins (FPOP): A Powerful Mass Spectrometry-based Structural Proteomics Tool. J. Biol. Chem. 2019, 294 (32), 11969-11979. DOI: 10.1074/jbc.REV119.006218

Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; Bridgland, A.; Meyers, C.; Kohl, S.A.A.; Ballard, A.J. Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A.W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D.. Highly Accurate Protein Structure Prediction with AlphaFold. Nature 2021, 596, 583-589. DOI: 10.1038/s41586-021-03819-2.

Kiselar, J. G.; Maleknia, S. D.; Sullivan, M.; Downard, K. M.; Chance, M. R. Hydroxyl Radical Probe of Protein Surfaces Using Synchrotron X-Ray Radiolysis and Mass Spectrometry. Int. J. Radiat. Biol. 2002, 78, 101—114. DOI: 10.1080/09553000110094805

Kong, A.T., et al., MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. Nat Methods, 2017. 14(5): p. 513-520.

Limpikirati, P.; Liu, T.; Vachet, R.W. Covalent Labeling-Mass Spectrometry with Non-Specific Reagents for Studying Protein Structure and Interactions. Methods 2018, 144, 79-93. DOI: 10.1016/j.ymeth.2018.04.002

Liu, X. R.; Zhang, M. M.; Gross, M. L. Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications. Chem. Rev. 2020, 120 (10), 4355-4454. DOI: 10.1021/acs.chemrev.9b00815

McKenzie-Coe, A.; Montes, N. S.; Jones, L. M. Hydroxyl Radical Protein Footprinting: A Mass Spectrometry-Based Structural Method for Studying the Higher Order Structure of Proteins. Chem. Rev. 2022, 122 (8), 7532-7561. DOI: 10.1021/acs.chemrev.1c00432

Minkoff, B. B.; Blatz, J. M.; Choudhury, F. A.; Benjamin, D.; Shohet, J. L.; Sussman, M. R. Plasma-Generated OH Radical Production for Analyzing Three-Dimensional Structure in Protein Therapeutics. Sci. Rep. 2017, 7 (1), 12946. DOI: 10.1038/s41598-017-13371-7

Mintseris, J.; Gygi, S. P. High-density Chemical Cross-linking for Modeling Protein Interactions. Proc. Natl. Acad. Sci. U.S. A. 2020, 117 (1), 93-102. DOI: 10.1073/pnas.1902931116

Qian, Y.; Schurmann, M.; Janning, P.; Hedberg, C.; Waldmann, H. Activity-Based Proteome Profiling Probes Based on Woodward's Reagent K with Distinct Target Selectivity. Angew. Chem. Int. Ed. Engl. 2016, 55 (27), 7766-7771. DOI: 10.1002/anie.201602666

Rostovtsev, V.V.; Green, L.G.; Fokin, V.V.; Sharpless, K.B. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective “Ligation” of Azides and Terminal Alkynes. Angew. Chem. Int. Ed. 2002, 41, 2596-2599. DOI: 10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4

Sanderson, R.J.; Mosbaugh, D.W. Identification of Specific Carboxyl Groups on Uracil-DNA Glycosylase Inhibitor Protein that are Required for Activity. J. Biol. Chem. 1996, 271 (46), 29170-29181. DOI: 10.1074/jbc.271.46.29170

Thingholm, T.E.; Jensen, O.N.; Larsen, M.R. Analytical Strategies for Phosphoproteomics. Proteomics 2009, 9 (6), 1451-1468. DOI: 10.1002/pmic.20080045?4

Tornøe, C.W.; Christensen, C.; Meldal, M. Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. J. Org. Chem. 2002, 67 (9), 3057-3064. DOI: 10.1021/jo011148j

West, G.M.; Tang, L.; Fitzgerald, M.C. Thermodynamic Analysis of Protein Stability and Ligand Binding Using a Chemical Modification- and Mass Spectrometry-Based Strategy. Anal. Chem. 2008, 80 (11), 4175-4185. DOI: 10.1021/ac702610a

Valeur, E.; Bradley, M. Amide Bond Formation: Beyond the Myth of Coupling Reagents. Chem. Soc. Rev. 2009, 38 (2), 606-631. DOI: 10.1039/b701677h

Varadi, M.; Bertoni, D.; Magana, P.; Paramval, U.; Pidruchna, I.; Radhakrishnan, M.; Tsenkov, M.; Nair, S.; Mirdita, M.; Yeo, J.; Kovaleskiy, O.; Tunyasuvunakool, K.; Laydon, A.; Zidek, A.; Tomlinson, H.; Hariharan, D.; Abrahamson, J.; Green, T.; Jumper, J.; Birney, E.; Steinegger, M.; Hassabis, D.; Velankar, S. AlphaFold Protein Structure Database in 2024: Providing Structure Coverage for Over 214 Million Protein Sequences. Nucleic Acids Res. 2024, 52 (D1), D368-D375. DOI: 10.1093/nar/gkad101

Yu, F.; Teo, G. C.; Kong, A. T.; Haynes, S. E.; Avtonomov, D. M.; Geiszler, D. J.; Nesvizhskii, A. I. Identification of Modified Peptides Using Localization-aware Open Search. Nat. Commun. 2020, 11 (1), 4065, DOI: 10.1038/s41467-020-17921-y

Zhang, H.; Wen, J.; Huang, R.Y-C.; Blankenship, R.E.; Gross, M.L. Mass Spectrometry-based Carboxyl Footprinting of Proteins: Method Evaluation. Int. J. Mass. Spectrom. 2012, 312, 78-86. DOI: 10.1016/j.ijms.2011.07.015

Zhang, H.; Liu, H.; Blankenship, R.E.; Gross, M.L. Isotope-Encoded Carboxyl Group Footprinting for Mass Spectrometry-Based Protein Conformational Studies. 2016, 27, 178-181. DOI: 10.1007/s13361-015-1260-5

CLAIMS

What is claimed is:

1. A method of labeling and, optionally, processing and analyzing a protein sample, the method comprising:

reacting the protein sample in a first labeling reaction with a carboxylic acid activating agent and a first labeling agent to label a first set of protein residues in the protein sample with the first labeling agent, wherein the first labeling agent has the structure of Nu-L^1-Y^1, wherein Nu is a nucleophilic group, L^1 is a first linker, and Y^1 is a first click chemistry handle.

2. The method of any prior claim, wherein the protein sample during the first labeling reaction comprises a protein mixture comprising at least 100 different proteins.

3. The method of any prior claim, wherein the protein sample during the first labeling reaction comprises a full cellular proteome.

4. The method of any prior claim, wherein the nucleophilic group is selected from the group consisting of substituted or unsubstituted amino, hydroxyl, and thiol.

5. The method of any prior claim, wherein the nucleophilic group is unsubstituted amino.

6. The method of any prior claim, wherein the carboxylic acid activating agent is selected from the group consisting of a carbodiimide coupling agent and 2-ethyl-5-phenylisoxazolium-3'-sulfonate.

7. The method of any prior claim, wherein the carbodiimide coupling agent is selected from the group consisting of 1-ethyl-3-(3'-dimethylaminopropyl)carbodiimide (EDC), 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide methiodide (EDC), dicyclohexylcarbodiimide (DCC), 1-ethyl-3-(3'-dimethylaminopropyl)carbodiimide (DIC), and N-cyclohexyl-N'-(2-morpholinylethyl)carbodiimide methyl-p-toluenesulfonate (CMC).

8. The method of any prior claim, wherein the first linker comprises one or more instances of one or more first linking moieties selected from the group consisting of optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroaralkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted arylene, optionally substituted heteroarylene, oxygen, sulfur, phosphorus, silicon, secondary amine, tertiary amine, carbonyl, and sulfonyl.

9. The method of any prior claim, wherein the first linker comprises one or more instances of one or more first linking moieties selected from the group consisting of:

$-(CH_2)_x-$;

$-(CH_2)_x-$, $-(CH_2CH_2O)_x-$, $-(OCH_2CH_2)_x-$, $-(CH_2O)_x-$, and $-(OCH_2)_x-$;

$-(CH_2)_x-$, $-C(O)NH-$, and $-NHC(O)-$; or

para-phenylene, *meta*-phenylene, and *ortho*-phenylene,

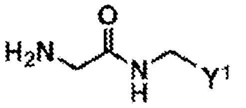
wherein x in each instance is independently an integer from 1 to 6.

10. The method of any prior claim, wherein the first linker comprises a backbone chain that is devoid of an ester group.

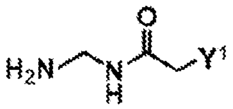
11. The method of any prior claim, wherein the first click chemistry handle is selected from the group consisting of a terminal alkyne, an azide, a strained alkyne, a diene, a dieneophile, an alkoxyamine, a phosphine, a hydrazide, a thiol, a tetrazine, and an alkene.

12. The method of any prior claim, wherein the first click chemistry handle is selected from the group consisting of a terminal alkyne and an azide.

13. The method of any prior claim, wherein the first labeling agent is selected from the group consisting of:



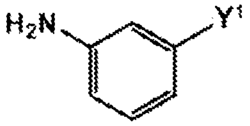
;



;

;

; and



,

wherein Y^1 is a terminal alkyne.

14. The method of any prior claim, further comprising reacting the protein sample in a second labeling reaction with a second labeling agent to label a second set of protein residues in the protein sample with the second labeling agent, wherein the second labeling agent has the structure of Y^2-L^2-Z, wherein Y^2 is a second click chemistry handle, L^2 is a second linker, Z is a label, and the second click chemistry handle is a partner click chemistry handle of the first click chemistry handle.

15. The method of claim 14, wherein the second click chemistry handle is selected from the group consisting of a terminal alkyne, an azide, a strained alkyne, a diene, a dieneophile, an alkoxyamine, a phosphine, a hydrazide, a thiol, a tetrazine, and an alkene.

16. The method of any one of claims 14-15, wherein the second linker is a cleavable linker.

17. The method of any one of claims 14-16, wherein the label comprises a ligand.

18. The method of any one of claims 14-17, wherein the second labeling agent comprises PC biotin azide (1-[4-[[22-[(3aR,4R,6aS)-Hexahydro-2-oxo-1H-thieno[3,4-d]imidazol-4-yl]-4,18-dioxo-8,11,14-trioxa-5,17-diazadocos-1-yl]oxy]-5-methoxy-2-nitrophenyl]ethyl N-(3-azidopropyl)carbamate).

19. The method of any one of claims 14-18, further comprising fragmenting the protein sample.

20. The method of claim 19, wherein the fragmenting is performed after the first labeling reaction and prior to the second labeling reaction.

21. The method of claim 19, wherein the fragmenting is performed after the first labeling reaction and after the second labeling reaction.

22. The method of any one of claims 14-21, further comprising separating labeled polypeptides from unlabeled polypeptides, wherein the labeled polypeptides are polypeptides comprising at least one protein residue from the second set of protein residues, and wherein the unlabeled polypeptides are polypeptides lacking any of the second set of protein residues.

23. The method of claim 22, further comprising, after the separating, cleaving the second linkers to separate the labeled polypeptides from the label.

24. The method of any one of claims 14-23, further comprising determining amino acid sequences of the labeled polypeptides.

25. The method of any prior claim, wherein the protein sample comprises multiple protein samples and the method further comprises sample-labeling each of the multiple protein samples with detectably different sample labels and subsequently combining the multiple protein samples.

26. The method of claim 25, wherein the combining the multiple protein samples is performed prior to determining the amino acid sequences of the labeled polypeptides.

27. The method of any one of claims 25-26, wherein the detectably different sample labels are mass labels.

28. The method of any prior claim, wherein the protein sample comprises a first protein sample and a second protein sample, wherein, during the first labeling reaction, a proteome of the first protein sample is identical to a proteome of the second protein sample, and wherein, during the first labeling reaction, the first protein sample comprises at least one molecule not present in the second protein sample.

29. The method of any prior claim, wherein the protein sample comprises a first protein sample and a second protein sample, wherein, during the first labeling reaction, the first protein sample comprises a first sequence variant of a protein and the second

sample comprises a second sequence variant of a protein.

30. The method of any one of claims 28 and 29, wherein the method further comprises comparing the sequences of the labeled polypeptides from the first protein sample with the sequences of the labeled polypeptides from the second protein sample.

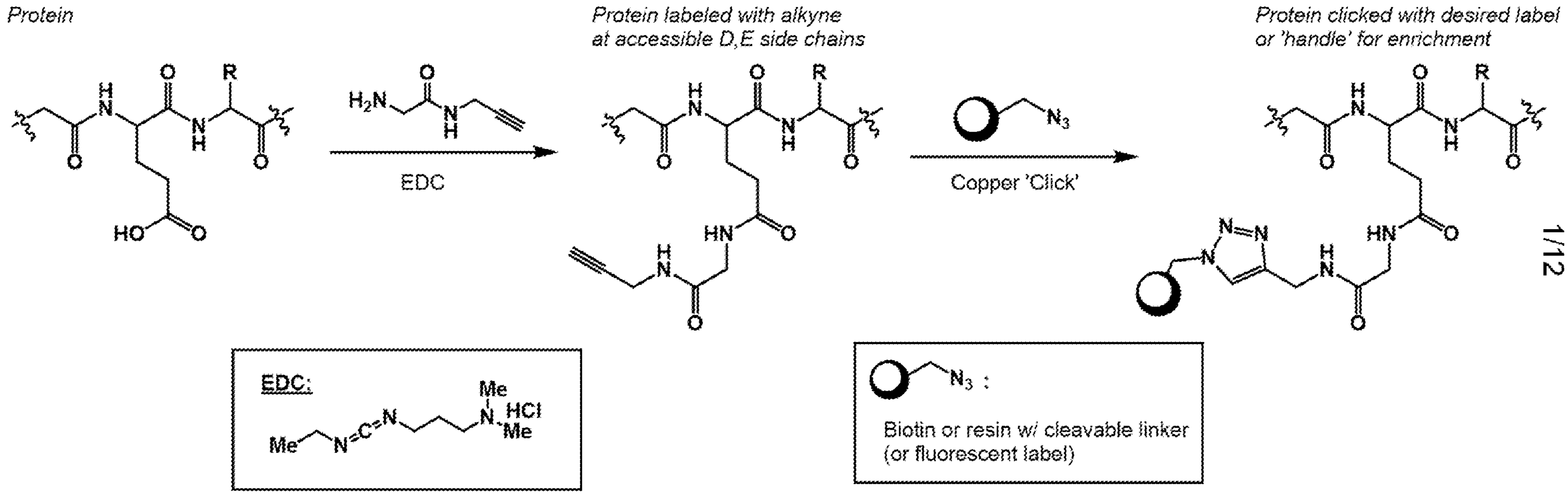

Protein
R
O
H
N
N
H
O
O
HO
O
H$_2$N
O
N
H
EDC
Protein labeled with alkyne
at accessible D,E side chains
HN
H
N
N$_3$
Copper 'Click'
Protein clicked with desired label
or 'handle' for enrichment
N=N
N
EDC:
Me
N
HCl
Me
Me
N=C=N
N$_3$:
Biotin or resin w/ cleavable linker
(or fluorescent label)

FIG. 1

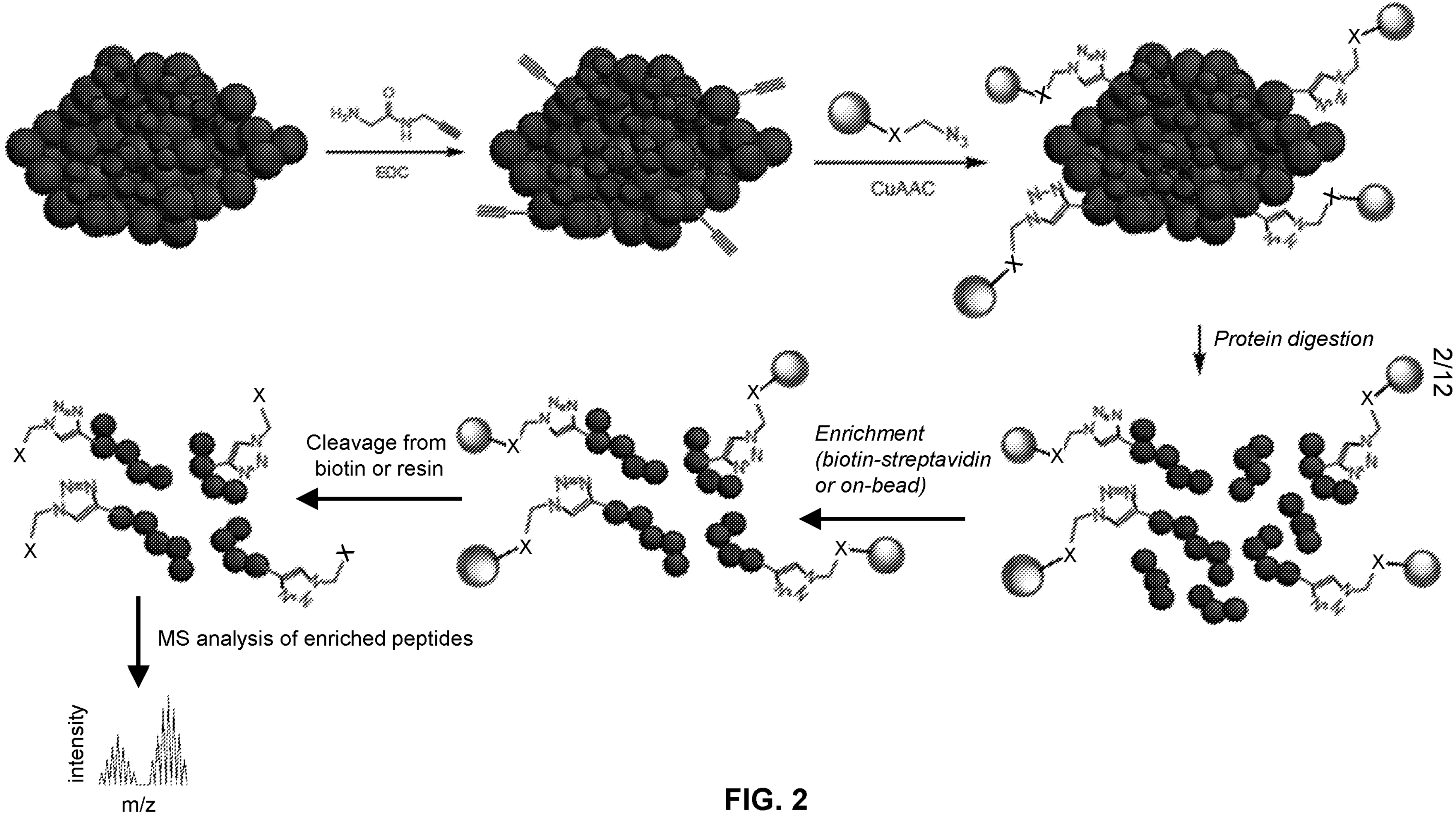

EDC
CuAAC
X
N_3
Protein digestion
Enrichment
(biotin-streptavidin
or on-bead)
Cleavage from
biotin or resin
MS analysis of enriched peptides
intensity
m/z

FIG. 2

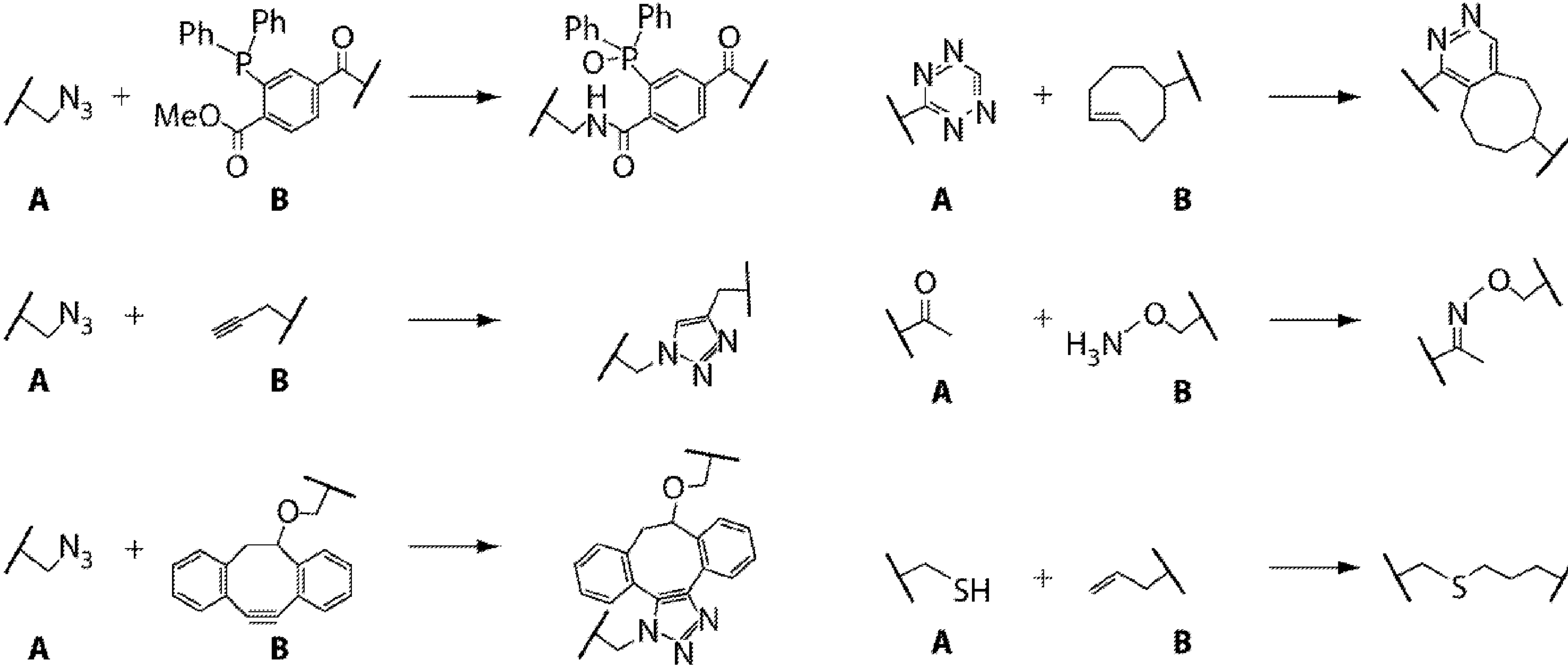


FIG. 3A

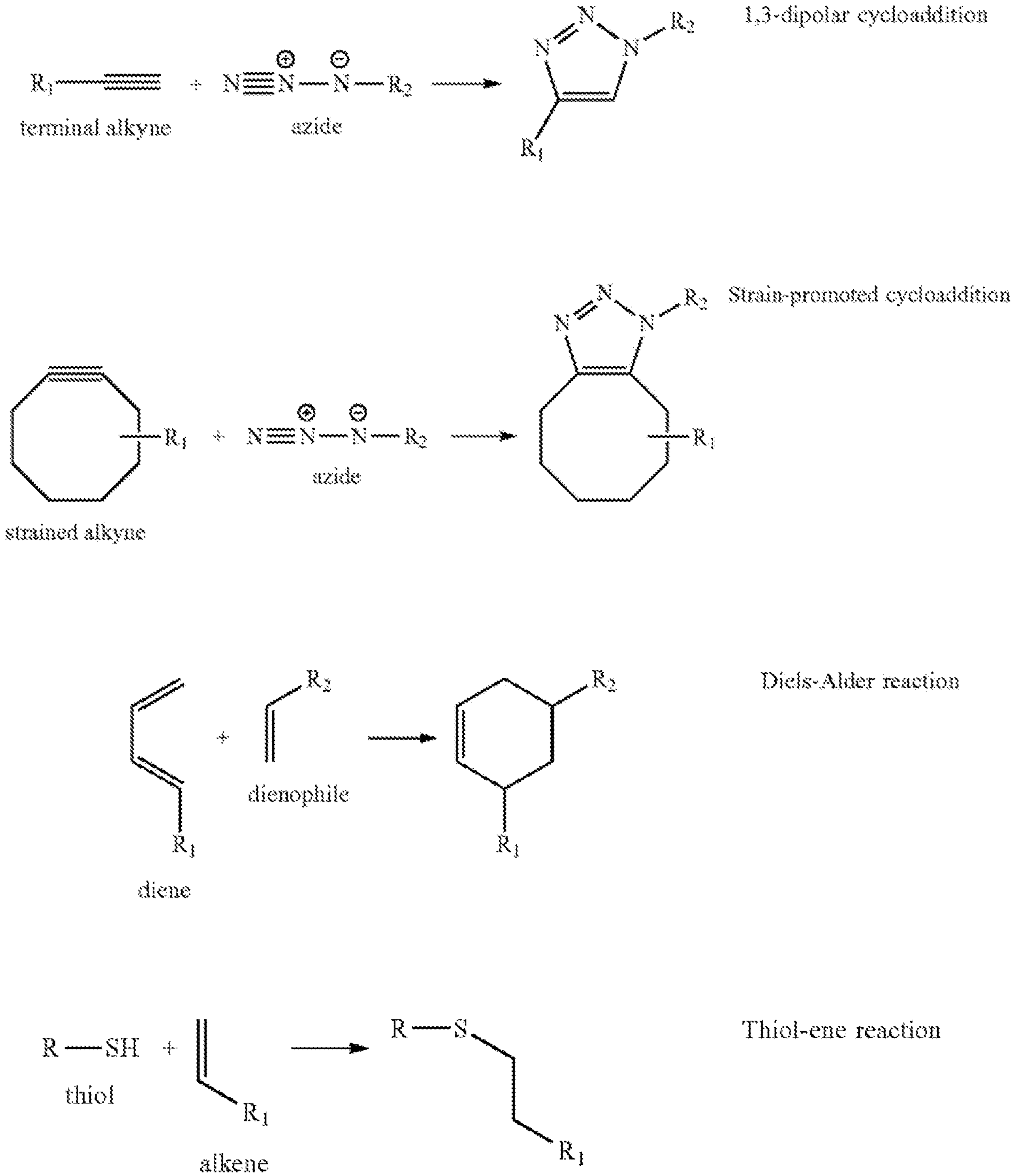

R1
terminal alkyne
+
N
N
N
R2
azide
N
N
N
R2
R1
1,3-dipolar cycloaddition
strained alkyne
R1
+
N
N
N
R2
azide
N
N
N
R2
R1
Strain-promoted cycloaddition
R1
diene
+
R2
dienophile
R2
R1
Diels-Alder reaction
R
SH
thiol
+
R1
alkene
R
S
R1
Thiol-ene reaction

FIG. 3B

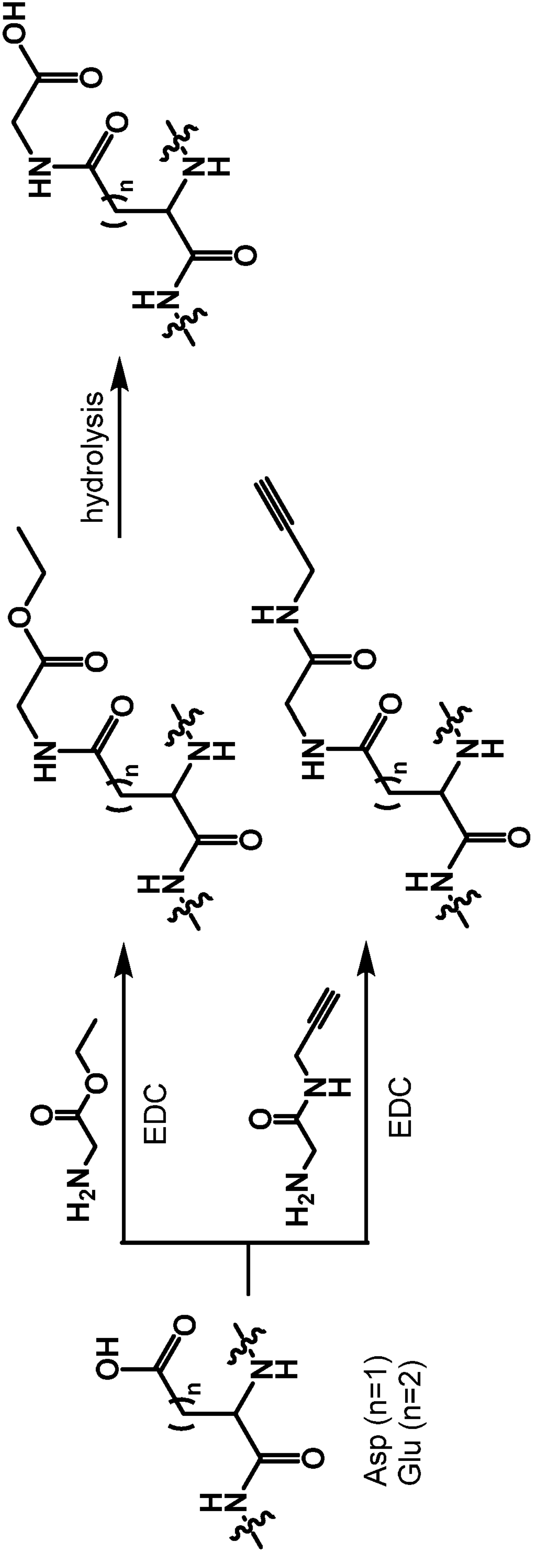
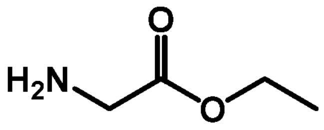
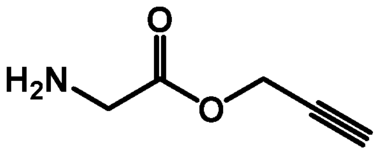


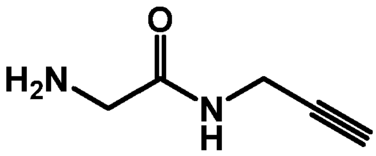
FIG. 4



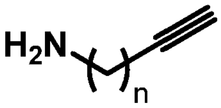
a



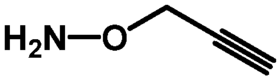
b



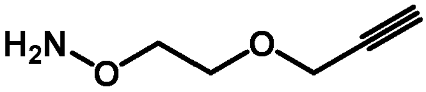
c



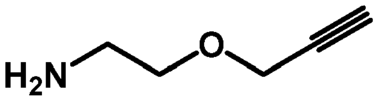
d, n=1
e, n=2
f, n=3



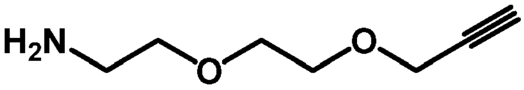
g



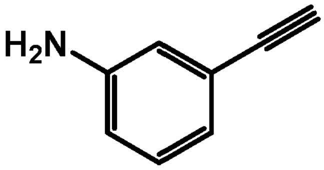
h



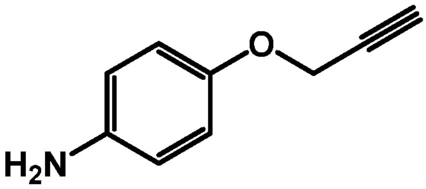
i



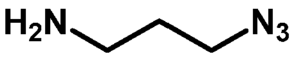
j



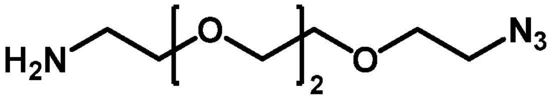
k



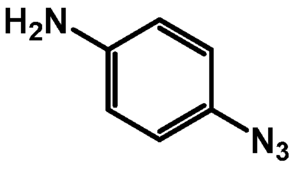
l



m



n



o

FIG. 5

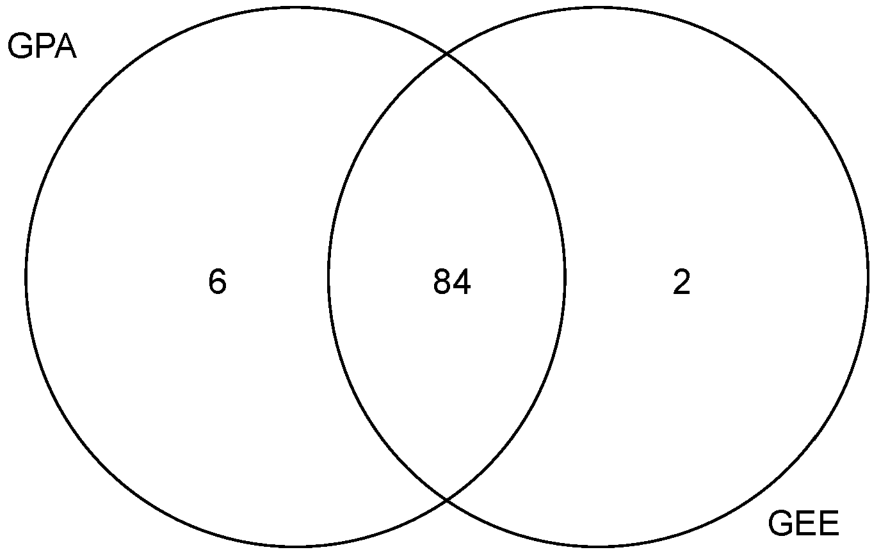
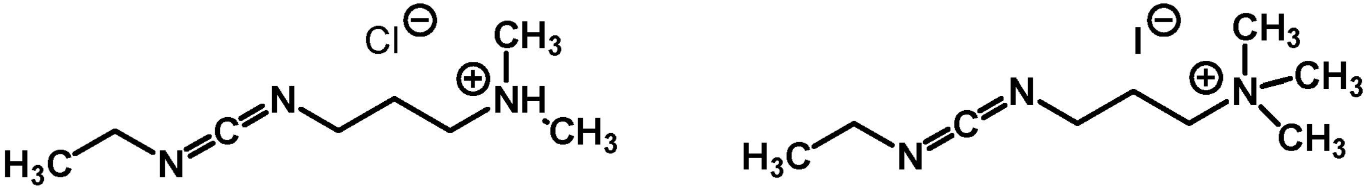
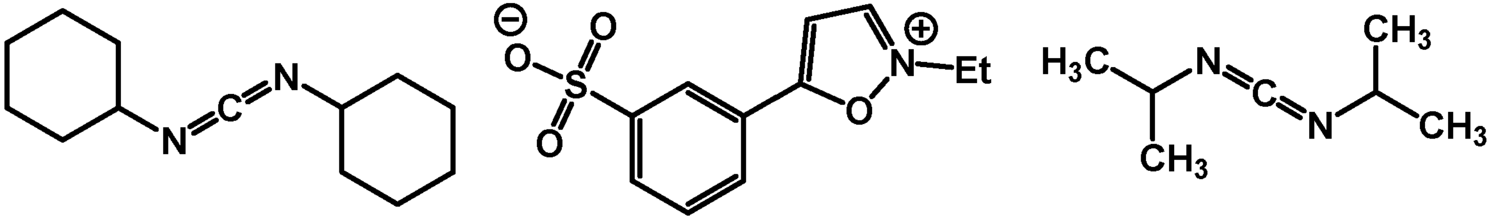


FIG. 6



p (EDC)

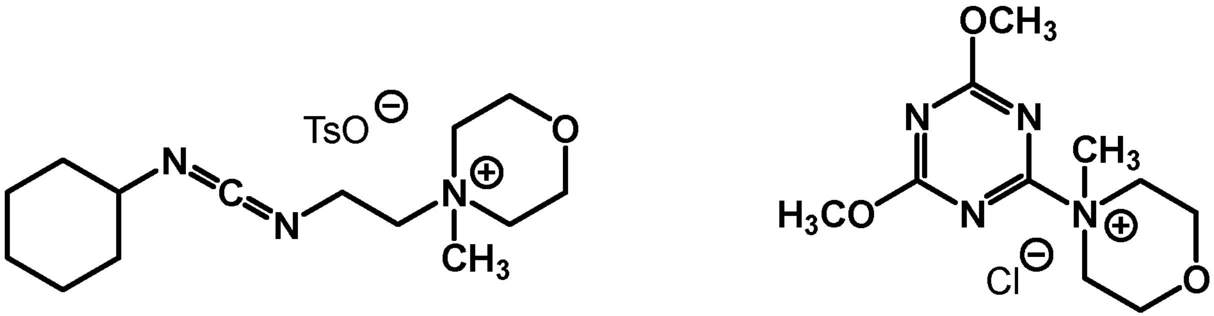
q (EDC methiodide)



r (DCC)

s (WRK)

t (DIC)



u (CMC)

v (DMTMM)

FIG. 7

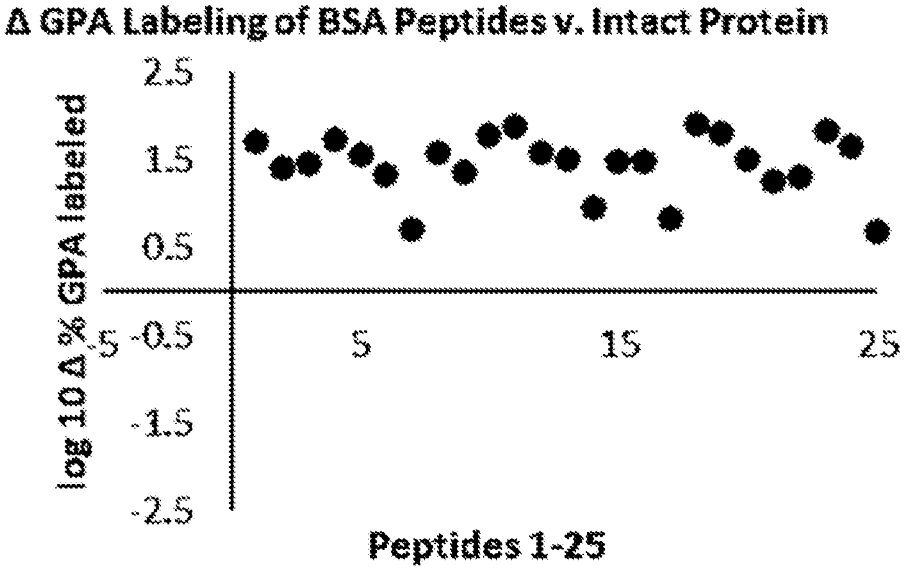
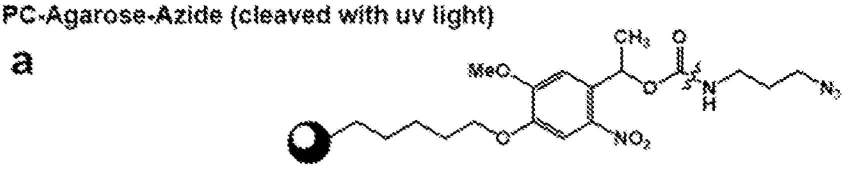
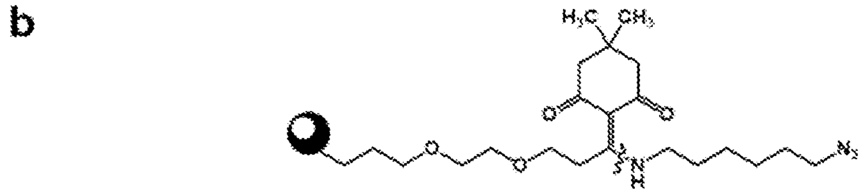

Δ GPA Labeling of BSA Peptides v. Intact Protein
log 10 Δ % GPA labeled
2.5
1.5
0.5
-0.5
-1.5
-2.5
-5
5
15
25
Peptides 1-25

FIG. 8

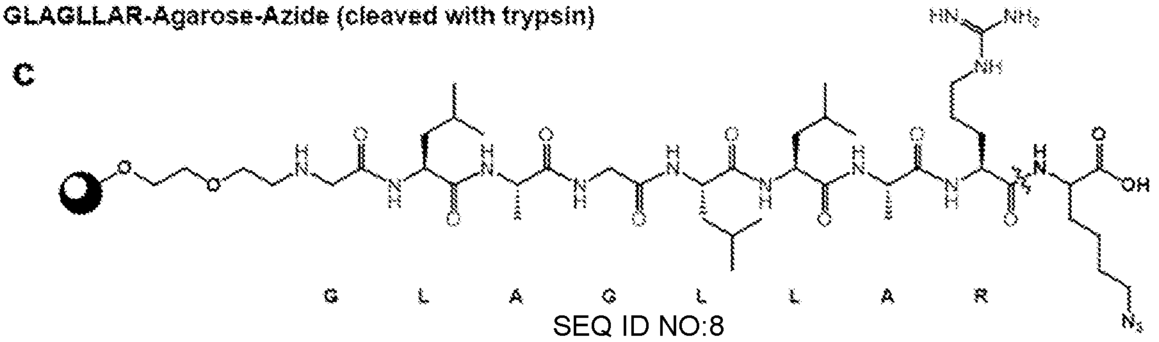
PC-Agarose-Azide (cleaved with uv light)



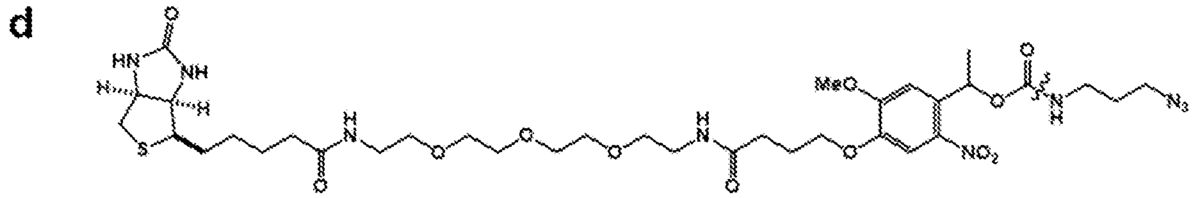
Dde-Agarose-azide (cleaved with 2% hydrazine)



GLAGLLAR-Agarose-Azide (cleaved with trypsin)



PC-Biotin-Azide (cleaved with uv light)



DADPS-Biotin-Azide (cleaved with 2% formic acid)

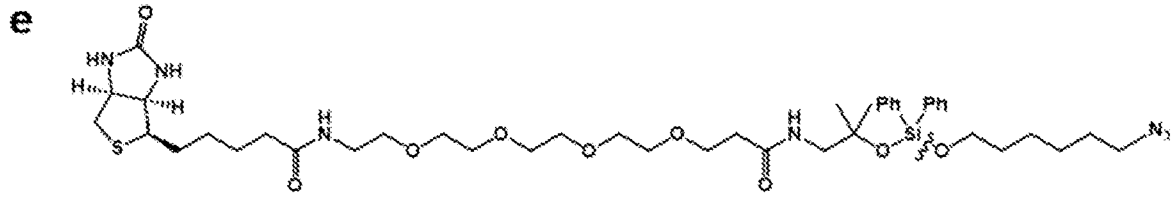


FIG. 9

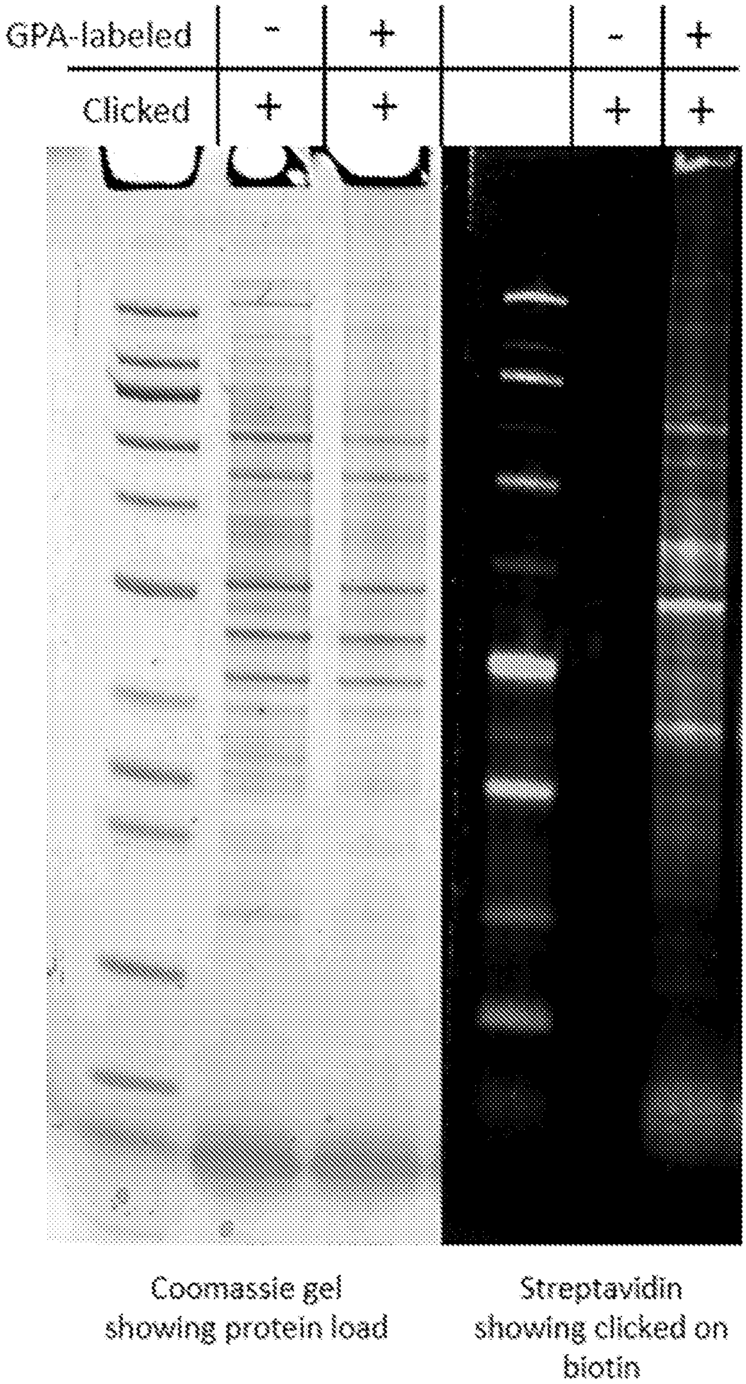

GPA-labeled
-
+
-
+
Clicked
+
+
+
+
Coomassie gel showing protein load
Streptavidin showing clicked on biotin

FIG. 10

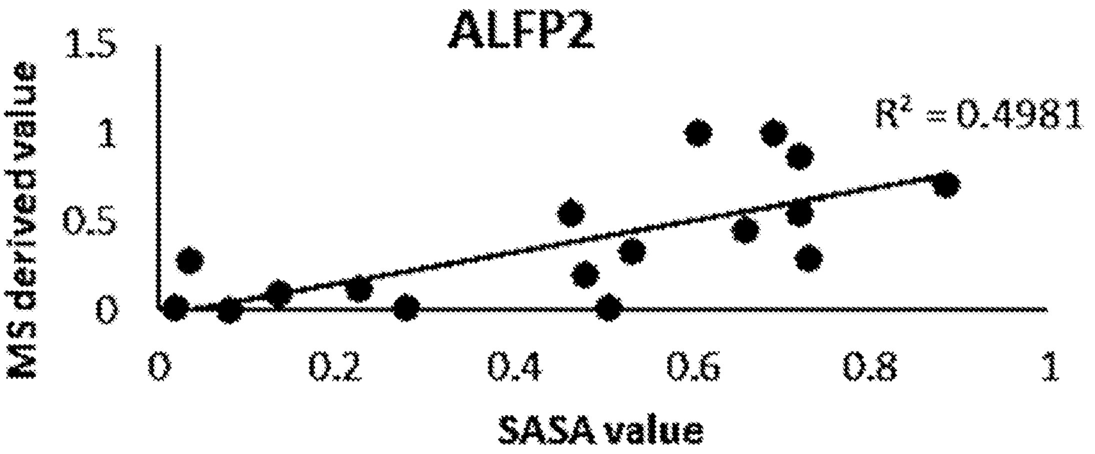

ALFP2
MS derived value
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
SASA value
$R^2 = 0.4981$

FIG. 11