(12) STANDARD PATENT (11) Application No. AU 2017259794 B2
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Macromolecule analysis employing nucleic acid encoding

(51) International Patent Classification(s)
C12N 15/10 (2006.01) ***C40B 40/04*** (2006.01)
C40B 20/04 (2006.01) ***C40B 70/00*** (2006.01)

(21) Application No: **2017259794** (22) Date of Filing: **2017.05.02**

(87) WIPO No: **WO17/192633**

(30) Priority Data

(31) Number	(32) Date	(33) Country
62/376,886	**2016.08.18**	**US**
62/339,071	**2016.05.19**	**US**
62/330,841	**2016.05.02**	**US**

(43) Publication Date: **2017.11.09**
(44) Accepted Journal Date: **2023.04.13**

(71) Applicant(s)
Encodia, Inc.

(72) Inventor(s)
Chee, Mark;Gunderson, Kevin;Weiner, Michael Phillip

(74) Agent / Attorney
Pizzeys Patent and Trade Mark Attorneys Pty Ltd, GPO Box 1374, Brisbane, QLD, 4001, AU

(56) Related Art
US 2009/0264300 A1
US 2008/0305957 A1

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

CORRECTED VERSION

(19) **World Intellectual Property Organization**
International Bureau

(43) **International Publication Date**
09 November 2017 (09.11.2017)

WIPO | PCT

(10) **International Publication Number**
WO 2017/192633 A9

(51) **International Patent Classification:**
C12N 15/10 (2006.01) *C40B 40/04* (2006.01)
C40B 20/04 (2006.01) *C40B 70/00* (2006.01)

(21) **International Application Number:** PCT/US2017/030702

(22) **International Filing Date:** 02 May 2017 (02.05.2017)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**

62/330,841	02 May 2016 (02.05.2016)	US
62/339,071	19 May 2016 (19.05.2016)	US
62/376,886	18 August 2016 (18.08.2016)	US

(71) **Applicant: PROCURE LIFE SCIENCES INC.** [US/US]; 11125 Flintkote Ave., Suite B, San Diego, California 92121 (US).

(72) **Inventors: CHEE, Mark**; 11125 Flintkote Ave., Suite B, San Diego, California 92121 (US). **GUNDERSON, Kevin**; 11125 Flintkote Ave., Suite B, San Diego, California 92121 (US). **WEINER, Michael, Phillip**; 52 Leighton Trail, Guilford, Connecticut 06437 (US).

(74) **Agent: SUN, Eileen, S.** et al.; Seed Intellectual Property Law Group LLP, Suite 5400, 701 Fifth Avenue, Seattle, Washington 98104-7064 (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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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,

(54) **Title:** MACROMOLECULE ANALYSIS EMPLOYING NUCLEIC ACID ENCODING

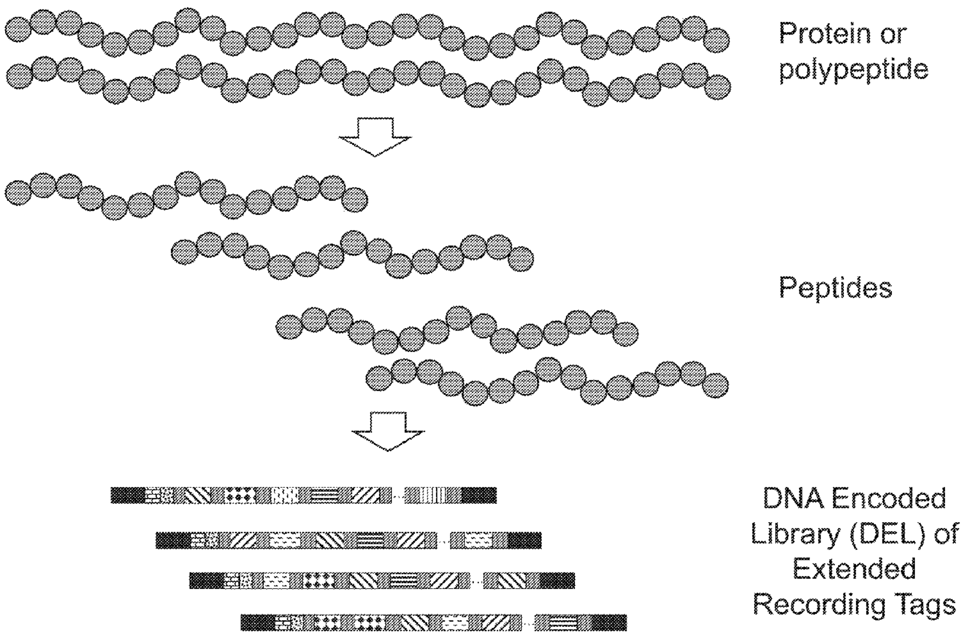


Fig. 1B

(57) **Abstract:** A method for analyzing macromolecules, including peptides, polypeptides, and proteins, employing nucleic acid encoding is disclosed.

[Continued on next page]

WO 2017/192633 A9

EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, 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).

Declarations under Rule 4.17:

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*
— *with information concerning incorporation by reference of missing parts and/or elements (Rule 20.6)*

(48) Date of publication of this corrected version:
14 December 2017 (14.12.2017)

(15) Information about Correction:
see Notice of 14 December 2017 (14.12.2017)

MACROMOLECULE ANALYSIS EMPLOYING NUCLEIC ACID ENCODING

STATEMENT REGARDING SEQUENCE LISTING

The Sequence Listing associated with this application is provided in text format in lieu of a paper copy, and is hereby incorporated by reference into the specification. The name of the text file containing the Sequence Listing is 760229_401WO_SEQUENCE_LISTING.txt. The text file is 38.7 KB, was created on May 2, 2017, and is being submitted electronically via EFS-Web.

BACKGROUND

Technical Field

This disclosure generally relates to analysis of macromolecules, including peptides, polypeptides, and proteins, employing barcoding and nucleic acid encoding of molecular recognition events.

Description of the Related Art

Proteins play an integral role in cell biology and physiology, performing and facilitating many different biological functions. The repertoire of different protein molecules is extensive, much more complex than the transcriptome, due to additional diversity introduced by post-translational modifications (PTMs). Additionally, proteins within a cell dynamically change (in expression level and modification state) in response to the environment, physiological state, and disease state. Thus, proteins contain a vast amount of relevant information that is largely unexplored, especially relative to genomic information. In general, innovation has been lagging in proteomics analysis relative to genomics analysis. In the field of genomics, next-generation sequencing (NGS) has transformed the field by enabling analysis of billions of DNA sequences in a single instrument run, whereas in protein analysis and peptide sequencing, throughput is still limited.

Yet this protein information is direly needed for a better understanding of proteome dynamics in health and disease and to help enable precision medicine. As such, there is great interest in developing “next-generation” tools to miniaturize and highly-parallelize collection of this proteomic information.

Highly-parallel macromolecular characterization and recognition of proteins is challenging for several reasons. The use of affinity-based assays is often difficult due to several key challenges. One significant challenge is multiplexing the readout of a collection of affinity agents to a collection of cognate macromolecules; another challenge is minimizing cross-reactivity between the affinity agents and off-target macromolecules; a third challenge is developing an efficient high-throughput read out platform. An example of this problem occurs in proteomics in which one goal is to identify and quantitate most or all the proteins in a sample. Additionally, it is desirable to characterize various post-translational modifications (PTMs) on the proteins at a single molecule level. Currently this is a formidable task to accomplish in a high-throughput way.

Molecular recognition and characterization of a protein or peptide macromolecule is typically performed using an immunoassay. There are many different immunoassay formats including ELISA, multiplex ELISA (*e.g.*, spotted antibody arrays, liquid particle ELISA arrays), digital ELISA (*e.g.*, Quanterix, Singulex), reverse phase protein arrays (RPPA), and many others. These different immunoassay platforms all face similar challenges including the development of high affinity and highly-specific (or selective) antibodies (binding agents), limited ability to multiplex at both the sample and analyte level, limited sensitivity and dynamic range, and cross-reactivity and background signals. Binding agent agnostic approaches such as direct protein characterization via peptide sequencing (Edman degradation or Mass Spectroscopy) provide useful alternative approaches. However, neither of these approaches is very parallel or high-throughput.

Peptide sequencing based on Edman degradation was first proposed by Pehr Edman in 1950; namely, stepwise degradation of the N-terminal amino acid on a peptide through a series of chemical modifications and downstream HPLC analysis (later replaced by mass spectrometry analysis). In a first step, the N-terminal amino acid is modified with phenyl isothiocyanate (PITC) under mildly basic conditions (NMP/methanol/H_2O) to form a phenylthiocarbamoyl (PTC) derivative. In a second step, the PTC-modified amino group is treated with acid (anhydrous TFA) to create a cleaved cyclic ATZ(2-anilino-5(4)- thiozolinone) modified amino acid, leaving a new N-terminus on the peptide. The cleaved cyclic ATZ-amino acid is converted to a PTH-amino acid derivative and analyzed by reverse phase HPLC. This process is continued in an iterative fashion until all or a partial number of the amino acids comprising a peptide sequence has been removed from the N-terminal end and identified. In general, Edman degradation peptide sequencing is slow and has a limited throughput of only a few peptides per day.

In the last 10-15 years, peptide analysis using MALDI, electrospray mass spectroscopy (MS), and LC-MS/MS has largely replaced Edman degradation. Despite the recent advances in MS instrumentation (Riley et al., 2016, Cell Syst 2:142-143), MS still suffers from several drawbacks including high instrument cost, requirement for a sophisticated user, poor quantification ability, and limited ability to make measurements spanning the dynamic range of the proteome. For example, since proteins ionize at different levels of efficiencies, absolute quantitation and even relative quantitation between sample is challenging. The implementation of mass tags has helped improve relative quantitation, but requires labeling of the proteome. Dynamic range is an additional complication in which concentrations of proteins within a sample can vary over a very large range (over 10 orders for plasma). MS typically only analyzes the more abundant species, making characterization of low abundance proteins challenging. Finally, sample throughput is typically limited to a few thousand peptides per run, and for data independent analysis (DIA), this throughput is inadequate for true bottoms-up high-throughput proteome analysis. Furthermore, there is a significant compute requirement to de-convolute thousands of complex MS spectra recorded for each sample.

Accordingly, there remains a need in the art for improved techniques relating to macromolecule sequencing and/or analysis, with applications to protein sequencing and/or analysis, as well as to products, methods and kits for accomplishing the same. There is a need for proteomics technology that is highly-parallelized, accurate, sensitive, and high-throughput. The present disclosure fulfills these and other needs.

These and other aspects of the invention will be apparent upon reference to the following detailed description. To this end, various references are set forth herein which describe in more detail certain background information, procedures, compounds and/or compositions, and are each hereby incorporated by reference in their entirety.

BRIEF SUMMARY

Embodiments of the present disclosure relate generally to methods of highly-parallel, high throughput digital macromolecule analysis, particularly peptide analysis.

In a first embodiment is a method for analyzing a macromolecule, comprising the steps of:

(a) providing a macromolecule and an associated recording tag joined to a solid support;

(b) contacting the macromolecule with a first binding agent capable of binding to the macromolecule, wherein the first binding agent comprises a first coding tag with identifying information regarding the first binding agent;

(c) transferring the information of the first coding tag to the recording tag to generate a first order extended recording tag;

(d) contacting the macromolecule with a second binding agent capable of binding to the macromolecule, wherein the second binding agent comprises a second coding tag with identifying information regarding the second binding agent;

(e) transferring the information of the second coding tag to the first order extended recording tag to generate a second order extended recording tag; and

(f) analyzing the second order extended recording tag.

In a second embodiment is the method of the first embodiment, wherein contacting steps (b) and (d) are performed in sequential order.

In a third embodiment is the method of the first embodiment, where wherein contacting steps (b) and (d) are performed at the same time.

In a fourth embodiment is the method of the first embodiment, further comprising, between steps (e) and (f), the following steps:

(x) repeating steps (d) and (e) one or more times by replacing the second binding agent with a third (or higher order) binding agent capable of binding to the macromolecule, wherein the third (or higher order) binding agent comprises a third (or higher order) coding tag with identifying information regarding the third (or higher order) bind agent; and

(y) transferring the information of the third (or higher order) coding tag to the second (or higher order) extended recording tag to generate a third (or higher order) extended recording tag;

and wherein the third (or higher order) extended recording tag is analyzed in step (f).

In a fifth embodiment is a method for analyzing a macromolecule, comprising the steps of:

(a) providing a macromolecule, an associated first recording tag and an associated second recording tag joined to a solid support;

(b) contacting the macromolecule with a first binding agent capable of binding to the macromolecule, wherein the first binding agent comprises a first coding tag with identifying information regarding the first binding agent;

(c) transferring the information of the first coding tag to the first recording tag to generate a first extended recording tag;

(d) contacting the macromolecule with a second binding agent capable of binding to the macromolecule, wherein the second binding agent comprises a second coding tag with identifying information regarding the second binding agent;

(e) transferring the information of the second coding tag to the second recording tag to generate a second extended recording tag; and

(f) analyzing the first and second extended recording tags.

In a sixth embodiment is the method of fifth embodiment, wherein contacting steps (b) and (d) are performed in sequential order.

In a seventh embodiment is the method of the fifth embodiment, wherein contacting steps (b) and (d) are performed at the same time.

In an eight embodiment is the method of fifth embodiment, wherein step (a) further comprises providing an associated third (or higher odder) recording tag joined to the solid support.

In a ninth embodiment is the method of the eighth embodiment, further comprising, between steps (e) and (f), the following steps:

(x) repeating steps (d) and (e) one or more times by replacing the second binding agent with a third (or higher order) binding agent capable of binding to the macromolecule, wherein the third (or higher order) binding agent comprises a third (or higher order) coding tag with identifying information regarding the third (or higher order) bind agent; and

(y) transferring the information of the third (or higher order) coding tag to the third (or higher order) recording tag to generate a third (or higher order) extended recording tag;

and wherein the first, second and third (or higher order) extended recording tags are analyzed in step (f).

In a 10^{th} embodiment is the method of any one of the 5th-9^{th} embodiments, wherein the first coding tag, second coding tag, and any higher order coding tags comprise a binding cycle specific spacer sequence.

In an 11^{th} embodiment is a method for analyzing a peptide, comprising the steps of:

(a) providing a peptide and an associated recording tag joined to a solid support;

(b) modifying the N-terminal amino acid (NTAA) of the peptide with a chemical agent;

(c) contacting the peptide with a first binding agent capable of binding to the modified NTAA, wherein the first binding agent comprises a first coding tag with identifying information regarding the first binding agent;

(d) transferring the information of the first coding tag to the recording tag to generate an extended recording tag; and

(e) analyzing the extended recording tag.

In a 12th embodiment is the method of 11th embodiment, wherein step (c) further comprises contacting the peptide with a second (or higher order) binding agent comprising a second (or higher order) coding tag with identifying information regarding the second (or higher order) binding agent, wherein the second (or higher order) binding agent is capable of binding to a modified NTAA other than the modified NTAA of step (b).

In a 13th embodiment is the method of the 12th embodiment, wherein contacting the peptide with the second (or higher order) binding agent occurs in sequential order following the peptide being contacted with the first binding agent.

In a 14th embodiment is the method of 12th embodiment, wherein contacting the peptide with the second (or higher order) binding agent occurs simultaneously with the peptide being contacted with the first binding agent.

In a 15th embodiment is the method of any one the 11th-14th embodiments, wherein the chemical agent is an isothiocyanate derivative, 2,4-dinitrobenzenesulfonic (DNBS), 4-sulfonyl-2-nitrofluorobenzene (SNFB) 1-fluoro-2,4-dinitrobenzene, dansyl chloride, 7-methoxycoumarin acetic acid, a thioacylation reagent, a thioacetylation reagent, or a thiobenzylation reagent.

In a 16th embodiment is a method for analyzing a peptide, comprising the steps of:

(a) providing a peptide and an associated recording tag joined to a solid support;

(b) modifying the N-terminal amino acid (NTAA) of the peptide with a chemical agent to yield a modified NTAA;

(c) contacting the peptide with a first binding agent capable of binding to the modified NTAA, wherein the first binding agent comprises a first coding tag with identifying information regarding the first binding agent;

(d) transferring the information of the first coding tag to the recording tag to generate a first extended recording tag;

(e) removing the modified NTAA to expose a new NTAA;

(f) modifying the new NTAA of the peptide with a chemical agent to yield a newly modified NTAA;

(g) contacting the peptide with a second binding agent capable of binding to the newly modified NTAA, wherein the second binding agent comprises a second coding tag with identifying information regarding the second binding agent;

(h) transferring the information of the second coding tag to the first extended recording tag to generate a second extended recording tag; and

(i) analyzing the second extended recording tag.

In a 17th embodiment is a method for analyzing a peptide, comprising the steps of:

(a) providing a peptide and an associated recording tag joined to a solid support;

(b) contacting the peptide with a first binding agent capable of binding to the N-terminal amino acid (NTAA) of the peptide, wherein the first binding agent comprises a first coding tag with identifying information regarding the first binding agent;

(c) transferring the information of the first coding tag to the recording tag to generate an extended recording tag; and

(d) analyzing the extended recording tag.

In an 18th embodiment is the method of the 17th embodiment, wherein step (b) further comprises contacting the peptide with a second (or higher order) binding agent comprising a second (or higher order) coding tag with identifying information regarding the second (or higher order) binding agent, wherein the second (or higher order) binding agent is capable of binding to a NTAA other than the NTAA of the peptide.

In a 19th embodiment is the method of the 18th embodiment, wherein contacting the peptide with the second (or higher order) binding agent occurs in sequential order following the peptide being contacted with the first binding agent.

In a 20th embodiment is the method of the 18th embodiment, wherein contacting the peptide with the second (or higher order) binding agent occurs simultaneously with the peptide being contacted with the first binding agent.

In a 21st embodiment is a method for analyzing a peptide, comprising the steps of:

(a) providing a peptide and an associated recording tag joined to a solid support;

(b) contacting the peptide with a first binding agent capable of binding to the N-terminal amino acid (NTAA) of the peptide, wherein the first binding agent comprises a first coding tag with identifying information regarding the first binding agent;

(c) transferring the information of the first coding tag to the recording tag to generate a first extended recording tag;

(d) removing the NTAA to expose a new NTAA of the peptide;

(e) contacting the peptide with a second binding agent capable of binding to the new NTAA, wherein the second binding agent comprises a second coding tag with identifying information regarding the second binding agent;

(h) transferring the information of the second coding tag to the first extended recording tag to generate a second extended recording tag; and

(i) analyzing the second extended recording tag.

In a 22nd embodiment is the method of any one of the 1st-10th embodiments, wherein the macromolecule is a protein, polypeptide or peptide.

In a 23rd embodiment is the method of any one of the 1st-10th embodiments, wherein the macromolecule is a peptide.

In a 24th embodiment is the method of any one of the 11th-23rd embodiments, wherein the peptide is obtained by fragmenting a protein from a biological sample.

In a 25th embodiment is the method of any one of the 1st-10th embodiments, wherein the macromolecule is a lipid, a carbohydrate, or a macrocycle.

In a 26th embodiment is the method of any one of the 1st-25th embodiments, wherein the recording tag is a DNA molecule, DNA with pseudo-complementary bases, an RNA molecule, a BNA molecule, an XNA molecule, a LNA molecule, a PNA molecule, a γPNA molecule, or a combination thereof.

In a 27th embodiment is the method of any one of the 1st-26th embodiments, wherein the recording tag comprises a universal priming site.

In a 28th embodiment is the method of the 27th embodiment, wherein the universal priming site comprises a priming site for amplification, sequencing, or both.

In a 29th embodiment is the method of the 1st -28th embodiments, where the recording tag comprises a unique molecule identifier (UMI).

In a 30th embodiment is the method of any one of the 1st-29th embodiments, wherein the recording tag comprises a barcode.

In a 31st embodiment is the method of any one of the 1st-30th embodiments, wherein the recording tag comprises a spacer at its 3'-terminus.

In a 32nd embodiment is the method of claim any one of the 1st-31st embodiments, wherein the macromolecule and the associated recording tag are covalently joined to the solid support.

In a 33rd embodiment is the method of any one of the 1st-32nd embodiments, wherein the solid support is a bead, a porous bead, a porous matrix, an array, a glass surface, a silicon surface, a plastic surface, a filter, a membrane, nylon, a silicon wafer chip, a flow through chip, a biochip including signal transducing electronics, a microtitre well, an ELISA plate, a spinning interferometry disc, a nitrocellulose membrane, a nitrocellulose-based polymer surface, a nanoparticle, or a microsphere.

In a 34th embodiment is the method of the 33rd embodiment, wherein the solid support is a polystyrene bead, a polymer bead, an agarose bead, an acrylamide bead, a solid core bead, a porous bead, a paramagnetic bead, glass bead, or a controlled pore bead.

In a 35th embodiment is the method of any one of the 1st-34th embodiments, wherein a plurality of macromolecules and associated recording tags are joined to a solid support.

In a 36th embodiment is the method of the 35th embodiment, wherein the plurality of macromolecules are spaced apart on the solid support at an average distance > 50 nm.

In a 37th embodiment is the method of any one of 1st-36th embodiments, wherein the binding agent is a polypeptide or protein.

In a 38th embodiment is the method of the 37th embodiment, wherein the binding agent is a modified aminopeptidase, a modified amino acyl tRNA synthetase, a modified anticalin, or a modified ClpS.

In a 39th embodiment is he method of any one of the 1st-38th embodiments, wherein the binding agent is capable of selectively binding to the macromolecule.

In a 40th embodiment is the method of any one of the 1st-39th embodiments, wherein the coding tag is DNA molecule, an RNA molecule, a BNA molecule, an XNA molecule, a LNA molecule, a PNA molecule, a γPNA molecule, or a combination thereof.

In a 41st embodiment is the method of any one of the 1st-40th embodiments, wherein the coding tag comprises an encoder sequence.

In a 42nd embodiment is the method of any one of the 1st-41st embodiments, wherein the coding tag further comprises a spacer, a binding cycle specific sequence, a unique molecular identifier, a universal priming site, or any combination thereof.

In a 43rd embodiment is the method of any one of the 1st-42nd embodiments, wherein the binding agent and the coding tag are joined by a linker.

In a 44th embodiment is the method of any one of the 1st-42nd embodiments, wherein the binding agent and the coding tag are joined by a SpyTag/SpyCatcher or SnoopTag/SnoopCatcher peptide-protein pair.

In a 45th embodiment is the method of any one of the 1st-44th embodiments, wherein transferring the information of the coding tag to the recording tag is mediated by a DNA ligase.

In a 46th embodiment is the method of any one of the 1st-44th embodiments, wherein transferring the information of the coding tag to the recording tag is mediated by a DNA polymerase.

In a 47th embodiment is the method of any one of the 1st-44th embodiments, wherein transferring the information of the coding tag to the recording tag is mediated by chemical ligation.

In a 48th embodiment is the method of any one of claims 1st-47th embodiments, wherein analyzing the extended recording tag comprises a nucleic acid sequencing method.

In a 49th embodiment is the method of the 48th embodiment, wherein the nucleic acid sequencing method is sequencing by synthesis, sequencing by ligation, sequencing by hybridization, polony sequencing, ion semiconductor sequencing, or pyrosequencing.

In a 50th embodiment is the method of the 48th embodiment, wherein the nucleic acid sequencing method is single molecule real-time sequencing, nanopore-based sequencing, or direct imaging of DNA using advanced microscopy.

In a 51st embodiment is the method of any one of the 1st-50th embodiments, wherein the extended recording tag is amplified prior to analysis.

In a 52nd embodiment is the method of any one of the 1st -51st embodiments, wherein the order of coding tag information contained on the extended recording tag provides information regarding the order of binding by the binding agents to the macromolecule.

In a 53rd embodiment is the method of any one of the 1st-52nd embodiments, wherein frequency of the coding tag information contained on the extended recording tag provides information regarding the frequency of binding by the binding agents to the macromolecule.

In a 54th embodiment is the method of any one of the 1st-53rd embodiments, wherein a plurality of extended recording tags representing a plurality of macromolecules are analyzed in parallel.

In a 55th embodiment is the method of the 54th embodiment, wherein the plurality of extended recording tags representing a plurality of macromolecules is analyzed in a multiplexed assay.

In a 56th embodiment is the method of any one of the 1st-55th embodiments, wherein the plurality of extended recording tags undergoes a target enrichment assay prior to analysis.

In a 57th embodiment is the method of any one of the 1st-56th embodiments, wherein the plurality of extended recording tags undergoes a subtraction assay prior to analysis.

In a 58th embodiment is the method of any one of the 1st-57th embodiments, wherein the plurality of extended recording tags undergoes a normalization assay to reduce highly abundant species prior to analysis.

In a 59th embodiment is the method of any one of the 1st-58th embodiments, wherein the NTAA is removed by a modified aminopeptidase, a modified amino acid tRNA synthetase, mild Edman degradation, Edmanase enzyme, or anhydrous TFA.

In a 60th embodiment is the method of any one of the 1st-59th embodiments, wherein at least one binding agent binds to a terminal amino acid residue.

In a 61st embodiment is the method of any one of the 1st-60th embodiments, wherein at least one binding agent binds to a post-translationally modified amino acid.

In a 62nd embodiment is a method for analyzing one or more peptides from a sample comprising a plurality of protein complexes, proteins, or polypeptides, the method comprising:

(a) partitioning the plurality of protein complexes, proteins, or polypeptides within the sample into a plurality of compartments, wherein each compartment comprises a plurality of compartment tags optionally joined to a solid support, wherein the plurality of compartment tags are the same within an individual compartment and are different from the compartment tags of other compartments;

(b) fragmenting the plurality of protein complexes, proteins, and/or polypeptides into a plurality of peptides;

(c) contacting the plurality of peptides to the plurality of compartment tags under conditions sufficient to permit annealing or joining of the plurality of peptides with the plurality of compartment tags within the plurality of compartments, thereby generating a plurality of compartment tagged peptides;

(d) collecting the compartment tagged peptides from the plurality of compartments; and

(e) analyzing one or more compartment tagged peptide according to a method of any one of the 1st-21st embodiments and 26th -61st embodiments.

In a 63rd embodiment is the method of the 62nd embodiment, wherein the compartment is a microfluidic droplet.

In a 64th embodiment is the method of the 62nd embodiment, wherein the compartment is a microwell.

In a 65th embodiment is the method of the 62nd embodiment, wherein the compartment is a separated region on a surface.

In a 66th embodiment is the method of any one of the 62nd-65th embodiments, wherein each compartment comprises on average a single cell.

In a 67th embodiment is a method for analyzing one or more peptides from a sample comprising a plurality of protein complexes, proteins, or polypeptides, the method comprising:

(a) labeling of the plurality of protein complexes, proteins, or polypeptides with a plurality of universal DNA tags;

(b) partitioning the plurality of labeled protein complexes, proteins, or polypeptides within the sample into a plurality of compartments, wherein each compartment comprises a plurality of compartment tags, wherein the plurality of compartment tags are the same within an individual compartment and are different from the compartment tags of other compartments;

(c) contacting the plurality of protein complexes, proteins, or polypeptides to the plurality of compartment tags under conditions sufficient to permit annealing or joining of the plurality of protein complexes, proteins, or polypeptides with the plurality of compartment tags within the plurality of compartments, thereby generating a plurality of compartment tagged protein complexes, proteins or polypeptides;

(d) collecting the compartment tagged protein complexes, proteins, or polypeptides from the plurality of compartments;

(e) optionally fragmenting the compartment tagged protein complexes, proteins, or polypeptides into a compartment tagged peptides; and

(f) analyzing one or more compartment tagged peptide according to a method of any one of the 1st-21st embodiments and 26th -61st embodiments.

In a 68th embodiment is the method of any one of the 62nd-67th embodiments, wherein compartment tag information is transferred to a recording tag associated with a peptide via primer extension or ligation.

In a 69th embodiment is the method of any one of the 62nd-68th embodiments, wherein the solid support comprises a bead.

In a 70th embodiment is the method of the 69th embodiment, wherein the bead is a polystyrene bead, a polymer bead, an agarose bead, an acrylamide bead, a solid core bead, a porous bead, a paramagnetic bead, glass bead, or a controlled pore bead.

In a 71st embodiment is the method of any one of the 62nd-70th embodiments, wherein the compartment tag comprises a single stranded or double stranded nucleic acid molecule.

In a 72nd embodiment is the method of any one of the 62nd-71st embodiments, wherein the compartment tag comprises a barcode and optionally a UMI.

In a 73rd embodiment is the method of the 72nd embodiment, wherein the solid support is a bead and the compartment tag comprises a barcode, further wherein beads comprising the plurality of compartment tags joined thereto are formed by split-and-pool synthesis.

In a 74th embodiment is the method of the 72nd embodiment, wherein the solid support is a bead and the compartment tag comprises a barcode, further wherein beads comprising a plurality of compartment tags joined thereto are formed by individual synthesis or immobilization.

In a 75th embodiment is the method of any one of the 62nd-74th embodiments, wherein the compartment tag is a component within a recording tag, wherein the recording tag optionally further comprises a spacer, a unique molecular identifier, a universal priming site, or any combination thereof.

In a 76th embodiment is the method of any one of the 62nd-75th embodiments, wherein the compartment tags further comprise a functional moiety capable of reacting with an internal amino acid or N-terminal amino acid on the plurality of protein complexes, proteins, or polypeptides.

In a 77th embodiment is the method of the 76th embodiment, wherein the functional moiety is an NHS group.

In a 78th embodiment is the method of the 76th embodiment, wherein the functional moiety is an aldehyde group.

In a 79th embodiment is the method of any one of the 62nd-78th embodiments, wherein the plurality of compartment tags is formed by: printing, spotting, ink-jetting the compartment tags into the compartment, or a combination thereof.

In an 80th embodiment is the method of any one of the 62nd-79th embodiments, wherein the compartment tag further comprises a peptide.

In an 81st embodiment is the method of the 80th embodiment, wherein the compartment tag peptide comprises a protein ligase recognition sequence.

In an 82nd embodiment is the method of the 81st embodiment, wherein the protein ligase is butelase I or a homolog thereof.

In an 83rd. embodiment is the method of any one of the 62nd-82nd embodiments, wherein the plurality of polypeptides is fragmented with a protease.

In an 84th embodiment is the method of the 83rd embodiment, wherein the protease is a metalloprotease.

In an 85th embodiment is the method of the 84th embodiment, wherein the activity of the metalloprotease is modulated by photo-activated release of metallic cations.

In an 86th embodiment is the method of any one of the 62nd-85th embodiments, further comprising subtraction of one or more abundant proteins from the sample prior to partitioning the plurality of polypeptides into the plurality of compartments.

In an 87th embodiment is the method of any one of the 62nd-86th embodiments, further comprising releasing the compartment tags from the solid support prior to joining of the plurality of peptides with the compartment tags.

In an 88th embodiment is the method of the 62nd embodiment, further comprising following step (d), joining the compartment tagged peptides to a solid support in association with recording tags.

In an 89th embodiment is the method of the 88th embodiment, further comprising transferring information of the compartment tag on the compartment tagged peptide to the associated recording tag.

In a 90th embodiment is the method of the 89th embodiment, further comprising removing the compartment tags from the compartment tagged peptides prior to step (e).

In a 91st embodiment is the method of any one of the 62nd-90th embodiments, further comprising determining the identity of the single cell from which the analyzed peptide derived based on the analyzed peptide's compartment tag sequence.

In a 92nd embodiment is the method of any one of the 62nd-90th embodiments, further comprising determining the identity of the protein or protein complex from which the analyzed peptide derived based on the analyzed peptide's compartment tag sequence.

In a 93rd embodiment is a method for analyzing a plurality of macromolecules, comprising the steps of:

(a) providing a plurality macromolecules and associated recording tags joined to a solid support;

(b) contacting the plurality of macromolecules with a plurality of binding agents capable of binding to the plurality of macromolecules, wherein each binding agent comprises a coding tag with identifying information regarding the binding agent;

(c) (i) transferring the information of the macromolecule associated recording tags to the coding tags of the binding agents that are bound to the macromolecules to generate extended coding tags; or (ii) transferring the information of macromolecule associated recording tags and coding tags of the binding agents that are bound to the macromolecules to a di-tag construct;

(d) collecting the extended coding tags or di-tag constructs;

(e) optionally repeating steps (b) – (d) for one or more binding cycles;

(f) analyzing the collection of extended coding tags or di-tag constructs.

In a 94th embodiment is the method of the 93rd embodiment, wherein the macromolecule is a protein.

In a 95th embodiment is the method of the 93rd embodiment, wherein the macromolecule is a peptide.

In a 96th embodiment is the method of the 95th embodiment, wherein the peptide is obtained by fragmenting a protein from a biological sample.

In a 97th embodiment ishe method of any one of the 93rd-96th embodiments, wherein the recording tag is a DNA molecule, an RNA molecule, a PNA molecule, a BNA molecule, an XNA, molecule, an LNA molecule, a γPNA molecule, or a combination thereof.

In a 98th embodiment is the method of any one of the 93rd-97th embodiments, wherein the recording tag comprises a unique molecular identifier (UMI).

In a 99th embodiment is the method of embodiments 93-98, wherein the recording tag comprises a compartment tag.

In a 100th embodiment is the method of any one of embodiments 93-99, wherein the recording tag comprises a universal priming site.

In a 101st embodiment is the method of any one of embodiment 93-100, wherein the recording tag comprises a spacer at its 3'-terminus.

In a 102nd embodiment is the method of any one of embodiment 93-101, wherein the 3'-terminus of the recording tag is blocked to prevent extension of the recording tag by a polymerase and the information of macromolecule associated recording tag and coding tag of the binding agent that is bound to the macromolecule is transferred to a di-tag construct.

In a 103rd embodiment is the method of any one of embodiment 93-102, wherein the coding tag comprises an encoder sequence.

In a 104th embodiment is the method of any one of embodiments 93-103, wherein the coding tag comprises a UMI.

In a 105th embodiment is the method of any one of embodiments 93-104, wherein the coding tag comprises a universal priming site.

In a 106th embodiment is the method of any one of embodiments 93-105, wherein the coding tag comprises a spacer at its 3'-terminus.

In a 107th embodiment is the method of any one of embodiments 93-106, wherein the coding tag comprises a binding cycle specific sequence.

In a 108th embodiment is the method of any one of embodiments 93-107, wherein the binding agent and the coding tag are joined by a linker.

In a 109th embodiment is the method of any one of embodiments 93-108, wherein transferring information of the recording tag to the coding tag is effected by primer extension.

In a 110th embodiment is the method of any one of embodiments 93-108, wherein transferring information of the recording tag to the coding tag is effected by ligation.

In an 111th embodiment is the method of any one of embodiments 93-108, wherein the di-tag construct is generated by gap fill, primer extension, or both.

In a 112th embodiment is the method of any one of embodiments 93-97, 107, 108, and 111, wherein the di-tag molecule comprises a universal priming site derived from the recording tag, a compartment tag derived from the recording tag, a unique molecular identifier derived from the recording tag, an optional spacer derived from the recording tag, an encoder sequence derived from the coding tag, a unique molecular identifier derived from the coding tag, an optional spacer derived from the coding tag, and a universal priming site derived from the coding tag.

In a 113th embodiment is the method of any one of embodiments 93-112, wherein the macromolecule and the associated recording tag are covalently joined to the solid support.

In a 114th embodiment is the method of embodiment 113, wherein the solid support is a bead, a porous bead, a porous matrix, an array, a glass surface, a silicon surface, a plastic surface, a filter, a membrane, nylon, a silicon wafer chip, a flow through chip, a biochip including signal transducing electronics, a microtitre well, an ELISA plate, a spinning interferometry disc, a nitrocellulose membrane, a nitrocellulose-based polymer surface, a nanoparticle, or a microsphere.

In a 115th embodiment is the method of embodiment 114, wherein the solid support is a polystyrene bead, a polymer bead, an agarose bead, an acrylamide bead, a solid core bead, a porous bead, a paramagnetic bead, glass bead, or a controlled pore bead.

In a 116th embodiment is the method of any one of embodiments 93-115, wherein the binding agent is a polypeptide or protein.

In a 117th embodiment is the method of embodiment 116, wherein the binding agent is a modified aminopeptidase, a modified amino acyl tRNA synthetase, a modified anticalin, or an antibody or binding fragment thereof.

In an 118th embodiment is the method of any one of embodiment 95-117 wherein the binding agent binds to a single amino acid residue, a dipeptide, a tripeptide or a post-translational modification of the peptide.

In a 119th embodiment is the method of embodiment 118, wherein the binding agent binds to an N-terminal amino acid residue, a C-terminal amino acid residue, or an internal amino acid residue.

In a 120th embodiment is the method of embodiment 118, wherein the binding agent binds to an N-terminal peptide, a C-terminal peptide, or an internal peptide.

In a 121st embodiment is method of embodiment 119, wherein the binding agent binds to the N-terminal amino acid residue and the N-terminal amino acid residue is cleaved after each binding cycle.

In a 122nd embodiment is the method of embodiment 119, wherein the binding agent binds to the C-terminal amino acid residue and the C-terminal amino acid residue is cleaved after each binding cycle.

Embodiment 123. The method of embodiment 121, wherein the N-terminal amino acid residue is cleaved via Edman degradation.

embodiment 124. The method of embodiment 93, wherein the binding agent is a site-specific covalent label of an amino acid or post-translational modification.

Embodiment 125. The method of any one of embodiment 93-124, wherein following step (b), complexes comprising the macromolecule and associated binding agents are dissociated from the solid support and partitioned into an emulsion of droplets or microfluidic droplets.

Embodiment 126. The method of embodiment 125, wherein each microfluidic droplet, on average, comprises one complex comprising the macromolecule and the binding agents.

Embodiment 127. The method of embodiment 125 or 126, wherein the recording tag is amplified prior to generating an extended coding tag or di-tag construct.

Embodiment 128. The method of any one of embodiments 125-127, wherein emulsion fusion PCR is used to transfer the recording tag information to the coding tag or to create a population of di-tag constructs.

Embodiment 129. The method of any one of embodiments 93-128, wherein the collection of extended coding tags or di-tag constructs are amplified prior to analysis.

Embodiment 130. The method of any one of embodiments 93-129, wherein analyzing the collection of extended coding tags or di-tag constructs comprises a nucleic acid sequencing method.

Embodiment 131. The method of embodiment 130, wherein the nucleic acid sequencing method is sequencing by synthesis, sequencing by ligation, sequencing by hybridization, polony sequencing, ion semiconductor sequencing, or pyrosequencing.

Embodiment 132. The method of embodiment 130, wherein the nucleic acid sequencing method is single molecule real-time sequencing, nanopore-based sequencing, or direct imaging of DNA using advanced microscopy.

Embodiment 133. The method of embodiment 130, wherein a partial composition of the macromolecule is determined by analysis of a plurality of extended coding tags or di-tag constructs using unique compartment tags and optionally UMIs.

Embodiment 134. The method of any one of embodiments 1-133, wherein the analysis step is performed with a sequencing method having a per base error rate of > 5%, > 10%, > 15%, > 20%, > 25%, or > 30%.

Embodiment 135. The method of any one of embodiments 1-134, wherein the identifying components of a coding tag, recording tag, or both comprise error correcting codes.

Embodiment 136. The method of embodiment 135, wherein the identifying components are selected from an encoder sequence, barcode, UMI, compartment tag, cycle specific sequence, or any combination thereof.

Embodiment 137. The method of embodiment 135 or 136, wherein the error correcting code is selected from Hamming code, Lee distance code, asymmetric Lee distance code, Reed-Solomon code, and Levenshtein-Tenengolts code.

Embodiment 138. The method of any one of embodiments1-134, wherein the identifying components of a coding tag, recording tag, or both are capable of generating a unique current or ionic flux or optical signature, wherein the analysis step comprises detection of the unique current or ionic flux or optical signature in order to identify the identifying components.

Embodiment 139. The method of embodiment 138, wherein the identifying components are selected from an encoder sequence, barcode, UMI, compartment tag, cycle specific sequence, or any combination thereof.

Embodiment 140. A method for analyzing a plurality of macromolecules, comprising the steps of:

(a) providing a plurality macromolecules and associated recording tags joined to a solid support;

(b) contacting the plurality of macromolecules with a plurality of binding agents capable of binding to cognate macromolecules, wherein each binding agent comprises a coding tag with identifying information regarding the binding agent;

(c) transferring the information of a first coding tag of a first binding agent to a first recording tag associated with the first macromolecule to generate a first order extended recording tag, wherein the first binding agent binds to the first macromolecule,;

(d) contacting the plurality of macromolecules with the plurality of binding agents capable of binding to cognate macromolecules;

(e) transferring the information of a second coding tag of a second binding agent to the first order extended recording tag to generate a second order extended recording tag, wherein the second binding agent binds to the first macromolecule;

(f) optionally repeating steps (d) – (e) for "n" binding cycles, wherein the information of each coding tag of each binding agent that binds to the first macromolecule is transferred to the extended recording tag generated from the previous binding cycle to generate an n^{th} order extended recording tag that represents the first macromolecule;

(g) analyzing the n^{th} order extended recording tag.

Embodiment 141. The method of embodiment 140, wherein a plurality of n^{th} order extended recording tags that represent a plurality of macromolecules are generated and analyzed.

Embodiment 142. The method of embodiment 140 or 141, wherein the macromolecule is a protein.

Embodiment 143. The method of embodiment 142, wherein the macromolecule is a peptide.

Embodiment 144. The method of embodiment 143, wherein the peptide is obtained by fragmenting proteins from a biological sample.

Embodiment 145. The method of any one of embodiments 140-144, wherein the plurality of macromolecules comprises macromolecules from multiple, pooled samples.

Embodiment 146. The method of any one of embodiments 140-145, wherein the recording tag is a DNA molecule, an RNA molecule, a PNA molecule, a BNA molecule, an XNA, molecule, an LNA molecule, a γPNA molecule, or a combination thereof.

Embodiment 147. The method of any one of embodiments 140-146, wherein the recording tag comprises a unique molecular identifier (UMI).

Embodiment 148. The method of embodiments 140-147, wherein the recording tag comprises a compartment tag.

Embodiment 149. The method of any one of embodiments 140-148, wherein the recording tag comprises a universal priming site.

Embodiment 150. The method of any one of embodiments 140-149, wherein the recording tag comprises a spacer at its 3'-terminus.

Embodiment 151. The method of any one of embodiments 140-150, wherein the coding tag comprises an encoder sequence.

Embodiment 152. The method of any one of embodiments 140-151, wherein the coding tag comprises a UMI.

Embodiment 153. The method of any one of embodiments 140-152, wherein the coding tag comprises a universal priming site.

Embodiment 154. The method of any one of embodiments 140-153, wherein the coding tag comprises a spacer at its 3'-terminus.

Embodiment 155. The method of any one of embodiments 140-154, wherein the coding tag comprises a binding cycle specific sequence.

Embodiment 156. The method of any one of embodiments 140-155, wherein the coding tag comprises a unique molecular identifier.

Embodiment 157. The method of any one of embodiments 140-156, wherein the binding agent and the coding tag are joined by a linker.

Embodiment 158. The method of any one of embodiments 140-157, wherein transferring information of the recording tag to the coding tag is mediated by primer extension.

Embodiment 159. The method of any one of embodiments 140-158, wherein transferring information of the recording tag to the coding tag is mediated by ligation.

Embodiment 160. The method of any one of embodiments 140-159, wherein the plurality of macromolecules, the associated recording tags, or both are covalently joined to the solid support.

Embodiment 161. The method of any one of embodiments 140-160, wherein the solid support is a bead, a porous bead, a porous matrix, an array, a glass surface, a silicon surface, a plastic surface, a filter, a membrane, nylon, a silicon wafer chip, a flow through chip, a biochip including signal transducing electronics, a microtitre well, an ELISA plate, a spinning interferometry disc, a nitrocellulose membrane, a nitrocellulose-based polymer surface, a nanoparticle, or a microsphere.

Embodiment 162. The method of embodiment 161, wherein the solid support is a polystyrene bead, a polymer bead, an agarose bead, an acrylamide bead, a solid core bead, a porous bead, a paramagnetic bead, glass bead, or a controlled pore bead.

Embodiment 163. The method of any one of embodiments 140-162, wherein the binding agent is a polypeptide or protein.

Embodiment 164. The method of embodiment 163, wherein the binding agent is a modified aminopeptidase, a modified amino acyl tRNA synthetase, a modified anticalin, or an antibody or binding fragment thereof.

Embodiment 165. The method of any one of embodiments 142-164 wherein the binding agent binds to a single amino acid residue, a dipeptide, a tripeptide or a post-translational modification of the peptide.

Embodiment 166. The method of embodiment 165, wherein the binding agent binds to an N-terminal amino acid residue, a C-terminal amino acid residue, or an internal amino acid residue.

Embodiment 167. The method of embodiment 165, wherein the binding agent binds to an N-terminal peptide, a C-terminal peptide, or an internal peptide.

Embodiment 168. The method of any one of embodiments 142-164, wherein the binding agent binds to a chemical label of a modified N-terminal amino acid residue, a modified C-terminal amino acid residue, or a modified internal amino acid residue.

Embodiment 169. The method of embodiment 166 or 168, wherein the binding agent binds to the N-terminal amino acid residue or the chemical label of the modified N-terminal amino acid residue, and the N-terminal amino acid residue is cleaved after each binding cycle.

Embodiment 170. The method of embodiment 166 or 168, wherein the binding agent binds to the C-terminal amino acid residue or the chemical label of the modified C-terminal amino acid residue, and the C-terminal amino acid residue is cleaved after each binding cycle.

Embodiment 171. The method of embodiment 169, wherein the N-terminal amino acid residue is cleaved via Edman degradation, Edmanase, a modified amino peptidase, or a modified acylpeptide hydrolase.

Embodiment 172. The method of embodiment 163, wherein the binding agent is a site-specific covalent label of an amino acid or post-translational modification.

Embodiment 173. The method of any one of embodiments 140-172, wherein the plurality of n^{th} order extended recording tags are amplified prior to analysis.

Embodiment 174. The method of any one of embodiments 140-173, wherein analyzing the n^{th} order extended recording tag comprises a nucleic acid sequencing method.

Embodiment 175. The method of embodiment 174, wherein a plurality of n^{th} order extended recording tags representing a plurality of macromolecules are analyzed in parallel.

Embodiment 176. The method of embodiment 174 or 175, wherein the nucleic acid sequencing method is sequencing by synthesis, sequencing by ligation, sequencing by hybridization, polony sequencing, ion semiconductor sequencing, or pyrosequencing.

Embodiment 177. The method of embodiment 174 or 175, wherein the nucleic acid sequencing method is single molecule real-time sequencing, nanopore-based sequencing, or direct imaging of DNA using advanced microscopy.

BRIEF DESCRIPTION OF THE FIGURES

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. For purposes of illustration, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention.

Figures 1A-B: Figure 1A illustrates key for functional elements shown in the figures. **Figure 1B** illustrates a general overview of transducing protein code to a DNA code where a plurality of proteins or polypeptides are fragmented into a plurality of peptides, which are then converted into a library of extended recording tags, representing the plurality of peptides. The extended recording tags constitute a DNA Encoded Library representing the peptide sequences. The library can be appropriately modified to sequence on any Next Generation Sequencing (NGS) platform.

Figures 2A-2D illustrate an example of protein macromolecule analysis according to the methods disclosed herein, using multiple cycles of binding agents (*e.g.*, antibodies, anticalins, N-recognins proteins (e.g., ATP-dependent Clp protease adaptor protein (ClpS)), aptamers, etc. and variants/homologues thereof) comprising coding tags interacting with an immobilized protein that is co-localized or co-labeled with a single or multiple recording tags. The recording tag is comprised of a universal priming site, a barcode (e.g., partition barcode, compartment barcode, fraction barcode), an optional unique molecular identifier (UMI) sequence, and a spacer sequence (Sp) used in information transfer of the coding tag. The spacer sequence (Sp) can be constant across all binding cycles, be binding agent specific, or be binding cycle number specific. The coding tag is comprised of an encoder sequence providing identifying information for the binding agent, an optional UMI, and a spacer sequence that hybridizes to the complementary spacer sequence on the recording tag, facilitating transfer of coding tag information to the recording tag (e.g., primer extension, also referred to herein as polymerase extension). **Figure 2A** illustrates a process of creating an extended recording tag through the cyclic binding of cognate binding agents to a protein, and corresponding information transfer from the binding agent's coding tag to the protein's recording tag. After a series of sequential binding and coding tag information transfer steps, the final extended recording tag is produced, containing binding agent coding tag information including encoder sequences from "n" binding cycles providing identifying information for the binding agents (e.g., antibody 1 (Ab1), antibody 2 (Ab2), antibody 3 (Ab3),…antibody "n" (Abn)), a barcode/optional UMI sequence from the recording tag, an optional UMI sequence from the binding agent's coding tag, and flanking universal priming sequences at each end of the library construct to facilitate amplification and analysis by digital next-generation sequencing. **Figure 2B** illustrates an example of a scheme for labeling a protein with DNA barcoded recording tags. In the top panel, N-hydroxysuccinimide (NHS) is an amine reactive coupling agent, and Dibenzocyclooctyl (DBCO) is a strained alkyne useful in “click” coupling to the surface of a solid substrate. In this scheme, the recording tags are coupled to ε amines of lysine (K) residues (and optionally N-terminal amino acids) of the protein via NHS moieties. In the bottom panel, a heterobifunctional linker, NHS-alkyne, is used to label the ε amines of lysine (K) residues to create an alkyne “click” moiety. Azide-labeled DNA recording tags can then easily be attached to these reactive alkyne groups via standard click chemistry. Moreover, the DNA recording tag can also be designed with an orthogonal methyltetrazine (mTet) moiety for downstream coupling to a TCO-derivatized sequencing substrate via an inverse iEDDA reaction. **Figure 2C** illustrates two examples of the protein analysis methods using recording tags. In the top panel, protein macromolecules are immobilized on a solid support via a capture agent and optionally cross-linked. Either the protein or capture agent may be labeled with a recording tag. In the bottom panel, proteins with associated recording tags are directly immobilized on a solid support. **Figure 2D** illustrates an example of an overall workflow for a simple protein immunoassay using DNA encoding of cognate binders and sequencing of the resultant extended recording tag. The proteins can be sample barcoded (*i.e.*, indexed) via recording tags and pooled prior to cyclic binding analysis, greatly increasing sample throughput and economizing on binding reagents. This approach is effectively a digital, simpler, and more scalable approach to performing reverse phase protein assays (RPPA).

Figures 3A-D illustrate a process for a degradation-based peptide sequencing assay by construction of a DNA extended recording tag representing the peptide sequence. This is accomplished through an Edman degradation-like approach using a cyclic process of N-terminal amino acid (NTAA) binding, coding tag information transfer to a recording tag attached to the peptide, NTAA cleavage, and repeating the process in a cyclic manner, all on a solid support. Provided is an overview of an exemplary construction of an extended recording tag from N-terminal degradation of a peptide: **(A)** N-terminal amino acid of a peptide is labeled (*e.g.*, with a phenylthiocarbamoyl (PTC), dinitrophenyl (DNP), sulfonyl nitrophenyl (SNP), acetyl, or guanidindyl moiety); **(B)** shows a binding agent and an associated coding tag bound to the labeled NTAA; **(C)** shows the peptide bound to a solid support (*e.g.*, bead) and associated with a recording tag (*e.g.*, via a trifunctional linker), wherein upon binding of the binding agent to the NTAA of the peptide, information of the coding tag is transferred to the recording tag (*e.g.*, via primer extension) to generate an extended recording tag; **(D)** the labeled NTAA is cleaved via chemical or enzymatic means to expose a new NTAA. As illustrated by the arrows, the cycle is repeated "n" times to generate a final extended recording tag. The final extended recording tag is optionally flanked by universal priming sites to facilitate downstream amplification and DNA sequencing. The forward universal priming site (e.g., Illumina's P5-S1 sequence) can be part of the original recording tag design and the reverse universal priming site (e.g., Illumina's P7-S2' sequence) can be added as a final step in the extension of the recording tag. This final step may be done independently of a binding agent.

Figures 4A-B illustrate exemplary protein sequencing workflows according to the methods disclosed herein. **Figure 4A** illustrates exemplary work flows with alternative modes outlined in light grey dashed lines, with a particular embodiment shown in boxes linked by arrows. Alternative modes for each step of the workflow are shown in boxes below the arrows. **Figure 4B** illustrates options in conducting a cyclic binding and coding tag information transfer step to improve the efficiency of information transfer. Multiple recording tags per molecule can be employed. Moreover, for a given binding event, the transfer of coding tag information to the recording tag can be conducted multiples times, or alternatively, a surface amplification step can be employed to create copies of the extended recording tag library, etc.

Figures 5A-B illustrate an overview of an exemplary construction of an extended recording tag using primer extension to transfer identifying information of a coding tag of a binding agent to a recording tag associated with a macromolecule (e.g., peptide) to generate an extended recording tag. A coding tag comprising a unique encoder sequence with identifying information regarding the binding agent is optionally flanked on each end by a common spacer sequence (Sp'). **Figure 5A** illustrates an NTAA binding agent comprising a coding tag binding to an NTAA of a recording-tag labeled peptide linked to a bead. The recording tag anneals to the coding tag via complementary spacer sequence (Sp), and a primer extension reaction mediates transfer of coding tag information to the recording tag using the spacer (Sp) as a priming site. The coding tag is illustrated as a duplex with a single stranded spacer (Sp') sequence at the terminus distal to the binding agent. This configuration minimizes hybridization of the coding tag to internal sites in the recording tag and favors hybridization of the recording tag's terminal spacer (Sp) sequence with the single stranded spacer overhang (Sp') of the coding tag. Moreover, the extended recording tag may be pre-annealed with oligonucleotides (complementary to encoder, spacer sequences) to block hybridization of the coding tag to internal recording tag sequence elements. **Figure 5B** shows a final extended recording tag produced after "n" cycles of binding ("***" represents intervening binding cycles not shown in the extended recording tag) and transfer of coding tag information and the addition of a universal priming site at the 3'-end.

Figure 6 illustrates coding tag information being transferred to an extended recording tag via enzymatic ligation. Two different macromolecules are shown with their respective recording tags, with recording tag extension proceeding in parallel. Ligation can be facilitated by designing the double stranded coding tags so that the spacer sequences (Sp) have a "sticky end" overhang that anneals with a complementary spacer (Sp') on the recording tag. The complementary strand of a double stranded coding tag transfers information to the recording tag. When ligation is used to extend the recording tag, the direction of extension can be 5' to 3' as illustrated, or optionally 3' to 5'.

Figure 7 illustrates a "spacer-less" approach of transferring coding tag information to a recording tag via chemical ligation to link the 3' nucleotide of a recording tag or extended recording tag to the 5' nucleotide of the coding tag (or its complement) without inserting a spacer sequence into the extended recording tag. The orientation of the extended recording tag and coding tag could also be inverted such that the 5' end of the recording tag is ligated to the 3' end of the coding tag (or complement). In the example shown, hybridization between complementary "helper" oligonucleotide sequences on the recording tag ("recording helper") and the coding tag are used to stabilize the complex to enable specific chemical ligation of the recording tag to coding tag complementary strand. The resulting extended recording tag is devoid of spacer sequences. Also illustrated is a "click chemistry" version of chemical ligation (e.g., using azide and alkyne moieties (shown as a triple line symbol)) which can employ DNA, PNA, or similar nucleic acid polymers.

Figures 8A-B illustrate an exemplary method of writing of post-translational modification (PTM) information of a peptide into an extended recording tag prior to N-terminal amino acid degradation. **Figure 8A**: A binding agent comprising a coding tag with identifying information regarding the binding agent (*e.g.*, a phosphotyrosine antibody comprising a coding tag with identifying information for phosphotyrosine antibody) is capable of binding to the peptide. If phosphotyrosine is present in the recording tag-labeled peptide, as illustrated, upon binding of the phosphotyrosine antibody to phosphotyrosine, the coding tag and recording tag anneal via complementary spacer sequences and the coding tag information is transferred to the recording tag to generate an extended recording tag. **Figure 8B:** An extended recording tag may comprise coding tag information for both primary amino acid sequence (*e.g.*, "aa_1", "aa_2", "aa_3",..., "aa_N") and post-translational modifications (*e.g.*, "PTM_1", "PTM_2") of the peptide.

Figures 9A-B illustrate a process of multiple cycles of binding of a binding agent to a macromolecule and transferring information of a coding tag that is attached to a binding agent to an individual recording tag among a plurality of recording tags co-localized at a site of a single macromolecule attached to a solid support (*e.g.*, a bead), thereby generating multiple extended recording tags that collectively represent the macromolecule. In this figure, for purposes of example only, the macromolecule is a peptide and each cycle involves binding a binding agent to an N-terminal amino acid (NTAA), recording the binding event by transferring coding tag information to a recording tag, followed by removal of the NTAA to expose a new NTAA. **Figure 9A** illustrates a plurality of recording tags (comprising universal forward priming sequence and a UMI) co-localized on a solid support with the macromolecule. Individual recording tags possess a common spacer sequence (Sp) complementary to a common spacer sequence within coding tags of binding agents, which can be used to prime an extension reaction to transfer coding tag information to a recording tag. **Figure 9B** illustrates different pools of cycle-specific NTAA binding agents that are used for each successive cycle of binding, each pool having cycle specific spacer sequences.

Figures 10A-C illustrate an exemplary mode comprising multiple cycles of transferring information of a coding tag that is attached to a binding agent to a recording tag among a plurality of recording tags co-localized at a site of a single macromolecule attached to a solid support (*e.g.*, a bead), thereby generating multiple extended recording tags that collectively represent the macromolecule. In this figure, for purposes of example only, the macromolecule is a peptide and each round of processing involves binding to an NTAA, recording the binding event, followed by removal of the NTAA to expose a new NTAA. **Figure 10A** illustrates a plurality of recording tags (comprising a universal forward priming sequence and a UMI) co-localized on a solid support with the macromolecule, preferably a single molecule per bead. Individual recording tags possess different spacer sequences at their 3'-end with different "cycle specific" sequences (*e.g.*, C_1, C_2, C_3,...C_n). Preferably, the recording tags on each bead share the same UMI sequence. In a first cycle of binding (Cycle 1), a plurality of NTAA binding agents is contacted with the macromolecule. The binding agents used in Cycle 1 possess a common 5'-spacer sequence (C'1) that is complementary to the Cycle 1 C_1 spacer sequence of the recording tag. The binding agents used in Cycle 1 also possess a 3'-spacer sequence (C'_2) that is complementary to the Cycle 2 spacer C_2. During binding Cycle 1, a first NTAA binding agent binds to the free N-terminus of the macromolecule, and the information of a first coding tag is transferred to a cognate recording tag via primer extension from the C_1 sequence hybridized to the complementary C'_1 spacer sequence. Following removal of the NTAA to expose a new NTAA, binding Cycle 2 contacts a plurality of NTAA binding agents that possess a Cycle 2 5'-spacer sequence (C'_2) that is identical to the 3'-spacer sequence of the Cycle 1 binding agents and a common Cycle 3 3'-spacer sequence (C'_3), with the macromolecule. A second NTAA binding agent binds to the NTAA of the macromolecule, and the information of a second coding tag is transferred to a cognate recording tag via primer extension from the complementary C_2 and C'_2 spacer sequences. These cycles are repeated up to "n" binding cycles, wherein the last extended recording tag is capped with a universal reverse priming sequence, generating a plurality of extended recording tags co-localized with the single macromolecule, wherein each extended recording tag possesses coding tag information from one binding cycle. Because each set of binding agents used in each successive binding cycle possess cycle specific spacer sequences in the coding tags, binding cycle information can be associated with binding agent information in the resulting extended recording tags. **Figure 10B** illustrates different pools of cycle-specific binding agents that are used for each successive cycle of binding, each pool having cycle specific spacer sequences. **Figure 10C** illustrates how the collection of extended recording tags that are co-localized at the site of the macromolecule can be assembled in a sequential order based on PCR assembly of the extended recording tags using cycle specific spacer sequences, thereby providing an ordered sequence of the macromolecule. In a preferred mode, multiple copies of each extended recording tag are generated via amplification prior to concatenation.

Figures 11A-B illustrate information transfer from recording tag to a coding tag or di-tag construct. Two methods of recording binding information are illustrated in (A) and (B). A binding agent may be any type of binding agent as described herein; an anti-phosphotyrosine binding agent is shown for illustration purposes only. For extended coding tag or di-tag construction, rather than transferring binding information from the coding tag to the recording tag, information is either transferred from the recording tag to the coding tag to generate an extended coding tag (A), or information is transferred from both the recording tag and coding tag to a third di-tag-forming construct (B). The di-tag and extended coding tag comprise the information of the recording tag (containing a barcode, an optional UMI sequence, and an optional compartment tag (CT) sequence (not illustrated)) and the coding tag. The di-tag and extended coding tag can be eluted from the recording tag, collected, and optionally amplified and read out on a next generation sequencer.

Figures 12A-D illustrate design of PNA combinatorial barcode/UMI recording tag and di-tag detection of binding events. In **Figure 12A**, the construction of a combinatorial PNA barcode/UMI via chemical ligation of four elementary PNA word sequences (A, A'-B, B'-C, and C') is illustrated. Hybridizing DNA arms are included to create a spacer-less combinatorial template for combinatorial assembly of a PNA barcode/UMI. Chemical ligation is used to stitch the annealed PNA "words" together. **Figure 12B** shows a method to transfer the PNA information of the recording tag to a DNA intermediate. The DNA intermediate is capable of transferring information to the coding tag. Namely, complementary DNA word sequences are annealed to the PNA and chemically ligated (optionally enzymatically ligated if a ligase is discovered that uses a PNA template). In **Figure 12C**, the DNA intermediate is designed to interact with the coding tag via a spacer sequence, Sp. A strand-displacing primer extension step displaces the ligated DNA and transfers the recording tag information from the DNA intermediate to the coding tag to generate an extended coding tag. A terminator nucleotide may be incorporated into the end of the DNA intermediate to prevent transfer of coding tag information to the DNA intermediate via primer extension. **Figure 12D:** Alternatively, information can be transferred from coding tag to the DNA intermediate to generate a di-tag construct. A terminator nucleotide may be incorporated into the end of the coding tag to prevent transfer of recording tag information from the DNA intermediate to the coding tag.

Figures 13A-E illustrate proteome partitioning on a compartment barcoded bead, and subsequent di-tag assembly via emulsion fusion PCR to generate a library of elements representing peptide sequence composition. The amino acid content of the peptide can be subsequently characterized through N-terminal sequencing or alternatively through attachment (covalent or non-covalent) of amino acid specific chemical labels or binding agents associated with a coding tag. The coding tag is comprised of universal priming sequence, as well as an encoder sequence for the amino acid identity, a compartment tag, and an amino acid UMI. After information transfer, the ditags are mapped back to the originating molecule via the recording tag UMI. In **Figure 13A**, the proteome is compartmentalized into droplets with barcoded beads. Peptides with associated recording tags (comprising compartment barcode information) are attached to the bead surface. The droplet emulsion is broken releasing barcoded beads with partitioned peptides. In **Figure 13B**, specific amino acid residues on the peptides are chemically labeled with DNA coding tags that are conjugated to site-specific labeling moieties. The DNA coding tags comprise amino acid barcode information and optionally an amino acid UMI. **Figure 13C**: Labeled peptide-recording tag complexes are released from the beads. **Figure 13D**: The labeled peptide-recording tag complexes are emulsified into nano or microemulsions such that there is, on average, less than one peptide-recording tag complex per compartment. **Figure 13E**: An emulsion fusion PCR transfers recording tag information (e.g., compartment barcode) to all of the DNA coding tags attached to the amino acid residues.

Figure 14 illustrates generation of extended coding tags from emulsified peptide recording tag - coding tags complex. The peptide complexes from **Figure 13C** are co-emulsified with PCR reagents into droplets with on average a single peptide complex per droplet. A three-primer fusion PCR approach is used to amplify the recording tag associated with the peptide, fuse the amplified recording tags to multiple binding agent coding tags or coding tags of covalently labeled amino acids, extend the coding tags via primer extension to transfer peptide UMI and compartment tag information from the recording tag to the coding tag, and amplify the resultant extended coding tags. There are multiple extended coding tag species per droplet, with a different species for each amino acid encoder sequence-UMI coding tag present. In this way, both the identity and count of amino acids within the peptide can be determined. The U1 universal primer and Sp primer are designed to have a higher melting Tm than the $U2_{tr}$ universal primer. This enables a two-step PCR in which the first few cycles are performed at a higher annealing temperature to amplify the recording tag, and then stepped to a lower Tm so that the recording tags and coding tags prime on each other during PCR to produce an extended coding tag, and the U1 and $U2_{tr}$ universal primers are used to prime amplification of the resultant extended coding tag product. In certain embodiments, premature polymerase extension from the $U2_{tr}$ primer can be prevented by using a photo-labile 3' blocking group (Young et al., 2008, Chem. Commun. (Camb) 4:462-464). After the first round of PCR amplifying the recording tags, and a second-round fusion PCR step in which the coding tag Sp_{tr} primes extension of the coding tag on the amplified Sp' sequences of the recording tag, the 3' blocking group of $U2_{tr}$ is removed, and a higher temperature PCR is initiated for amplifying the extended coding tags with U1 and $U2_{tr}$ primers.

Figure 15 illustrates use of proteome partitioning and barcoding facilitating enhanced mappability and phasing of proteins. In peptide sequencing, proteins are typically digested into peptides. In this process, information about the relationship between individual peptides that originated from a parent protein molecule, and their relationship to the parent protein molecule is lost. In order to reconstruct this information, individual peptide sequences are mapped back to a collection of protein sequences from which they may have derived. The task of finding a unique match in such a set is rendered more difficult with short and/or partial peptide sequences, and as the size and complexity of the collection (e.g., proteome sequence complexity) increases. The partitioning of the proteome into barcoded (*e.g.*, compartment tagged) compartments or partitions, subsequent digestion of the protein into peptides, and the joining of the compartment tags to the peptides reduces the "protein" space to which a peptide sequence needs to be mapped to, greatly simplifying the task in the case of complex protein samples. Labeling of a protein with unique molecular identifier (UMI) prior to digestion into peptides facilitates mapping of peptides back to the originating protein molecule and allows annotation of phasing information between post-translational modified (PTM) variants derived from the same protein molecule and identification of individual proteoforms. **Figure 15A** shows an example of proteome partitioning comprising labeling proteins with recording tags comprising a partition barcode and subsequent fragmentation into recording-tag labeled peptides. **Figure 15B:** For partial peptide sequence information or even just composition information, this mapping is highly-degenerate. However, partial peptide sequence or composition information coupled with information from multiple peptides from the same protein, allow unique identification of the originating protein molecule.

Figure 16 illustrates exemplary modes of compartment tagged bead sequence design. The compartment tags comprise a barcode of $X_{5\text{-}20}$ to identify an individual compartment and a unique molecular identifier (UMI) of $N_{5\text{-}10}$ to identify the peptide to which the compartment tag is joined, where X and N represent degenerate nucleobases or nucleobase words. Compartment tags can be single stranded (upper depictions) or double stranded (lower depictions). Optionally, compartment tags can be a chimeric molecule comprising a peptide sequence with a recognition sequence for a protein ligase (*e.g.*, butelase I) for joining to a peptide of interest (left depictions). Alternatively, a chemical moiety can be included on the compartment tag for coupling to a peptide of interest (e.g., azide as shown in right depictions).

Figures 17A-B illustrate: **(A)** a plurality of extended recording tags representing a plurality of peptides; and **(B)** an exemplary method of target peptide enrichment via standard hybrid capture techniques. For example, hybrid capture enrichment may use one or more biotinylated "bait" oligonucleotides that hybridize to extended recording tags representing one or more peptides of interest ("target peptides") from a library of extended recording tags representing a library of peptides. The bait oligonucleotide:target extended recording tag hybridization pairs are pulled down from solution via the biotin tag after hybridization to generate an enriched fraction of extended recording tags representing the peptide or peptides of interest. The separation ("pull down") of extended recording tags can be accomplished, for example, using streptavidin-coated magnetic beads. The biotin moieties bind to streptavidin on the beads, and separation is accomplished by localizing the beads using a magnet while solution is removed or exchanged. A non-biotinylated competitor enrichment oligonucleotide that competitively hybridizes to extended recording tags representing undesirable or over-abundant peptides can optionally be included in the hybridization step of a hybrid capture assay to modulate the amount of the enriched target peptide. The non-biotinylated competitor oligonucleotide competes for hybridization to the target peptide, but the hybridization duplex is not captured during the capture step due to the absence of a biotin moiety. Therefore, the enriched extended recording tag fraction can be modulated by adjusting the ratio of the competitor oligonucleotide to the biotinylated "bait" oligonucleotide over a large dynamic range. This step will be important to address the dynamic range issue of protein abundance within the sample.

Figures 18A-B illustrate exemplary methods of single cell and bulk proteome partitioning into individual droplets, each droplet comprising a bead having a plurality of compartment tags attached thereto to correlate peptides to their originating protein complex, or to proteins originating from a single cell. The compartment tags comprise barcodes. Manipulation of droplet constituents after droplet formation: **(A)** Single cell partitioning into an individual droplet followed by cell lysis to release the cell proteome, and proteolysis to digest the cell proteome into peptides, and inactivation of the protease following sufficient proteolysis; **(B)** Bulk proteome partitioning into a plurality of droplets wherein an individual droplet comprises a protein complex followed by proteolysis to digest the protein complex into peptides, and inactivation of the protease following sufficient proteolysis. A heat labile metallo-protease can be used to digest the encapsulated proteins into peptides after photo-release of photo-caged divalent cations to activate the protease. The protease can be heat inactivated following sufficient proteolysis, or the divalent cations may be chelated. Droplets contain hybridized or releasable compartment tags comprising nucleic acid barcodes (separate from recording tag) capable of being ligated to either an N- or C- terminal amino acid of a peptide.

Figures 19A-B illustrate exemplary methods of single cell and bulk proteome partitioning into individual droplets, each droplet comprising a bead having a plurality of bifunctional recording tags with compartment tags attached thereto to correlate peptides to their originating protein or protein complex, or proteins to originating single cell. Manipulation of droplet constituents after post droplet formation: **(A)** Single cell partitioning into an individual droplet followed by cell lysis to release the cell proteome, and proteolysis to digest the cell proteome into peptides, and inactivation of the protease following sufficient proteolysis; **(B)** Bulk proteome partitioning into a plurality of droplets wherein an individual droplet comprises a protein complex followed by proteolysis to digest the protein complex into peptides, and inactivation of the protease following sufficient proteolysis. A heat labile metallo-protease can be used to digest the encapsulated proteins into peptides after photo-release of photo-caged divalent cations (e.g., Zn2+). The protease can be heat inactivated following sufficient proteolysis or the divalent cations may be chelated. Droplets contain hybridized or releasable compartment tags comprising nucleic acid barcodes (separate from recording tag) capable of being ligated to either an N- or C-terminal amino acid of a peptide.

Figures 20A-L illustrate generation of compartment barcoded recording tags attached to peptides. Compartment barcoding technology (*e.g.*, barcoded beads in microfluidic droplets, etc.) can be used to transfer a compartment-specific barcode to molecular contents encapsulated within a particular compartment. **(A)** In a particular embodiment, the protein molecule is denatured, and the ε-amine group of lysine residues (K) is chemically conjugated to an activated universal DNA tag molecule (comprising a universal priming sequence (U1)), shown with NHS moiety at the 5' end). After conjugation of universal DNA tags to the polypeptide, excess universal DNA tags are removed. **(B)** The universal DNA tagged-polypeptides are hybridized to nucleic acid molecules bound to beads, wherein the nucleic acid molecules bound to an individual bead comprise a unique population of compartment tag (barcode) sequences. The compartmentalization can occur by separating the sample into different physical compartments, such as droplets (illustrated by the dashed oval). Alternatively, compartmentalization can be directly accomplished by the immobilization of the labeled polypeptides on the bead surface, e.g., via annealing of the universal DNA tags on the polypeptide to the compartment DNA tags on the bead, without the need for additional physical separation. A single polypeptide molecule interacts with only a single bead (e.g., a single polypeptide does not span multiple beads). Multiple polypeptides, however, may interact with the same bead. In addition to the compartment barcode sequence (BC), the nucleic acid molecules bound to the bead may be comprised of a common Sp (spacer) sequence, a unique molecular identifier (UMI), and a sequence complementary to the polypeptide DNA tag, U1'. **(C)** After annealing of the universal DNA tagged polypeptides to the compartment tags bound to the bead, the compartment tags are released from the beads via cleavage of the attachment linkers. **(D)** The annealed U1 DNA tag primers are extended via polymerase-based primer extension using the compartment tag nucleic acid molecule originating from the bead as template. The primer extension step may be carried out after release of the compartment tags from the bead as shown in (C) or, optionally, while the compartment tags are still attached to the bead (not shown). This effectively writes the barcode sequence from the compartment tags on the bead onto the U1 DNA-tag sequence on the polypeptide. This new sequence constitutes a recording tag. After primer extension, a protease, *e.g.*, Lys-C (cleaves on C-terminal side of lysine residues), Glu-C (cleaves on C-terminal side of glutamic acid residues and to a lower extent glutamic acid residues), or random protease such as Proteinase K, is used to cleave the polypeptide into peptide fragments. **(E)** Each peptide fragment is labeled with an extended DNA tag sequence constituting a recording tag on its C-terminal lysine for downstream peptide sequencing as disclosed herein. **(F)** The recording tagged peptides are coupled to azide beads through a strained alkyne label, DBCO. The azide beads optionally also contain a capture sequence complementary to the recording tag to facilitate the efficiency of DBCO-azide immobilization. It should be noted that removing the peptides from the original beads and re-immobilizing to a new solid support (*e.g.*, beads) permits optimal intermolecular spacing between peptides to facilitate peptide sequencing methods as disclosed herein. **Figure 20G-L** illustrates a similar concept as illustrated in Figures20A-F except using click chemistry conjugation of DNA tags to an alkyne pre-labeled polypeptide (as described in Figure 2B). The Azide and mTet chemistries are orthogonal allowing click conjugation to DNA tags and click iEDDA conjugation (mTet and TCO) to the sequencing substrate .

Figure 21 illustrates an exemplary method using flow-focusing T-junction for single cell and compartment tagged (*e.g.*, barcode) compartmentalization with beads. With two aqueous flows, cell lysis and protease activation (Zn^{2+} mixing) can easily be initiated upon droplet formation.

Figures 22A-B illustrate exemplary tagging details. **(A)** A compartment tag (DNA-peptide chimera) is attached onto the peptide using peptide ligation with Butelase I. **(B)** Compartment tag information is transferred to an associated recording tag prior to commencement of peptide sequencing. Optionally, an endopeptidase AspN, which selectively cleaves peptide bonds N-terminal to aspartic acid residues, can be used to cleave the compartment tag after information transfer to the recording tag.

Figures 23A-C: Array-based barcodes for a spatial proteomics-based analysis of a tissue slice. **(A)** An array of spatially-encoded DNA barcodes (feature barcodes denoted by BC_{ij}), is combined with a tissue slice (FFPE or frozen). In one embodiment, the tissue slice is fixed and permeabilized. In a preferred embodiment, the array feature size is smaller than the cell size (~10 µm for human cells). **(B)** The array-mounted tissue slice is treated with reagents to reverse cross-linking (*e.g.*, antigen retrieval protocol w/ citraconic anhydride (Namimatsu, Ghazizadeh et al. 2005), and then the proteins therein are labeled with site-reactive DNA labels, that effectively label all protein molecules with DNA recording tags (*e.g.*, lysine labeling, liberated after antigen retrieval). After labeling and washing, the array bound DNA barcode sequences are cleaved and allowed to diffuse into the mounted tissue slice and hybridize to DNA recording tags attached to the proteins therein. **(C)** The array-mounted tissue is now subjected to polymerase extension to transfer information of the hybridized barcodes to the DNA recording tags labeling the proteins. After transfer of the barcode information, the array-mounted tissue is scraped from the slides, optionally digested with a protease, and the proteins or peptides extracted into solution.

Figures 24A-B illustrate two different exemplary DNA target macromolecules (AB and CD) that are immobilized on beads and assayed by binding agents attached to coding tags. This model system serves to illustrate the single molecule behavior of coding tag transfer from a bound agent to a proximal reporting tag. In the preferred embodiment, the coding tags are incorporated into an extended recoding tag via primer extension. **Figure 24A** illustrates the interaction of an AB macromolecule with an A-specific binding agent ("A'", an oligonucleotide sequence complementary to the "A" component of the AB macromolecule) and transfer of information of an associated coding tag to a recording tag via primer extension, and a B-specific binding agent ("B'", an oligonucleotide sequence complementary to the "B" component of the AB macromolecule) and transfer of information of an associated coding tag to a recoding tag via primer extension. Coding tags A and B are of different sequence, and for ease of identification in this illustration, are also of different length. The different lengths facilitate analysis of coding tag transfer by gel electrophoresis, but are not required for analysis by next generation sequencing. The binding of A' and B' binding agents are illustrated as alternative possibilities for a single binding cycle. If a second cycle is added, the extended recording tag would be further extended. Depending on which of A' or B' binding agents are added in the first and second cycles, the extended recording tags can contain coding tag information of the form AA, AB, BA, and BB. Thus, the extended recording tag contains information on the order of binding events as well as the identity of binders. Similarly, **Figure 24B** illustrates the interaction of a CD macromolecule with a C-specific binding agent ("C'", an oligonucleotide sequence complementary to the "C" component of the CD macromolecule) and transfer of information of an associated coding tag to a recording tag via primer extension, and a D-specific binding agent ("D'", an oligonucleotide sequence complementary to the "D" component of the CD macromolecule) and transfer of information of an associated coding tag to a recording tag via primer extension. Coding tags C and D are of different sequence and for ease of identification in this illustration are also of different length. The different lengths facilitate analysis of coding tag transfer by gel electrophoresis, but are not required for analysis by next generation sequencing. The binding of C' and D' binding agents are illustrated as alternative possibilities for a single binding cycle. If a second cycle is added, the extended recording tag would be further extended. Depending on which of C' or D' binding agents are added in the first and second cycles, the extended recording tags can contain coding tag information of the form CC, CD, DC, and DD. Coding tags may optionally comprise a UMI. The inclusion of UMIs in coding tags allows additional information to be recorded about a binding event; it allows binding events to be distinguished at the level of individual binding agents. This can be useful if an individual binding agent can participate in more than one binding event (*e.g.* its binding affinity is such that it can disengage and re-bind sufficiently frequently to participate in more than one event). It can also be useful for error-correction. For example, under some circumstances a coding tag might transfer information to the recording tag twice or more in the same binding cycle. The use of a UMI would reveal that these were likely repeated information transfer events all linked to a single binding event.

Figure 25 illustrates exemplary DNA target macromolecules (AB) and immobilized on beads and assayed by binding agents attached to coding tags. An A-specific binding agent ("A'", oligonucleotide complementary to A component of AB macromolecule) interacts with an AB macromolecule and information of an associated coding tag is transferred to a recording tag by ligation. A B-specific binding agent ("B'", an oligonucleotide complementary to B component of AB macromolecule) interacts with an AB macromolecule and information of an associated coding tag is transferred to a recording tag by ligation. Coding tags A and B are of different sequence and for ease of identification in this illustration are also of different length. The different lengths facilitate analysis of coding tag transfer by gel electrophoresis, but are not required for analysis by next generation sequencing.

Figures 26A-B illustrate exemplary DNA-peptide macromolecules for binding/coding tag transfer via primer extension. **Figure 26A** illustrates an exemplary oligonucleotide-peptide target macromolecule ("A" oligonucleotide-cMyc peptide) immobilized on beads. A cMyc-specific binding agent (*e.g.* antibody) interacts with the cMyc peptide portion of the macromolecule and information of an associated coding tag is transferred to a recording tag. The transfer of information of the cMyc coding tag to a recording tag may be analyzed by gel electrophoresis. **Figure 26B** illustrates an exemplary oligonucleotide-peptide target macromolecule ("C" oligonucleotide-hemagglutinin (HA) peptide) immobilized on beads. An HA-specific binding agent (*e.g.*, antibody) interacts with the HA peptide portion of the macromolecule and information of an associated coding tag is transferred to a recording tag. The transfer of information of the coding tag to a recording tag may be analyzed by gel electrophoresis. The binding of cMyc antibody-coding tag and HA antibody-coding tag are illustrated as alternative possibilities for a single binding cycle. If a second binding cycle is performed, the extended recording tag would be further extended. Depending on which of cMyc antibody-coding tag or HA antibody-coding tag are added in the first and second binding cycles, the extended recording tags can contain coding tag information of the form cMyc-HA, HA-cMyc, cMyc-cMyc, and HA-HA. Although not illustrated, additional binding agents can also be introduced to enable detection of the A and C oligonucleotide components of the macromolecules. Thus, hybrid macromolecules comprising different types of backbone can be analyzed via transfer of information to a recording tag and readout of the extended recording tag, which contains information on the order of binding events as well as the identity of the binding agents.

Figures 27A-D. Generation of Error-Correcting Barcodes. **(A)** A subset of 65 error-correcting barcodes (SEQ ID NOS:1-65) were selected from a set of 77 barcodes derived from the R software package ‘DNABarcodes’ (https://bioconductor.riken.jp/packages/3.3/bioc/manuals/DNABarcodes/man/DNABarcodes.pdf) using the command parameters [create.dnabarcodes(n=15,dist=10)]. This algorithm generates 15-mer “Hamming” barcodes that can correct substitution errors out to a distance of four substitutions, and detect errors out to nine substitutions. The subset of 65 barcodes was created by filtering out barcodes that didn’t exhibit a variety of nanopore current levels (for nanopore-based sequencing) or that were too correlated with other members of the set. **(B)** A plot of the predicted nanopore current levels for the 15-mer barcodes passing through the pore. The predicted currents were computed by splitting each 15-mer barcode word into composite sets of 11 overlapping 5-mer words, and using a 5-mer R9 nanopore current level look-up table (template_median68pA.5mers.model (https://github.com/jts/nanopolish/tree/master/etc/r9-models) to predict the corresponding current level as the barcode passes through the nanopore, one base at a time. As can be appreciated from (B), this set of 65 barcodes exhibit unique current signatures for each of its members. **(C)** Generation of PCR products as model extended recording tags for nanopore sequencing is shown using overlapping sets of DTR and DTR primers. PCR amplicons are then ligated to form a concatenated extended recording tag model. **(D)** Nanopore sequencing read of exemplary "extended recording tag" model (read length 734 bases) generated as shown in Figure 27C. The MinIon R9.4 Read has a quality score of 7.2 (poor read quality). However, barcode sequences can easily be identified using *lalign* even with a poor quality read (Qscore = 7.2). A 15-mer spacer element is underlined. Barcodes can align in either forward or reverse orientation, denoted by BC or BC' designation.

Figures 28A-D. Analyte-specific labeling of proteins with recording tags. (A) A binding agent targeting a protein analyte of interest in its native conformation comprises an analyte-specific barcode (BC_A') that hybridizes to a complementary analyte-specific barcode (BC_A) on a DNA recording tag. Alternatively, the DNA recording tag could be attached to the binding agent via a cleavable linker, and the DNA recording tag is "clicked" to the protein directly and is subsequently cleaved from the binding agent (via the cleavable linker). The DNA recording tag comprises a reactive coupling moiety (such as a click chemistry reagent (e.g., azide, mTet, etc.) for coupling to the protein of interest, and other functional components (e.g., universal priming sequence (P1), sample barcode (BC_S), analyte specific barcode (BC_A), and spacer sequence (Sp)). A sample barcode (BC_S) can also be used to label and distinguish proteins from different samples. The DNA recording tag may also comprise an orthogonal coupling moiety (e.g., mTet) for subsequent coupling to a substrate surface. For click chemistry coupling of the recording tag to the protein of interest, the protein is pre-labeled with a click chemistry coupling moiety cognate for the click chemistry coupling moiety on the DNA recording tag (e.g., alkyne moiety on protein is cognate for azide moiety on DNA recording tag). Examples of reagents for labeling the DNA recording tag with coupling moieties for click chemistry coupling include alkyne-NHS reagents for lysine labeling, alkyne-benzophenone reagents for photoaffinity labeling, etc. **(B)** After the binding agent binds to a proximal target protein, the reactive coupling moiety on the recording tag (e.g., azide) covalently attaches to the cognate click chemistry coupling moiety (shown as a triple line symbol) on the proximal protein. **(C)** After the target protein analyte is labeled with the recording tag, the attached binding agent is removed by digestion of uracils (U) using a uracil-specific excision reagent (e.g., USER™). **(D)** The DNA recording tag labeled target protein analyte is immobilized to a substrate surface using a suitable bioconjugate chemistry reaction, such as click chemistry (alkyne-azide binding pair, methyl tetrazine (mTET)- *trans*-cyclooctene (TCO) binding pair, etc.). In certain embodiments, the entire target protein-recording tag labeling assay is performed in a single tube comprising many different target protein analytes using a pool of binding agents and a pool of recording tags. After targeted labeling of protein analytes within a sample with recording tags comprising a sample barcode (BC_S), multiple protein analyte samples can be pooled before the immobilization step in (D). Accordingly, in certain embodiments, up to thousands of protein analytes across hundreds of samples can be labeled and immobilized in a single tube next generation protein assay (NGPA), greatly economizing on expensive affinity reagents (e.g., antibodies).

Figures 29A-E. Conjugation of DNA recording tags to polypeptides. **(A)** A denatured polypeptide is labeled with a bifunctional click chemistry reagent, such as alkyne-NHS ester (acetylene-PEG-NHS ester) reagent or alkyne-benzophenone to generate an alkyne-labeled (triple line symbol) polypeptide. An alkyne can also be a strained alkyne, such as cyclooctynes including Dibenzocyclooctyl (DBCO), etc. **(B)** An example of a DNA recording tag design that is chemically coupled to the alkyne-labeled polypeptide is shown. The recording tag comprises a universal priming sequence (P1), a barcode (BC), and a spacer sequence (Sp). The recording tag is labeled with a mTet moiety for coupling to a substrate surface and an azide moiety for coupling with the alkyne moiety of the labeled polypeptide. **(C)** A denatured, alkyne-labeled protein or polypeptide is labeled with a recording tag via the alkyne and azide moieties. Optionally, the recording tag-labeled polypeptide can be further labeled with a compartment barcode, e.g., via annealing to complementary sequences attached to a compartment bead and primer extension (also referred to as polymerase extension), or a shown in Figures 20H-J. **(D)** Protease digestion of the recording tag-labeled polypeptide creates a population of recording tag-labeled peptides. In some embodiments, some peptides will not be labeled with any recording tags. In other embodiments, some peptides may have one or more recording tags attached. **(E)** Recording tag-labeled peptides are immobilized onto a substrate surface using an inverse electron demand Diels-Alder (iEDDA) click chemistry reaction between the substrate surface functionalized with TCO groups and the mTet moieties of the recording tags attached to the peptides. In certain embodiments, clean-up steps may be employed between the different stages shown. The use of orthogonal click chemistries (e.g., azide-alkyne and mTet-TCO) allows both click chemistry labeling of the polypeptides with recording tags, and click chemistry immobilization of the recording tag-labeled peptides onto a substrate surface (*see*, McKay et al., 2014, Chem. Biol. 21:1075-1101, incorporated by reference in its entirety).

Figures 30A-E. Writing sample barcodes into recording tags after initial DNA tag labeling of polypeptides. **(A)** A denatured polypeptide is labeled with a bifunctional click chemistry reagent such as an alkyne-NHS reagent or alkyne-benzophenone to generate an alkyne-labeled polypeptide. **(B)** After alkyne (or alternative click chemistry moiety) labeling of the polypeptide, DNA tags comprising a universal priming sequence (P1) and labeled with an azide moiety and an mTet moiety are coupled to the polypeptide via the azide-alkyne interaction. It is understood that other click chemistry interactions may be employed. **(C)** A recording tag DNA construct comprising a sample barcode information (BC_S') and other recording tag functional components (e.g., universal priming sequence (P1'), spacer sequence (Sp')) anneals to the DNA tag-labeled polypeptide via complementary universal priming sequences (P1-P1'). Recording tag information is transferred to the DNA tag by polymerase extension. **(D)** Protease digestion of the recording tag-labeled polypeptide creates a population of recording tag-labeled peptides. **(E)** Recording tag-labeled peptides are immobilized onto a substrate surface using an inverse electron demand Diels-Alder (iEDDA) click chemistry reaction between a surface functionalized with TCO groups and the mTet moieties of the recording tags attached to the peptides. In certain embodiments, clean-up steps may be employed between the different stages shown. The use of orthogonal click chemistries (e.g., azide-alkyne and mTet-TCO) allows both click chemistry labeling of the polypeptides with recording tags, and click chemistry immobilization of the recording tag-labeled polypeptides onto a substrate surface (*see*, McKay et al., 2014, Chem. Biol. 21:1075-1101, incorporated by reference in its entirety).

Figures 31A-E. Bead compartmentalization for barcoding polypeptides. **(A)** A polypeptide is labeled in solution with a heterobifunctional click chemistry reagent using standard bioconjugation or photoaffinity labeling techniques. Possible labeling sites include ε-amine of lysine residues (e.g., with NHS-alkyne as shown) or the carbon backbone of the peptide (e.g., with benzophenone-alkyne). **(B)** Azide-labeled DNA tags comprising a universal priming sequence (P1) are coupled to the alkyne moieties of the labeled polypeptide. **(C)** The DNA tag-labeled polypeptide is annealed to DNA recording tag labeled beads via complementary DNA sequences (P1 and P1'). The DNA recording tags on the bead comprises a spacer sequence (Sp'), a compartment barcode sequence (BC_P'), an optional unique molecular identifier (UMI), and a universal sequence (P1'). The DNA recording tag informationis transferred to the DNA tags on the polypeptide via polymerase extension (alternatively, ligation could be employed). After information transfer, the resulting polypeptide comprises multiple recording tags containing several functional elements including compartment barcodes. **(D)** Protease digestion of the recording tag-labeled polypeptide creates a population of recording tag-labeled peptides. The recording tag-labeled peptides are dissociated from the beads, and **(E)** re-immobilized onto a sequencing substrate (e.g., using iEDDA click chemistry between mTet and TCO moieties as shown).

Figures 32A-H. Example of workflow for Next Generation Protein Assay (NGPA). A protein sample is labeled with a DNA recording tag comprised of several functional units, e.g., a universal priming sequence (P1), a barcode sequence (BC), an optional UMI sequence, and a spacer sequence (Sp) (enables information transfer with a binding agent coding tag). **(A)** The labeled proteins are immobilized (passively or covalently) to a substrate (e.g., bead, porous bead or porous matrix). **(B)** The substrate is blocked with protein and, optionally, competitor oligonucleotides (Sp') complementary to the spacer sequence are added to minimize non-specific interaction of the analyte recording tag sequence. **(C)** Analyte-specific antibodies (w/ associated coding tags) are incubated with substrate-bound protein. The coding tag may comprise a uracil base for subsequent uracil specific cleavage. **(D)** After antibody binding, excess competitor oligonucleotides (Sp'), if added, are washed away. The coding tag transiently anneals to the recording tag via complementary spacer sequences, and the coding tag information is transferred to the recording tag in a primer extension reaction to generate an extended recording tag. If the immobilized protein is denatured, the bound antibody and annealed coding tag can be removed under alkaline wash conditions such as with 0.1N NaOH. If the immobilized protein is in a native conformation, then milder conditions may be needed to remove the bound antibody and coding tag. An example of milder antibody removal conditions is outlined in panels E-H. **(E)** After information transfer from the coding tag to the recording tag, the coding tag is nicked (cleaved) at its uracil site using a uracil-specific excision reagent (e.g., USER™) enzyme mix. **(F)** The bound antibody is removed from the protein using a high-salt, low/high pH wash. The truncated DNA coding tag remaining attached to the antibody is short and rapidly elutes off as well. The longer DNA coding tag fragment may or may not remain annealed to the recording tag. **(G)** A second binding cycle commences as in steps (B)-(D) and a second primer extension step transfers the coding tag information from the second antibody to the extended recording tag via primer extension. **(H)** The result of two binding cycles is a concatenate of binding information from the first antibody and second antibody attached to the recording tag.

Figures 33A-D. Single-step Next Generation Protein Assay (NGPA) using multiple binding agents and enzymatically-mediated sequential information transfer. NGPA assay with immobilized protein molecule simultaneously bound by two cognate binding agents (e.g., antibodies). After multiple cognate antibody binding events, a combined primer extension and DNA nicking step is used to transfer information from the coding tags of bound antibodies to the recording tag. The caret symbol (^) in the coding tags represents a double stranded DNA nicking endonuclease site. **(A)** In the example shown, the coding tag of the antibody bound to epitope 1 (Epi#1) of a protein transfers coding tag information (e.g., encoder sequence) to the recording tag in a primer extension step following hybridization of complementary spacer sequences. **(B)** Once the double stranded DNA duplex between the extended recording tag and coding tag is formed, a nicking endonuclease that cleaves only one strand of DNA on a double-stranded DNA substrate, such as Nt.BsmAI, which is active at 37 °C, is used to cleave the coding tag. Following the nicking step, the duplex formed from the truncated coding tag-binding agent and extended recording tag is thermodynamically unstable and dissociates. The longer coding tag fragment may or may not remain annealed to the recording tag. **(C)** This allows the coding tag from the antibody bound to epitope #2 (Epi#2) of the protein to anneal to the extended recording tag via complementary spacer sequences, and the extended recording tag to be further extended by transferring information from the coding tag of Epi#2 antibody to the extended recording tag via primer extension. **(D)** Once again, after a double stranded DNA duplex is formed between the extended recording tag and coding tag of Epi#2 antibody, the coding tag is nicked by a nicking endonuclease, such Nb.BssSI. In certain embodiments, use of a non-strand displacing polymerase during primer extension (also referred to as polymerase extension) is preferred. A non-strand displacing polymerase prevents extension of the cleaved coding tag stub that remains annealed to the recording tag by more than a single base. The process of (A)-(D) can repeat itself until all the coding tags of proximal bound binding agents are "consumed" by the hybridization, information transfer to the extended recording tag, and nicking steps. The coding tag can comprise an encoder sequence identical for all binding agents (e.g., antibodies) specific for a given analyte (e.g., cognate protein), can comprise an epitope-specific encoder sequence, or can comprise a unique molecular identifier (UMI) to distinguish between different molecular events.

Figures 34A-C: Controlled density of recording tag -peptide immobilization using titration of reactive moieties on substrate surface. (A) Peptide density on a substrate surface may be titrated by controlling the density of functional coupling moieties on the surface of the substrate. This can be accomplished by derivitizing the surface of the substrate with an appropriate ratio of active coupling molecules to "dummy" coupling molecules. In the example shown, NHS—PEG-TCO reagent (active coupling molecule) is combined with NHS-mPEG (dummy molecule) in a defined ratio to derivitize an amine surface with TCO. Functionalized PEGs come in various molecular weights from 300 to over 40,000. **(B)** A bifunctional 5' amine DNA recording tag (mTet is other functional moiety) is coupled to a N-terminal Cys residue of a peptide using a succinimidyl 4-(N-maleimidomethyl)cyclohexane-1 (SMCC) bifunctional cross-linker. The internal mTet-dT group on the recording tag is created from an azide-dT group using mTetrazine-Azide. **(C)** The recording tag labeled peptides are immobilized to the activated substrate surface from (A) using the iEDDA click chemistry reaction with mTet and TCO. The mTet-TCO iEDDA coupling reaction is extremely fast, efficient, and stable (mTet-TCO is more stable than Tet-TCO).

Figures 35A-C. Next Generation Protein Sequencing (NGPS) Binding Cycle-Specific Coding Tags. (A) Design of NGPS assay with a cycle-specific N-terminal amino acid (NTAA) binding agent coding tags. An NTAA binding agent (e.g., antibody specific for N-terminal DNP-labeled tyrosine) binds to a DNP-labeled NTAA of a peptide associated with a recording tag comprising a universal priming sequence (P1), barcode (BC) and spacer sequence (Sp). When the binding agent binds to a cognate NTAA of the peptide, the coding tag associated with the NTAA binding agent comes into proximity of the recording tag and anneals to the recording tag via complementary spacer sequences. Coding tag information is transferred to the recording tag via primer extension. To keep track of which binding cycle a coding tag represents, the coding tag can comprise of a cycle-specific barcode. In certain embodiments, coding tags of binding agents that bind to an analyte have the same encoder barcode independent of cycle number, which is combined with a unique binding cycle-specific barcode. In other embodiments, a coding tag for a binding agent to an analyte comprises a unique encoder barcode for the combined analyte-binding cycle information. In either approach, a common spacer sequence can be used for binding agents' coding tags in each binding cycle. **(B)** In this example, binding agents from each binding cycle have a short binding cycle-specific barcode to identify the binding cycle, which together with the encoder barcode that identifies the binding agent, provides a unique combination barcode that identifies a particular binding agent-binding cycle combination. **(C)** After completion of the binding cycles, the extended recording tag can be converted into an amplifiable library using a capping cycle step where, for example, a cap comprising a universal priming sequence P1' linked to a universal priming sequence P2 and spacer sequence Sp' initially anneals to the extended recording tag via complementary P1 and P1' sequences to bring the cap in proximity to the extended recording tag. The complementary Sp and Sp' sequences in the extended recording tag and cap anneal and primer extension adds the second universal primer sequence (P2) to the extended recording tag.

Figures 36A-E. **DNA based model system for demonstrating information transfer from coding tags to recording tags.** Exemplary binding and intra-molecular writing was demonstrated by an oligonucleotide model system. The targeting agent A' and B' in coding tags were designed to hybridize to target binding regions A and B in recording tags. Recording tag (RT) mix was prepared by pooling two recoding tags, saRT_Abc_v2 (A target) and saRT_Bbc_V2 (B target), at equal concentrations. Recording tags are biotinylated at their 5' end and contain a unique target binding region, a universal forward primer sequence, a unique DNA barcode, and an 8 base common spacer sequence (Sp). The coding tags contain unique encoder barcodes base flanked by 8 base common spacer sequences (Sp'), one of which is covalently linked to A or B target agents via polyethylene glycol linker. **(A)** Biotinylated recording tag oligonucleotides (saRT_Abc_v2 and saRT_Bbc_V2) along with a biotinylated Dummy-T10 oligonucleotide were immobilized to streptavidin beads. The recording tags were designed with A or B capture sequences (recognized by cognate binding agents – A' and B', respectively), and corresponding barcodes (rtA_BC and rtB_BC) to identify the binding target. All barcodes in this model system were chosen from the set of 65 15-mer barcodes (SEQ ID NOS:1-65). In some cases, 15-mer barcodes were combined to constitute a longer barcode for ease of gel analysis. In particular, rtA_BC = BC_1 + BC_2; rtB_BC = BC_3. Two coding tags for binding agents cognate to the A and B sequences of the recording tags, namely CT_A'-bc (encoder barcode = BC_5) and CT_B'-bc (encoder barcode = BC_5+BC_6) were also synthesized. Complementary blocking oligos (DupCT_A'BC and DupCT_AB'BC) to a portion of the coding tag sequence (leaving a single stranded Sp' sequence) were optionally pre-annealed to the coding tags prior to annealing of coding tags to the bead-immobilized recording tags. A strand displacing polymerase removes the blocking oligo during polymerase extension. A barcode key (inset) indicates the assignment of 15-mer barcodes to the functional barcodes in the recording tags and coding tags. **(B)** The recording tag barcode design and coding tag encoder barcode design provide an easy gel analysis of "intra-moleculer" vs. "inter-molecular" interactions between recording tags and coding tags. In this design, undesired "inter-molecular" interactions (A recording tag with B' coding tag, and B recording tag with A' coding tag) generate gel products that are wither 15 bases longer or shorter than the desired "intra-molecular" (A recording tag with A' coding tag; B recording tag with B' coding tag) interaction products. The primer extension step changes the A' and B' coding tag barcodes (ctA'_BC, ctB'_BC) to the reverse complement barcodes (ctA_BC and ctB_BC). **(C)** A primer extension assay demonstrated information transfer from coding tags to recording tags, and addition of adapter sequences via primer extension on annealed EndCap oligo for PCR analysis. **(D)** Optimization of "intra-molecular" information transfer via titration of surface density of recording tags via use of Dummy-T20 oligo. Biotinylated recording tag oligos were mixed with biotinylated Dummy-T20 oligo at various ratios from 1:0, 1:10, all the way down to 1:10000. At reduced recording tag density ($1:10^3$ and $1:10^4$), "intra-molecular" interactions predominate over "inter-molecular" interactions. **(F)** As a simple extension of the DNA model system, a simple protein binding system comprising Nano-Tag_{15} peptide-Streptavidin binding pair is illustrated (K_D ~4 nM) (Perbandt et al., 2007, Proteins 67:1147-1153), but any number of peptide-binding agent model systems can be employed. Nano-Tag_{15} peptide sequence is (fM)DVEAWLGARVPLVET (SEQ ID NO:131) (fM = formyl-Met). Nano-Tag_{15} peptide further comprises a short, flexible linker peptide (GGGGS) and a cysteine residue for coupling to the DNA recording tag. Other examples peptide tag – cognate binding agent pairs include: calmodulin binding peptide (CBP)-calmodulin (K_D ~2 pM) (Mukherjee et al., 2015, J. Mol. Biol. 427: 2707-2725), amyloid-beta (Aβ16–27) peptide-US7/Lcn2 anticalin (0.2 nM) (Rauth et al., 2016, Biochem. J. 473: 1563-

1578), PA tag/NZ-1 antibody (K_D ~ 400 pM), FLAG-M2 Ab (28 nM), HA-4B2 Ab (1.6 nM), and Myc-9E10 Ab (2.2 nM) (Fujii et al., 2014, Protein Expr. Purif. 95:240-247). **(E)** As a test of intra-molecular information transfer from the binding agent's coding tag to the recording tag via primer extension, an oligonucleotide "binding agent" that binds to complementary DNA sequence "A" can be used in testing and development. This hybridization event has essentially greater than fM affinity. Streptavidin may be used as a test binding agent for the Nano-tag$_{15}$ peptide epitope. The peptide tag – binding agent interaction is high affinity, but can easily be disrupted with an acidic and/or high salt washes (Perbandt et al., supra).

Figures 37A-B. **Use of nano- or micro- emulsion PCR to transfer information from UMI-labeled N or C terminus to DNA tags labeling body of polypeptide.** **(A)** A polypeptide is labeled, at its N- or C- terminus with a nucleic acid molecule comprising a unique molecular identifier (UMI). The UMI may be flanked by sequences that are used to prime subsequent PCR. The polypeptide is then "body labeled" at internal sites with a separate DNA tag comprising sequence complementary to a priming sequence flanking the UMI. **(B)** The resultant labeled polypeptides are emulsified and undergo an emulsion PCR (ePCR) (alternatively, an emulsion in vitro transcription-RT-PCR (IVT-RT-PCR) reaction or other suitable amplification reaction can be performed) to amplify the N- or C-terminal UMI. A microemulsion or nanoemulsion is formed such that the average droplet diameter is 50-1000 nm, and that on average there is fewer than one polypeptide per droplet. A snapshot of a droplet content pre-and post PCR is shown in the left panel and right panel, respectively. The UMI amplicons hybridize to the internal polypeptide body DNA tags via complementary priming sequences and the UMI information is transferred from the amplicons to the internal polypeptide body DNA tags via primer extension.

Figure 38. Single Cell Proteomics. Cells are encapsulated and lysed in droplets containing polymer-forming subunits (e.g., acrylamide). The polymer-forming subunits are polymerized (e.g., polyacrylamide), and proteins are cross-linked to the polymer matrix. The emulsion droplets are broken and polymerized gel beads that

contain a single cell protein lysate attached to the permeable polymer matrix are released. The proteins are cross-linked to the polymer matrix in either their native conformation or in a denatured state by including a denaturant such as urea in the lysis and encapsulation buffer. Recording tags comprising a compartment barcode and other recording tag components (e.g., universal priming sequence (P1), spacer sequence (Sp), optional unique molecular identifier (UMI)) are attached to the proteins using a number of methods known in the art and disclosed herein, including emulsification with barcoded beads, or combinatorial indexing. The polymerized gel bead containing the single cell protein can also be subjected to proteinase digest after addition of the recording tag to generate recording tag labeled peptides suitable for peptide sequencing. In certain embodiments, the polymer matrix can be designed such that is dissolves in the appropriate additive such as disulfide cross-linked polymer that break upon exposure to a reducing agent such as tris(2-carboxyethyl)phosphine (TCEP) or dithiothreitol (DTT).

Figures 39A-E. Enhancement of amino acid cleavage reaction using a bifunctional N-terminal amino acid (NTAA) modifier and a chimeric cleavage reagent. **(A)** and **(B)** A peptide attached to a solid-phase substrate is modified with a bifunctional NTAA modifier, such as biotin-phenyl isothiocyanate (PITC). **(C)** A low affinity Edmanase (> μM Kd) is recruited to biotin-PITC labeled NTAAs using a streptavidin-Edmanase chimeric protein. **(D)** The efficiency of Edmanase cleavage is greatly improved due to the increase in effective local concentration as a result of the biotin-strepavidin interaction. **(E)** The cleaved biotin-PITC labeled NTAA and associated streptavidin-Edmanase chimeric protein diffuse away after cleavage. A number of other bioconjugation recruitment strategies can also be employed. An azide modified PITC is commercially available (4-Azidophenyl isothiocyanate, Sigma), allowing a number of simple transformations of azide-PITC into other bioconjugates of PITC, such as biotin-PITC via a click chemistry reaction with alkyne-biotin.

Figures 40A-I: Generation of C-terminal recording tag-labeled peptides from protein lysate (may be encapsulated in a gel bead). **(A)** A denatured polypeptide is reacted with an acid anhydride to label lysine residues. In one embodiment, a mix of alkyne (mTet)-substituted citraconic anhydride + proprionic anhydride is used to label the lysines with mTet. (shown as striped rectangles) **(B)** The result is an alkyne (mTet)-labeled polypeptide, with a fraction of lysines blocked with a proprionic group (shown as squares on the polypeptide chain). The alkyne (mTet) moiety is useful in click-chemistry based DNA labeling. **(C)** DNA tags (shown as solid rectangles) are attached by click chemistry using azide or *trans*-cyclooctene (TCO) labels for alkyne or mTet moieties, respectively. **(D)** Barcodes and functional elements such as a spacer (Sp) sequence and universal priming sequence are appended to the DNA tags using a primer extension step as shown in Figure 31 to produce recording tag-labeled polypeptide. The barcodes may be a sample barcode, a partition barcode, a compartment barcode, a spatial location barcode, etc., or any combination thereof. **(E)** The resulting recording tag-labeled polypeptide is fragmented into recording tag-labeled peptides with a protease or chemically. **(F)** For illustration, a peptide fragment labeled with two recording tags is shown. **(G)** A DNA tag comprising universal priming sequence that is complementary to the universal priming sequence in the recording tag is ligated to the C-terminal end of the peptide. The C-terminal DNA tag also comprises a moiety for conjugating the peptide to a surface. **(H)** The complementary universal priming sequences in the C-terminal DNA tag and a stochastically selected recording tag anneal. An intra-molecular primer extension reaction is used to transfer information from the recording tag to the C-terminal DNA tag. **(I)** The internal recording tags on the peptide are coupled to lysine residues via maleic anhydride, which coupling is reversible at acidic pH. The internal recording tags are cleaved from the peptide's lysine residues at acidic pH, leaving the C-terminal recording tag. The newly exposed lysine residues can optionally be blocked with a non-hydrolyzable anhydride, such as proprionic anhydride.

Figure 41. Workflow for a Preferred Embodiment of NGPS Assay.

Figures 42A-D. Exemplary Steps of NGPS Sequencing assay. An N-terminal amino acid (NTAA) acetylation or amidination step on a recording tag-labeled, surface bound peptide can occur before or after binding by an NTAA binding agent, depending on whether NTAA binding agents have been engineered to bind to acetylated

NTAAs or native NTAAs. In the first case, **(A)** the peptide is initially acetylated at the NTAA by chemical means using acetic anhydride or enzymatically with an N-terminal acetyltransferase (NAT). **(B)** The NTAA is recognized by an NTAA binding agent, such as an engineered anticalin, aminoacyl tRNA synthetase (aaRS), ClpS, etc. A DNA coding tag is attached to the binding agent and comprises a barcode encoder sequence that identifies the particular NTAA binding agent. **(C)** After binding of the acetylated NTAA by the NTAA binding agent, the DNA coding tag transiently anneals to the recording tag via complementary sequences and the coding tag information is transferred to the recording tag via polymerase extension. In an alternative embodiment, the recording tag information is transferred to the coding tag via polymerase extension. **(D)** The acetylated NTAA is cleaved from the peptide by an engineered acylpeptide hydrolase (APH), which catalyzes the hydrolysis of terminal acetylated amino acid from acetylated peptides. After cleavage of the acetylated NTAA, the cycle repeats itself starting with acetylation of the newly exposed NTAA. N-terminal acetylation is used as an exemplary mode of NTAA modification/cleavage, but other N-terminal moieties, such as a guanyl moiety can be substituted with a concomitant change in cleavage chemistry. If guanidination is employed, the guanylated NTAA can be cleaved under mild conditions using 0.5-2% NaOH solution (see Hamada, 2016, incorporated by reference in its entirety). APH is a serine peptidase able to catalyse the removal of Nα-acetylated amino acids from blocked peptides and it belongs to the prolyl oligopeptidase (POP) family (clan SC, family S9). It is a crucial regulator of N-terminally acetylated proteins in eukaryal, bacterial and archaeal cells.

Figures 43A-B. **Exemplary recording tag – coding tag design features.** **(A)** Structure of an exemplary recording tag associated protein (or peptide) and bound binding agent (e.g., anticalin) with associated coding tag. A thymidine (T) base is inserted between the spacer (Sp') and barcode (BC') sequence on the coding tag to accommodate a stochastic non-templated 3' terminal adenosine (A) addition in the primer extension reaction. **(B)** DNA coding tag is attached to a binding agent (e.g., anticalin) via SpyCatcher-SpyTag protein-peptide interaction.

Figures 44A-E. Enhancement of NTAA Cleavage Reaction Using Hybridization of Cleavage Agent to Recording Tag (A) and (B). A recording tag-labeled peptide attached to a solid-phase substrate (e.g., bead) is modified or labeled at the NTAA (Mod), e.g., with PITC, DNP, SNP, an acetyl modifier, guanidinylation, etc. **(C)** A cleavage enzyme (e.g., acylpeptide hydrolase (APH), amino peptidase (AP), Edmanase,etc.) is attached to a DNA tag comprising a universal priming sequence complementary to the universal priming sequence on the recording tag. The cleavage enzyme is recruited to the modified NTAA via hybridization of complementary universal priming sequences on the cleavage enzyme's DNA tag and the recording tag. **(D)** This hybridization step greatly improves the effective affinity of the cleavage enzyme for the NTAA. **(E)** The cleaved NTAA diffuses away and associated cleavage enzyme can be removed by stripping the hybridized DNA tag.

Figure 45. Cyclic degradation peptide sequencing using peptide ligase + protease + diaminopeptidase. Butelase I ligates the TEV-Butelase I peptide substrate (TENLYFQNHV, SEQ ID NO:132) to the NTAA of the query peptide. Butelase requires an NHV motif at the C-terminus of the peptide substrate. After ligation, Tobacco Etch Virus (TEV) protease is used to cleave the chimeric peptide substrate after the glutamine (Q) residue, leaving a chimeric peptide having an asparagine (N) residue attached to the N-terminus of the query peptide. Diaminopeptidase (DAP) or Dipeptidyl-peptidase, which cleaves two amino acid residues from the N-terminus, shortens the N-added query peptide by two amino acids effectively removing the asparagine residue (N) and the original NTAA on the query peptide. The newly exposed NTAA is read using binding agents as provided herein, and then the entire cycle is repeated "n" times for "n" amino acids sequenced. The use of a streptavidin-DAP metalloenzyme chimeric protein and tethering a biotin moiety to the N-terminal asparagine residue may allow control of DAP processivity.

DETAILED DESCRIPTION

Terms not specifically defined herein should be given the meanings that would be given to them by one of skill in the art in light of the disclosure and the context. As used in the specification, however, unless specified to the contrary, the terms have the meaning indicated.

I. Introduction

The present disclosure provides, in part, methods of highly-parallel, high throughput digital macromolecule characterization and quantitation, with direct applications to protein and peptide characterization and sequencing (see, Figure 1B, Figure 2A). The methods described herein use binding agents comprising a coding tag with identifying information in the form of a nucleic acid molecule or sequenceable polymer, wherein the binding agents interact with a macromolecule of interest. Multiple, successive binding cycles, each cycle comprising exposing a plurality macromolecules, preferably representing pooled samples, immobilized on a solid support to a plurality of binding agents, are performed. During each binding cycle, the identity of each binding agent that binds to the macromolecule, and optionally binding cycle number, is recorded by transferring information from the binding agent coding tag to a recording tag co-localized with the macromolecule. In an alternative embodiment, information from the recording tag comprising identifying information for the associated macromolecule may be transferred to the coding tag of the bound binding agent (*e.g.*, to form an extended coding tag) or to a third “di-tag” construct. Multiple cycles of binding events build historical binding information on the recording tag co-localized with the macromolecule, thereby producing an extended recording tag comprising multiple coding tags in co-linear order representing the temporal binding history for a given macromolecule. In addition, cycle-specific coding tags can be employed to track information from each cycle, such that if a cycle is skipped for some reason, the extended recording tag can continue to collect information in subsequent cycles, and identify the cycle with missing information.

Alternatively, instead of writing or transferring information from the coding tag to recording tag, information can be transferred from a recording tag comprising identifying information for the associated macromolecule to the coding tag forming an extended coding tag or to a third di-tag construct. The resulting extended coding tags or di-tags can be collected after each binding cycle for subsequent sequence analysis. The identifying information on the recording tags comprising barcodes (e.g., partition tags, compartment tags, sample tags, fraction tags, UMIs, or any combination thereof) can be used to map the extended coding tag or di-tag sequence reads back to the originating macromolecule. In this manner, a nucleic acid encoded library representation of the binding history of the macromolecule is generated. This nucleic acid encoded library can be amplified, and analyzed using very high-throughput next generation digital sequencing methods, enabling millions to billions of molecules to be analyzed per run. The creation of a nucleic acid encoded library of binding information is useful in another way in that it enables enrichment, subtraction, and normalization by DNA-based techniques that make use of hybridization. These DNA-based methods are easily and rapidly scalable and customizable, and more cost-effective than those available for direct manipulation of other types of macromolecule libraries, such as protein libraries. Thus, nucleic acid encoded libraries of binding information can be processed prior to sequencing by one or more techniques to enrich and/or subtract and/or normalize the representation of sequences. This enables information of maximum interest to be extracted much more efficiently, rapidly and cost-effectively from very large libraries whose individual members may initially vary in abundance over many orders of magnitude. Importantly, these nucleic-acid based techniques for manipulating library representation are orthogonal to more conventional methods, and can be used in combination with them. For example, common, highly abundant proteins, such as albumin, can be subtracted using protein-based methods, which may remove the majority but not all the undesired protein. Subsequently, the albumin-specific members of an extended recording tag library can also be subtracted, thus achieving a more complete overall subtraction.

In one aspect, the present disclosure provides a highly-parallelized approach for peptide sequencing using a Edman-like degradation approach, allowing the sequencing from a large collection of DNA recording tag-labeled peptides (e.g., millions to billions). These recording tag labeled peptides are derived from a proteolytic digest or limited hydrolysis of a protein sample, and the recording tag labeled peptides are immobilized randomly on a sequencing substrate (e.g., porous beads) at an appropriate inter-molecular spacing on the substrate. Modification of N-terminal amino acid (NTAA) residues of the peptides with small chemical moieties, such as phenylthiocarbamoyl (PTC), dinitrophenol (DNP), sulfonyl nitrophenol (SNP), dansyl, 7-methoxy coumarin, acetyl, or guanidinyl, that catalyze or recruit an NTAA cleavage reaction allows for cyclic control of the Edman-like degradation process. The modifying chemical moieties may also provide enhanced binding affinity to cognate NTAA binding agents. The modified NTAA of each immobilized peptide is identified by the binding of a cognate NTAA binding agent comprising a coding tag, and transferring coding tag information (e.g., encoder sequence providing identifying information for the binding agent) from the coding tag to the recording tag of the peptide (e.g, primer extension or ligation). Subsequently, the modified NTAA is removed by chemical methods or enzymatic means. In certain embodiments, enzymes (e.g., Edmanase) are engineered to catalyze the removal of the modified NTAA. In other embodiments, naturally occurring exopeptidases, such as aminopeptidases or acyl peptide hydrolases, can be engineered to cleave a terminal amino acid only in the presence of a suitable chemical modification.

II. Definitions

In the following description, certain specific details are set forth in order to provide a thorough understanding of various embodiments. However, one skilled in the art will understand that the present compounds may be made and used without these details. In other instances, well-known structures have not been shown or described in detail to avoid unnecessarily obscuring descriptions of the embodiments. Unless the context requires otherwise, throughout the specification and claims which follow, the word “comprise” and variations thereof, such as, “comprises” and “comprising,” are to be construed in an open, inclusive sense, that is, as “including, but not limited to.” In addition, the term “comprising” (and related terms such as “comprise” or “comprises” or “having” or “including”) is not intended to exclude that in other certain embodiments, for example, an embodiment of any composition of matter, composition, method, or process, or the like, described herein, may "consist of" or "consist essentially of" the described features. Headings provided herein are for convenience only and do not interpret the scope or meaning of the claimed embodiments.

Reference throughout this specification to "one embodiment" or "an embodiment" means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment. Thus, the appearances of the phrases "in one embodiment" or "in an embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures, or characteristics may be combined in any suitable manner in one or more embodiments.

As used herein, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a peptide" includes one or more peptides, or mixtures of peptides. Also, and unless specifically stated or obvious from context, as used herein, the term "or" is understood to be inclusive and covers both "or" and "and".

As used herein, the term "macromolecule" encompasses large molecules composed of smaller subunits. Examples of macromolecules include, but are not limited to peptides, polypeptides, proteins, nucleic acids, carbohydrates, lipids, macrocycles. A macromolecule also includes a chimeric macromolecule composed of a combination of two or more types of macromolecules, covalently linked together (*e.g.*, a peptide linked to a nucleic acid). A macromolecule may also include a "macromolecule assembly", which is composed of non-covalent complexes of two or more macromolecules. A macromolecule assembly may be composed of the same type of macromolecule (*e.g.*, protein-protein) or of two more different types of macromolecules (*e.g.*, protein-DNA).

As used herein, the term "peptide" encompasses peptides, polypeptides and proteins, and refers to a molecule comprising a chain of two or more amino acids joined by peptide bonds. In general terms, a peptide having more than 20-30 amino acids is commonly referred to as a polypeptide, and one having more than 50 amino acids is commonly referred to as a protein. The amino acids of the peptide are most typically L-amino acids, but may also be D-amino acids, modified amino acids, amino acid analogs, amino acid mimetics, or any combination thereof. Peptides may be naturally occurring, synthetically produced, or recombinantly expressed. Peptides may also comprise additional groups modifying the amino acid chain, for example, functional groups added *via* post-translational modification.

As used herein, the term “amino acid” refers to an organic compound comprising an amine group, a carboxylic acid group, and a side-chain specific to each amino acid, which serve as a monomeric subunit of a peptide. An amino acid includes the 20 standard, naturally occurring or canonical amino acids as well as non-standard amino acids. The standard, naturally-occurring amino acids include Alanine (A or Ala), Cysteine (C or Cys), Aspartic Acid (D or Asp), Glutamic Acid (E or Glu), Phenylalanine (F or Phe), Glycine (G or Gly), Histidine (H or His), Isoleucine (I or Ile), Lysine (K or Lys), Leucine (L or Leu), Methionine (M or Met), Asparagine (N or Asn), Proline (P or Pro), Glutamine (Q or Gln), Arginine (R or Arg), Serine (S or Ser), Threonine (T or Thr), Valine (V or Val), Tryptophan (W or Trp), and Tyrosine (Y or Tyr). An amino acid may be an L-amino acid or a D-amino acid. Non-standard amino acids may be modified amino acids, amino acid analogs, amino acid mimetics, non-standard proteinogenic amino acids, or non-proteinogenic amino acids that occur naturally or are chemically synthesized. Examples of non-standard amino acids include, but are not limited to, selenocysteine, pyrrolysine, and N-formylmethionine, β-amino acids, Homo-amino acids, Proline and Pyruvic acid derivatives, 3-substituted alanine derivatives, glycine derivatives, ring-substituted phenylalanine and tyrosine derivatives, linear core amino acids, N-methyl amino acids.

As used herein, the term “post-translational modification” refers to modifications that occur on a peptide after its translation by ribosomes is complete. A post-translational modification may be a covalent modification or enzymatic modification. Examples of post-translation modifications include, but are not limited to, acylation, acetylation, alkylation (including methylation), biotinylation, butyrylation, carbamylation, carbonylation, deamidation, deiminiation, diphthamide formation, disulfide bridge formation, eliminylation, flavin attachment, formylation, gamma-carboxylation, glutamylation, glycylation, glycosylation, glypiation, heme C attachment, hydroxylation, hypusine formation, iodination, isoprenylation, lipidation, lipoylation, malonylation, methylation, myristolylation, oxidation, palmitoylation, pegylation, phosphopantetheinylation, phosphorylation, prenylation, propionylation, retinylidene Schiff base formation, S-glutathionylation, S-nitrosylation, S-sulfenylation, selenation, succinylation, sulfination, ubiquitination, and C-terminal amidation. A post-translational modification includes modifications of the amino terminus and/or the carboxyl terminus of a peptide. Modifications of the terminal amino group include, but are not limited to, des-amino, N-lower alkyl, N-di-lower alkyl, and N-acyl modifications. Modifications of the terminal carboxy group include, but are not limited to, amide, lower alkyl amide, dialkyl amide, and lower alkyl ester modifications (*e.g.*, wherein lower alkyl is C_1-C_4 alkyl). A post-translational modification also includes modifications, such as but not limited to those described above, of amino acids falling between the amino and carboxy termini. The term post-translational modification can also include peptide modifications that include one or more detectable labels.

As used herein, the term “binding agent” refers to a nucleic acid molecule, a peptide, a polypeptide, a protein, carbohydrate, or a small molecule that binds to, associates, unites with, recognizes, or combines with a macromolecule or a component or feature of a macromolecule. A binding agent may form a covalent association or non-covalent association with the macromolecule or component or feature of a macromolecule. A binding agent may also be a chimeric binding agent, composed of two or more types of molecules, such as a nucleic acid molecule-peptide chimeric binding agent or a carbohydrate-peptide chimeric binding agent. A binding agent may be a naturally occurring, synthetically produced, or recombinantly expressed molecule. A binding agent may bind to a single monomer or subunit of a macromolecule (*e.g.*, a single amino acid of a peptide) or bind to a plurality of linked subunits of a macromolecule (*e.g.*, a di-peptide , tri-peptide, or higher order peptide of a longer peptide, polypeptide, or protein molecule). A binding agent may bind to a linear molecule or a molecule having a three-dimensional structure (also referred to as conformation). For example, an antibody binding agent may bind to linear peptide, polypeptide, or protein, or bind to a conformational peptide, polypeptide, or protein. A binding agent may bind to an N-terminal peptide, a C-terminal peptide, or an intervening peptide of a peptide, polypeptide, or protein molecule. A binding agent may bind to an N-terminal amino acid, C-terminal amino acid, or an intervening amino acid of a peptide molecule. A binding agent may preferably bind to a chemically modified or labeled amino acid over a non-modified or unlabeled amino acid. For example, a binding agent may preferably bind to an amino acid that has been modified with an acetyl moiety, guanyl moiety, dansyl moiety, PTC moiety, DNP moiety, SNP moiety, etc., over an amino acid that does not possess said moiety. A binding agent may bind to a post-translational modification of a peptide molecule. A binding agent may exhibit selective binding to a component or feature of a macromolecule (*e.g.*, a binding agent may selectively bind to one of the 20 possible natural amino acid residues and with bind with very low affinity or not at all to the other 19 natural amino acid residues). A binding agent may exhibit less selective binding, where the binding agent is capable of binding a plurality of components or features of a macromolecule (*e.g.*, a binding agent may bind with similar affinity to two or more different amino acid residues). A binding agent comprises a coding tag, which is joined to the binding agent by a linker.

As used herein, the term “linker” refers to one or more of a nucleotide, a nucleotide analog, an amino acid, a peptide, a polypeptide, or a non-nucleotide chemical moiety that is used to join two molecules. A linker may be used to join a binding agent with a coding tag, a recording tag with a macromolecule (e.g., peptide), a macromolecule with a solid support, a recording tag with a solid support, etc. In certain embodiments, a linker joins two molecules via enzymatic reaction or chemistry reaction (e.g., click chemistry).

As used herein, the term “proteomics” refers to quantitative analysis of the proteome within cells, tissues, and bodily fluids, and the corresponding spatial distribution of the proteome within the cell and within tissues. Additionally, proteomics studies include the dynamic state of the proteome, continually changing in time as a function of biology and defined biological or chemical stimuli.

As used herein, the term "non-cognate binding agent" refers to a binding agent that is not capable of binding or binds with low affinity to a macromolecule feature, component, or subunit being interrogated in a particular binding cycle reaction as compared to a "cognate binding agent", which binds with high affinity to the corresponding macromolecule feature, component, or subunit. For example, if a tyrosine residue of a peptide molecule is being interrogated in a binding reaction, non-cognate binding agents are those that bind with low affinity or not at all to the tyrosine residue, such that the non-cognate binding agent does not efficiently transfer coding tag information to the recording tag under conditions that are suitable for transferring coding tag information from cognate binding agents to the recording tag. Alternatively, if a tyrosine residue of a peptide molecule is being interrogated in a binding reaction, non-cognate binding agents are those that bind with low affinity or not at all to the tyrosine residue, such that recording tag information does not efficiently transfer to the coding tag under suitable conditions for those embodiments involving extended coding tags rather than extended recording tags.

The terminal amino acid at one end of the peptide chain that has a free amino group is referred to herein as the "N-terminal amino acid" (NTAA). The terminal amino acid at the other end of the chain that has a free carboxyl group is referred to herein as the "C-terminal amino acid" (CTAA). The amino acids making up a peptide may be numbered in order, with the peptide being "*n*" amino acids in length. As used herein, NTAA is considered the n^{th} amino acid (also referred to herein as the "*n* NTAA"). Using this nomenclature, the next amino acid is the *n-1* amino acid, then the *n-2* amino acid, and so on down the length of the peptide from the N-terminal end to C-terminal end. In certain embodiments, an NTAA, CTAA, or both may be modified or labeled with a chemical moiety.

As used herein, the term "barcode" refers to a nucleic acid molecule of about 2 to about 30 bases (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30 bases) providing a unique identifier tag or origin information for a macromolecule (e.g., protein, polypeptide, peptide), a binding agent, a set of binding agents from a binding cycle, a sample macromolecules, a set of samples, macromolecules within a compartment (e.g., droplet, bead, or separated location), macromolecules within a set of compartments, a fraction of macromolecules, a set of macromolecule fractions, a spatial region or set of spatial regions, a library of macromolecules, or a library of binding agents. A barcode can be an artificial sequence or a naturally occurring sequence. In certain embodiments, each barcode within a population of barcodes is different. In other embodiments, a portion of barcodes in a population of barcodes is different, *e.g*, at least about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, or 99% of the barcodes in a population of barcodes is different. A population of barcodes may be randomly generated or non-randomly generated. In certain embodiments, a population of barcodes are error correcting barcodes. Barcodes can be used to computationally deconvolute the multiplexed sequencing data and identify sequence reads derived from an individual macromolecule, sample, library, etc. A barcode can also be used for deconvolution of a collection of macromolecules that have been distributed into small compartments for enhanced mapping. For example, rather than mapping a peptide back to the proteome, the peptide is mapped back to its originating protein molecule or protein complex.

A “sample barcode”, also referred to as “sample tag” identifies from which sample a macromolecule derives.

A “spatial barcode” which region of a 2-D or 3-D tissue section from which a macromolecule derives. Spatial barcodes may be used for molecular pathology on tissue sections. A spatial barcode allows for multiplex sequencing of a plurality of samples or libraries from tissue section(s).

As used herein, the term “coding tag” refers to a nucleic acid molecule of about 2 bases to about 100 bases, including any integer including 2 and 100 and in between, that comprises identifying information for its associated binding agent. A “coding tag” may also be made from a “sequencable polymer” (*see, e.g*., Niu et al., 2013, Nat. Chem. 5:282-292; Roy et al., 2015, Nat. Commun. 6:7237; Lutz, 2015, Macromolecules 48:4759-4767; each of which are incorporated by reference in its entirety). A coding tag comprises an encoder sequence, which is optionally flanked by one spacer on one side or flanked by a spacer on each side. A coding tag may also be comprised of an optional UMI and/or an optional binding cycle-specific barcode. A coding tag may be single stranded or double stranded. A double stranded coding tag may comprise blunt ends, overhanging ends, or both. A coding tag may refer to the coding tag that is directly attached to a binding agent, to a complementary sequence hybridized to the coding tag directly attached to a binding agent (*e.g.*, for double stranded coding tags), or to coding tag information present in an extended recording tag. In certain embodiments, a coding tag may further comprise a binding cycle specific spacer or barcode, a unique molecular identifier, a universal priming site, or any combination thereof.

As used herein, the term “encoder sequence” or “encoder barcode” refers to a nucleic acid molecule of about 2 bases to about 30 bases (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30 bases) in length that provides identifying information for its associated binding agent. The encoder sequence may uniquely identify its associated binding agent. In certain embodiments, an encoder sequence is provides identifying information for its associated binding agent and for the binding cycle in which the binding agent is used. In other embodiments, an encoder sequence is combined with a separate binding cycle-specific barcode within a coding tag. Alternatively, the encoder sequence may identify its associated binding agent as belonging to a member of a set of two or more different binding agents. In some embodiments, this level of identification is sufficient for the purposes of analysis. For example, in some embodiments involving a binding agent that binds to an amino acid, it may be sufficient to know that a peptide comprises one of two possible amino acids at a particular position, rather than definitively identify the amino acid residue at that position. In another example, a common encoder sequence is used for polyclonal antibodies, which comprises a mixture of antibodies that recognize more than one epitope of a protein target, and have varying specificities. In other embodiments, where an encoder sequence identifies a set of possible binding agents, a sequential decoding approach can be used to produce unique identification of each binding agent. This is accomplished by varying encoder sequences for a given binding agent in repeated cycles of binding (*see*, Gunderson et al., 2004, Genome Res. 14:870-7). The partially identifying coding tag information from each binding cycle, when combined with coding information from other cycles, produces a unique identifier for the binding agent, *e.g.*, the particular combination of coding tags rather than an individual coding tag (or encoder sequence) provides the uniquely identifying information for the binding agent. Preferably, the encoder sequences within a library of binding agents possess the same or a similar number of bases.

As used herein the term “binding cycle specific tag”, “binding cycle specific barcode”, or “binding cycle specific sequence” refers to a unique sequence used to identify a library of binding agents used within a particular binding cycle. A binding cycle specific tag may comprise about 2 bases to about 8 bases (e.g., 2, 3, 4, 5, 6, 7, or 8 bases) in length. A binding cycle specific tag may be incorporated within a binding agent’s coding tag as part of a spacer sequence, part of an encoder sequence, part of a UMI, or as a separate component within the coding tag.

As used herein, the term “spacer” (Sp) refers to a nucleic acid molecule of about 1 base to about 20 bases (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 bases) in length that is present on a terminus of a recording tag or coding tag. In certain embodiments, a spacer sequence flanks an encoder sequence of a coding tag on one end or both ends. Following binding of a binding agent to a macromolecule, annealing between complementary spacer sequences on their associated coding tag and recording tag, respectively, allows transfer of binding information through a primer extension reaction or ligation to the recording tag, coding tag, or a di-tag construct. Sp’ refers to spacer sequence complementary to Sp. Preferably, spacer sequences within a library of binding agents possess the same number of bases. A common (shared or identical) spacer may be used in a library of binding agents. A spacer sequence may have a “cycle specific” sequence in order to track binding agents used in a particular binding cycle. The spacer sequence (Sp) can be constant across all binding cycles, be specific for a particular class of macromolecules, or be binding cycle number specific. Macromolecule class-specific spacers permit annealing of a cognate binding agent’s coding tag information present in an extended recording tag from a completed binding/extension cycle to the coding tag of another binding agent recognizing the same class of macromolecules in a subsequent binding cycle via the class-specific spacers. Only the sequential binding of correct cognate pairs results in interacting spacer elements and effective primer extension. A spacer sequence may comprise sufficient number of bases to anneal to a complementary spacer sequence in a recording tag to initiate a primer extension (also referred to as polymerase extension) reaction, or provide a "splint" for a ligation reaction, or mediate a "sticky end" ligation reaction. A spacer sequence may comprise a fewer number of bases than the encoder sequence within a coding tag.

As used herein, the term "recording tag" refers to a nucleic acid molecule or sequenceable polymer molecule (*see, e.g.*, Niu et al., 2013, Nat. Chem. 5:282-292; Roy et al., 2015, Nat. Commun. 6:7237; Lutz, 2015, Macromolecules 48:4759-4767; each of which are incorporated by reference in its entirety) that comprises identifying information for a macromolecule to which it is associated. In certain embodiments, after a binding agent binds a macromolecule, information from a coding tag linked to a binding agent can be transferred to the recording tag associated with the macromolecule while the binding agent is bound to the macromolecule. In other embodiments, after a binding agent binds a macromolecule, information from a recording tag associated with the macromolecule can be transferred to the coding tag linked to the binding agent while the binding agent is bound to the macromolecule. A recoding tag may be directly linked to a macromolecule, linked to a macromolecule via a multifunctional linker, or associated with a macromolecule by virtue of its proximity (or co-localization) on a solid support. A recording tag may be linked via its 5' end or 3' end or at an internal site, as long as the linkage is compatible with the method used to transfer coding tag information to the recording tag or vice versa. A recording tag may further comprise other functional components, e.g., a universal priming site, unique molecular identifier, a barcode (*e.g.*, a sample barcode, a fraction barcode, spatial barcode, a compartment tag, etc.), a spacer sequence that is complementary to a spacer sequence of a coding tag, or any combination thereof. The spacer sequence of a recording tag is preferably at the 3'-end of the recording tag in embodiments where polymerase extension is used to transfer coding tag information to the recording tag.

As used herein, the term “primer extension”, also referred to as “polymerase extension”, refers to a reaction catalyzed by a nucleic acid polymerase (e.g., DNA polymerase) whereby a nucleic acid molecule (e.g., oligonucleotide primer, spacer sequence) that anneals to a complementary strand is extended by the polymerase, using the complementary strand as template.

As used herein, the term “unique molecular identifier” or “UMI” refers to a nucleic acid molecule of about 3 to about 40 bases (3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 bases in length providing a unique identifier tag for each macromolecule (*e.g.*, peptide) or binding agent to which the UMI is linked. A macromolecule UMI can be used to computationally deconvolute sequencing data from a plurality of extended recording tags to identify extended recording tags that originated from an individual macromolecule. A binding agent UMI can be used to identify each individual binding agent that binds to a particular macromolecule. For example, a UMI can be used to identify the number of individual binding events for a binding agent specific for a single amino acid that occurs for a particular peptide molecule. It is understood that when UMI and barcode are both referenced in the context of a binding agent or macromolecule, that the barcode refers to identifying information other that the UMI for the individual binding agent or macromolecule (*e.g.*, sample barcode, compartment barcode, binding cycle barcode).

As used herein, the term “universal priming site” or “universal primer” or “universal priming sequence” refers to a nucleic acid molecule, which may be used for library amplification and/or for sequencing reactions. A universal priming site may include, but is not limited to, a priming site (primer sequence) for PCR amplification, flow cell adaptor sequences that anneal to complementary oligonucleotides on flow cell surfaces enabling bridge amplification in some next generation sequencing platforms, a sequencing priming site, or a combination thereof. Universal priming sites can be used for other types of amplification, including those commonly used in conjunction with next generation digital sequencing. For example, extended recording tag molecules may be circularized and a universal priming site used for rolling circle amplification to form DNA nanoballs that can be used as sequencing templates (Drmanac et al., 2009, Science 327:78-81). Alternatively, recording tag molecules may be circularized and sequenced directly by polymerase extension from universal priming sites (Korlach et al., 2008, Proc. Natl. Acad. Sci. 105:1176-1181). The term "forward" when used in context with a "universal priming site" or "universal primer" may also be referred to as "5'" or "sense". The term "reverse" when used in context with a "universal priming site" or "universal primer" may also be referred to as "3'" or "antisense".

As used herein, the term "extended recording tag" refers to a recording tag to which information of at least one binding agent's coding tag (or its complementary sequence) has been transferred following binding of the binding agent to a macromolecule. Information of the coding tag may be transferred to the recording tag directly (*e.g.*, ligation) or indirectly (*e.g.*, primer extension). Information of a coding tag may be transferred to the recording tag enzymatically or chemically. An extended recording tag may comprise binding agent information of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200 or more coding tags. The base sequence of an extended recording tag may reflect the temporal and sequential order of binding of the binding agents identified by their coding tags, may reflect a partial sequential order of binding of the binding agents identified by the coding tags, or may not reflect any order of binding of the binding agents identified by the coding tags. In certain embodiments, the coding tag information present in the extended recording tag represents with at least 25%, 30%, 35% , 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97% 98%, 99%, or 100% identity the macromolecule sequence being analyzed. In certain embodiments where the extended recording tag does not represent the macromolecule sequence being analyzed with 100% identity, errors may be due to off-target binding by a binding agent, or to a "missed" binding cycle (e.g., because a binding agent fails to bind to a macromolecule during a binding cycle, because of a failed primer extension reaction), or both.

As used herein, the term “extended coding tag” refers to a coding tag to which information of at least one recording tag (or its complementary sequence) has been transferred following binding of a binding agent, to which the coding tag is joined, to a macromolecule, to which the recording tag is associated. Information of a recording tag may be transferred to the coding tag directly (*e.g.*, ligation), or indirectly (*e.g.*, primer extension). Information of a recording tag may be transferred enzymatically or chemically. In certain embodiments, an extended coding tag comprises information of one recording tag, reflecting one binding event. As used herein, the term “di-tag” or “di-tag construct” or “di-tag molecule” refers to a nucleic acid molecule to which information of at least one recording tag (or its complementary sequence) and at least one coding tag (or its complementary sequence) has been transferred following binding of a binding agent, to which the coding tag is joined, to a macromolecule, to which the recording tag is associated (see, Figure 11B). Information of a recording tag and coding tag may be transferred to the di-tag indirectly (*e.g.*, primer extension). Information of a recording tag may be transferred enzymatically or chemically. In certain embodiments, a di-tag comprises a UMI of a recording tag, a compartment tag of a recording tag, a universal priming site of a recording tag, a UMI of a coding tag, an encoder sequence of a coding tag, a binding cycle specific barcode, a universal priming site of a coding tag, or any combination thereof.

As used herein, the term “solid support”, “solid surface”, or “solid substrate” or “substrate” refers to any solid material, including porous and non-porous materials, to which a macromolecule (*e.g.*, peptide) can be associated directly or indirectly, by any means known in the art, including covalent and non-covalent interactions, or any combination thereof. A solid support may be two-dimensional (e.g., planar surface) or three-dimensional (e.g., gel matrix or bead). A solid support can be any support surface including, but not limited to, a bead, a microbead, an array, a glass surface, a silicon surface, a plastic surface, a filter, a membrane, nylon, a silicon wafer chip, a flow through chip, a flow cell, a biochip including signal transducing electronics, a channel, a microtiter well, an ELISA plate, a spinning interferometry disc, a nitrocellulose membrane, a nitrocellulose-based polymer surface, a polymer matrix, a nanoparticle, or a microsphere. Materials for a solid support include but are not limited to acrylamide, agarose, cellulose, nitrocellulose, glass, gold, quartz, polystyrene, polyethylene vinyl acetate, polypropylene, polymethacrylate, polyethylene, polyethylene oxide, polysilicates, polycarbonates, Teflon, fluorocarbons, nylon, silicon rubber, polyanhydrides, polyglycolic acid, polyactic acid, polyorthoesters, functionalized silane, polypropylfumerate, collagen, glycosaminoglycans, polyamino acids, dextran, or any combination thereof. Solid supports further include thin film, membrane, bottles, dishes, fibers, woven fibers, shaped polymers such as tubes, particles, beads, microspheres, microparticles, or any combination thereof. For example, when solid surface is a bead, the bead can include, but is not limited to, a a ceramic bead, polystyrene bead, a polymer bead, a methylstyrene bead, an agarose bead, an acrylamide bead, a solid core bead, a porous bead, a paramagnetic bead, a glass bead, or a controlled pore bead. A bead may be spherical or an irregularly shaped. A bead's size may range from nanometers, e.g.. 100 nm, to millimeters, e.g., 1 mm. In certain embodiments, beads range in size from about 0.2 micron to about 200 microns, or from about 0.5 micron to about 5 micron. n some embodiments, beads can be about 1, 1.5, 2, 2.5, 2.8, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 15, or 20 μm in diameter. In certain embodiments, "a bead" solid support may refer to an individual bead or a plurality of beads.

As used herein, the term "nucleic acid molecule" or "polynucleotide" refers to a single- or double-stranded polynucleotide containing deoxyribonucleotides or ribonucleotides that are linked by 3'-5' phosphodiester bonds, as well as polynucleotide analogs. A nucleic acid molecule includes, but is not limited to, DNA, RNA, and cDNA. A polynucleotide analog may possess a backbone other than a standard phosphodiester linkage found in natural polynucleotides and, optionally, a modified sugar moiety or moieties other than ribose or deoxyribose. Polynucleotide analogs contain bases capable of hydrogen bonding by Watson-Crick base pairing to standard polynucleotide bases, where the analog backbone presents the bases in a manner to permit such hydrogen bonding in a sequence-specific fashion between the oligonucleotide analog molecule and bases in a standard polynucleotide. Examples of polynucleotide analogs include, but are not limited to xeno nucleic acid (XNA), bridged nucleic acid (BNA), glycol nucleic acid (GNA), peptide nucleic acids (PNAs), γPNAs, morpholino polynucleotides, locked nucleic acids (LNAs), threose nucleic acid (TNA), 2'-O-Methyl polynucleotides, 2'-O-alkyl ribosyl substituted polynucleotides, phosphorothioate polynucleotides, and boronophosphate polynucleotides. A polynucleotide analog may possess purine or pyrimidine analogs, including for example, 7-deaza purine analogs, 8-halopurine analogs, 5-halopyrimidine analogs, or universal base analogs that can pair with any base, including hypoxanthine, nitroazoles, isocarbostyril analogues, azole carboxamides, and aromatic triazole analogues, or base analogs with additional functionality, such as a biotin moiety for affinity binding.

As used herein, "nucleic acid sequencing" means the determination of the order of nucleotides in a nucleic acid molecule or a sample of nucleic acid molecules.

As used herein, "next generation sequencing" refers to high-throughput sequencing methods that allow the sequencing of millions to billions of molecules in parallel. Examples of next generation sequencing methods include sequencing by synthesis, sequencing by ligation, sequencing by hybridization, polony sequencing, ion semiconductor sequencing, and pyrosequencing. By attaching primers to a solid substrate and a complementary sequence to a nucleic acid molecule, a nucleic acid molecule can be hybridized to the solid substrate via the primer and then multiple copies can be generated in a discrete area on the solid substrate by using polymerase to amplify (these groupings are sometimes referred to as polymerase colonies or polonies). Consequently, during the sequencing process, a nucleotide at a particular position can be sequenced multiple times (*e.g.*, hundreds or thousands of times) – this depth of coverage is referred to as "deep sequencing." Examples of high throughput nucleic acid sequencing technology include platforms provided by Illumina, BGI, Qiagen, Thermo-Fisher, and Roche, including formats such as parallel bead arrays, sequencing by synthesis, sequencing by ligation, capillary electrophoresis, electronic microchips, "biochips," microarrays, parallel microchips, and single-molecule arrays, as reviewed by Service (*Science* 311:1544-1546, 2006).

As used herein, "single molecule sequencing" or "third generation sequencing" refers to next-generation sequencing methods wherein reads from single molecule sequencing instruments are generated by sequencing of a single molecule of DNA. Unlike next generation sequencing methods that rely on amplification to clone many DNA molecules in parallel for sequencing in a phased approach, single molecule sequencing interrogates single molecules of DNA and does not require amplification or synchronization. Single molecule sequencing includes methods that need to pause the sequencing reaction after each base incorporation ('wash-and-scan' cycle) and methods which do not need to halt between read steps. Examples of single molecule sequencing methods include single molecule real-time sequencing (Pacific Biosciences), nanopore-based sequencing (Oxford Nanopore), duplex interrupted nanopore sequencing, and direct imaging of DNA using advanced microscopy.

As used herein, "analyzing" the macromolecule means to quantify, characterize, distinguish, or a combination thereof, all or a portion of the components of the macromolecule. For example, analyzing a peptide, polypeptide, or protein includes determining all or a portion of the amino acid sequence (contiguous or non-continuous) of the peptide. Analyzing a macromolecule also includes partial identification of a component of the macromolecule. For example, partial identification of amino acids in the macromolecule protein sequence can identify an amino acid in the protein as belonging to a subset of possible amino acids. Analysis typically begins with analysis of the *n* NTAA, and then proceeds to the next amino acid of the peptide (*i.e.*, *n-1*, *n-2*, *n-3*, and so forth). This is accomplished by cleavage of the *n* NTAA, thereby converting the *n-1* amino acid of the peptide to an N-terminal amino acid (referred to herein as the "*n-1* NTAA"). Analyzing the peptide may also include determining the presence and frequency of post-translational modifications on the peptide, which may or may not include information regarding the sequential order of the post-translational modifications on the peptide. Analyzing the peptide may also include determining the presence and frequency of epitopes in the peptide, which may or may not include information regarding the sequential order or location of the epitopes within the peptide. Analyzing the peptide may include combining different types of analysis, for example obtaining epitope information, amino acid sequence information, post-translational modification information, or any combination thereof.

As used herein, the term “compartment” refers to a physical area or volume that separates or isolates a subset of macromolecules from a sample of macromolecules. For example, a compartment may separate an individual cell from other cells, or a subset of a sample’s proteome from the rest of the sample’s proteome. A compartment may be an aqueous compartment (*e.g.*, microfluidic droplet), a solid compartment (*e.g.*, picotiter well or microtiter well on a plate, tube, vial, gel bead), or a separated region on a surface. A compartment may comprise one or more beads to which macromolecules may be immobilized.

As used herein, the term “compartment tag” or “compartment barcode” refers to a single or double stranded nucleic acid molecule of about 4 bases to about 100 bases (including 4 bases, 100 bases, and any integer between) that comprises identifying information for the constituents (*e.g.*, a single cell’s proteome), within one or more compartments (*e.g.*, microfluidic droplet). A compartment barcode identifies a subset of macromolecules in a sample, e.g., a subset of protein sample, that have been separated into the same physical compartment or group of compartments from a plurality (e.g., millions to billions) of compartments. Thus, a compartment tag can be used to distinguish constituents derived from one or more compartments having the same compartment tag from those in another compartment having a different compartment tag, even after the constituents are pooled together. By labeling the proteins and/or peptides within each compartment or within a group of two or more compartments with a unique compartment tag, peptides derived from the same protein, protein complex, or cell within an individual compartment or group of compartments can be identified. A compartment tag comprises a barcode, which is optionally flanked by a spacer sequence on one or both sides, and an optional universal primer. The spacer sequence can be complementary to the spacer sequence of a recording tag, enabling transfer of compartment tag information to the recording tag. A compartment tag may also comprise a universal priming site, a unique molecular identifier (for providing identifying information for the peptide attached thereto), or both, particularly for embodiments where a compartment tag comprises a recording tag to be used in downstream peptide analysis methods described herein. A compartment tag can comprise a functional moiety (*e.g.*, aldehyde, NHS, mTet, alkyne, etc.) for coupling to a peptide. Alternatively, a compartment tag can comprise a peptide comprising a recognition sequence for a protein ligase to allow ligation of the compartment tag to a peptide of interest. A compartment can comprise a single compartment tag, a plurality of identical compartment tags save for an optional UMI sequence, or two or more different compartment tags. In certain embodiments each compartment comprises a unique compartment tag (one-to-one mapping). In other embodiments, multiple compartments from a larger population of compartments comprise the same compartment tag (many-to-one mapping). A compartment tag may be joined to a solid support within a compartment (*e.g.*, bead) or joined to the surface of the compartment itself (*e.g.*, surface of a picotiter well). Alternatively, a compartment tag may be free in solution within a compartment.

As used herein, the term “partition” refers to random assignment of a unique barcode to a subpopulation of macromolecules from a population of macromolecules within a sample. In certain embodiments, partitioning may be achieved by distributing macromolecules into compartments. A partition may be comprised of the macromolecules within a single compartment or the macromolecules within multiple compartments from a population of compartments.

As used herein, a “partition tag” or “partition barcode” refers to a single or double stranded nucleic acid molecule of about 4 bases to about 100 bases (including 4 bases, 100 bases, and any integer between) that comprises identifying information for a partition. In certain embodiments, a partition tag for a macromolecule refers to identical compartment tags arising from the partitioning of macromolecules into compartment(s) labeled with the same barcode.

As used herein, the term “fraction” refers to a subset of macromolecules (e.g., proteins) within a sample that have been sorted from the rest of the sample or organelles using physical or chemical separation methods, such as fractionating by size, hydrophobicity, isoelectric point, affinity, and so on. Separation methods include HPLC separation, gel separation, affinity separation, cellular fractionation, cellular organelle fractionation, tissue fractionation, etc. Physical properties such as fluid flow, magnetism, electrical current, mass, density, or the like can also be used for separation.

As used herein, the term “fraction barcode” refers to a single or double stranded nucleic acid molecule of about 4 bases to about 100 bases (including 4 bases, 100 bases, and any integer therebetween) that comprises identifying information for the macromolecules within a fraction.

III. Methods of Analysing Macromolecules

The methods described herein provide a highly-parallelized approach for macromolecule analysis. Highly multiplexed macromolecule binding assays are converted into a nucleic acid molecule library for readout by next generation sequencing. The methods provided herein are particularly useful for protein or peptide sequencing.

In a preferred embodiment, protein samples are labeled at the single molecule level with at least one nucleic acid recording tag that includes a barcode (e.g., sample barcode, compartment barcode) and an optional unique molecular identifier. The protein samples undergo proteolytic digest to produce a population of recording tag labeled peptides (e.g., millions to billions). These recording tag labeled peptides are pooled and immobilized randomly on a solid support (e.g., porous beads). The pooled, immobilized, recording tag labeled peptides are subjected to multiple, successive binding cycles, each binding cycle comprising exposure to a plurality of binding agents (e.g., binding agents for all twenty of the naturally occurring amino acids) that are labeled with coding tags comprising an encoder sequence that identifies the associated binding agent. During each binding cycle, information about the binding of a binding agent to the peptide is captured by transferring a binding agent’s coding tag information to the recording tag (or transferring the recording tag information to the coding tag or transferring both recording tag information and coding tag information to a separate di-tag construct). Upon completion of binding cycles, a library of extended recording tags (or extended coding tags or di-tag constructs) is generated that represents the binding histories of the assayed peptides, which can be analyzed using very high-throughput next generation digital sequencing methods. The use of nucleic acid barcodes in the recording tag allows deconvolution of a massive amount of peptide sequencing data, e.g., to identify which sample, cell, subset of proteome, or protein, a peptide sequence originated from.

In one aspect, a method for analysing a macromolecule is provided comprising: (a) providing a macromolecule and an associated or co-localized recording tag joined to a solid support; (b) contacting the macromolecule with a first binding agent capable of binding to the macromolecule, wherein the first binding agent comprises a first coding tag with identifying information regarding the first binding agent; (c) transferring the information of the first coding tag to the recording tag to generate a first order extended recording tag; (d) contacting the macromolecule with a second binding agent capable of binding to the macromolecule, wherein the second binding agent comprises a second coding tag with identifying information regarding the second binding agent; (e) transferring the information of the second coding tag is transferred to the first order extended recording tag to generate a second order extended recording tag; and (f) analysing the second order extended tag (*see, e.g.*, Figures 2A-D).

In certain embodiments, the contacting steps (b) and (d) are performed in sequential order, e.g., the first binding agent and the second binding agent are contacted with the macromolecule in separate binding cycle reactions. In other embodiments, the contacting steps (b) and (d) are performed at the same time, e.g., as in a single binding cycle reaction comprising the first binding agent, the second binding agent, and optionally additional binding agents. In a preferred embodiment, the contacting steps (b) and (d) each comprise contacting the macromolecule with a plurality of binding agents.

In certain embodiments, the method further comprises between steps (e) and (f) the following steps: (x) repeating steps (d) and (e) one or more times by replacing the second binding agent with a third (or higher order) binding agent capable of binding to the macromolecule, wherein the third (or higher order) binding agent comprises a third (or higher order) coding tag with identifying information regarding the third (or higher order) bind agent; and (y) transferring the information of the third (or higher order) coding tag to the second (or higher order) extended recording tag to generate a third (or higher order) extended recording tag; and (z) analysing the third (or higher order) extended recording tag.

The third (or higher order) binding agent may be contacted with the macromolecule in a separate binding cycle reaction from the first binding agent and the second binding agent. Alternatively, the third (or higher order) binding agent may be contacted with the macromolecule in a single binding cycle reaction with the first binding agent, and the second binding agent.

In a second aspect, a method for analyzing a macromolecule is provided comprising the steps of: (a) providing a macromolecule, an associated first recording tag and an associated second recording tag joined to a solid support; (b) contacting the macromolecule with a first binding agent capable of binding to the macromolecule, wherein the first binding agent comprises a first coding tag with identifying information regarding the first binding agent; (c) transferring the information of the first coding tag to the first recording tag to generate a first extended recording tag; (d) contacting the macromolecule with a second binding agent capable of binding to the macromolecule, wherein the second binding agent comprises a second coding tag with identifying information regarding the second binding agent; (e) transferring the information of the second coding tag to the second recording tag to generate a second extended recording tag; and (f) analyzing the first and second extended recording tags.

In certain embodiments, contacting steps (b) and (d) are performed in sequential order, e.g., the first binding agent and the second binding agent are contacted with the macromolecule in separate binding cycle reactions. In other embodiments, contacting steps (b) and (d) are performed at the same time, e.g., as in a single binding cycle reaction comprising the first binding agent, the second binding agent, and optionally additional binding agents.

In certain embodiments, step (a) further comprises providing an associated third (or higher order) recording tag joined to the solid support. In further embodiments, the method further comprises, between steps (e) and (f), the following steps: (x) repeating steps (d) and (e) one or more times by replacing the second binding agent with a third (or higher order) binding agent capable of binding to the macromolecule, wherein the third (or higher order) binding agent comprises a third (or higher order) coding tag with identifying information regarding the third (or higher order) bind agent; and (y) transferring the information of the third (or higher order) coding tag to the third (or higher order) recording tag to generate a third (or higher order) extended recording tag; and (z) analysing the first, second and third (or higher order) extended recording tags.

The third (or higher order) binding agent may be contacted with the macromolecule in a separate binding cycle reaction from the first binding agent and the second binding agent. Alternatively, the third (or higher order) binding agent may be contacted with the macromolecule in a single binding cycle reaction with the first binding agent, and the second binding agent.

In certain embodiments, the first coding tag, second coding tag, and any higher order coding tags each have a binding cycle specific sequence.

In a third aspect, a method of analyzing a peptide is provided comprising the steps of: (a) providing a peptide and an associated recording tag joined to a solid support; (b) modifying the N-terminal amino acid (NTAA) of the peptide with a chemical moiety to produce a modified NTAA; (c) contacting the peptide with a first binding agent capable of binding to the modified NTAA, wherein the first binding agent comprises a first coding tag with identifying information regarding the first binding agent; (d) transferring the information of the first coding tag to the recording tag to generate an extended recording tag; and (e) analyzing the extended recording tag (*see, e.g.* Figure 3).

In certain embodiments, step (c) further comprises contacting the peptide with a second (or higher order) binding agent comprising a second (or higher order) coding tag with identifying information regarding the second (or higher order) binding agent, wherein the second (or higher order) binding agent is capable of binding to a modified NTAA other than the modified NTAA of step (b). In further embodiments, contacting the peptide with the second (or higher order) binding agent occurs in sequential order following the peptide being contacted with the first binding agent, e.g., the first binding agent and the second (or higher order) binding agent are contacted with the peptide in separate binding cycle reactions. In other embodiments, contacting the peptide with the second (or higher order) binding agent occurs simultaneously with the peptide being contacted with the first binding agent, e.g., as in a single binding cycle reaction comprising the first binding agent and the second (or higher order) binding agent).

In certain embodiments, the chemical moiety is add to the NTAA via chemical reaction or enzymatic reaction.

In certain embodiments, the chemical moiety used for modifying the NTAA is a phenylthiocarbamoyl (PTC), dinitrophenol (DNP) moiety; a sulfonyloxynitrophenyl (SNP) moiety, a dansyl moiety; a 7-methoxy coumarin moiety; a thioacyl moiety; a thioacetyl moiety; an acetyl moiety; a guanidnyl moiety; or a thiobenzyl moiety.

A chemical moiety may be added to the NTAA using a chemical agent. In certain embodiments, the chemical agent for modifying an NTAA with a PTC moiety is a phenyl isothiocyanate or derivative thereof; the chemical agent for modifying an NTAA with a DNP moiety is 2,4-dinitrobenzenesulfonic acid (DNBS) or an aryl halide such as 1-Fluoro-2,4-dinitrobenzene (DNFB); the chemical agent for modifying an NTAA with a sulfonyloxynitrophenyl (SNP) moiety is 4-sulfonyl-2-nitrofluorobenzene (SNFB); the chemical agent for modifying an NTAA with a dansyl group is a sulfonyl chloride such as dansyl chloride; the chemical agent for modifying an NTAA with a 7-methoxy coumarin moiety is 7-methoxycoumarin acetic acid (MCA); the chemical agent for modifying an NTAA with a thioacyl moiety is a thioacylation reagent; the chemical agent for modifying an NTAA with a thioacetyl moiety is a thioacetylation reagent; the chemical agent for modifying an NTAA with an acetyl moiety is an acetylating reagent (e.g., acetic anhydride); the chemical agent for modifying an NTAA with a guanidnyl (amidinyl) moiety is a guanidinylating reagent, or the chemical agent for modifying an NTAA with a thiobenzyl moiety is a thiobenzylation reagent.

In a fourth aspect the present disclosure provides, a method for analyzing a peptide is provided comprising the steps of: (a) providing a peptide and an associated recording tag joined to a solid support; (b) modifying the N-terminal amino acid (NTAA) of the peptide with a chemical moiety to produce a modified NTAA; (c) contacting the peptide with a first binding agent capable of binding to the modified NTAA, wherein the first binding agent comprises a first coding tag with identifying information regarding the first binding agent; (d) transferring the information of the first coding tag to the recording tag to generate a first extended recording tag; (e) removing the modified NTAA to expose a new NTAA; (f) modifying the new NTAA of the peptide with a chemical moiety to produce a newly modified NTAA; (g) contacting the peptide with a second binding agent capable of binding to the newly modified NTAA, wherein the second binding agent comprises a second coding tag with identifying information regarding the second binding agent; (h) transferring the information of the second coding tag to the first extended recording tag to generate a second extended recording tag; and (i) analyzing the second extended recording tag.

In certain embodiments, the contacting steps (c) and (g) are performed in sequential order, e.g., the first binding agent and the second binding agent are contacted with the peptide in separate binding cycle reactions.

In certain embodiments, the method further comprises between steps (h) and (i) the following steps: (x) repeating steps (e), (f), and (g) one or more times by replacing the second binding agent with a third (or higher order) binding agent capable of binding to the modified NTAA, wherein the third (or higher order) binding agent comprises a third (or higher order) coding tag with identifying information regarding the third (or higher order) bind agent; and (y) transferring the information of the third (or higher order) coding tag to the second (or higher order) extended recording tag to generate a third (or higher order) extended recording tag; and (z) analysing the third (or higher order) extended recording tag.

In certain embodiments, the chemical moiety is add to the NTAA via chemical reaction or enzymatic reaction.

In certain embodiments, the chemical moiety is a phenylthiocarbamoyl (PTC), dinitrophenol (DNP) moiety; a sulfonyloxynitrophenyl (SNP) moiety, a dansyl moiety; a 7-methoxy coumarin moiety; a thioacyl moiety; a thioacetyl moiety; an acetyl moiety; a guanyl moiety; or a thiobenzyl moiety.

A chemical moiety may be added to the NTAA using a chemical agent. In certain embodiments, the chemical agent for modifying an NTAA with a PTC moiety is a phenyl isothiocyanate or derivative thereof; the chemical agent for modifying an NTAA with a DNP moiety is 2,4-dinitrobenzenesulfonic acid (DNBS) or an aryl halide such as 1-Fluoro-2,4-dinitrobenzene (DNFB); the chemical agent for modifying an NTAA with a sulfonyloxynitrophenyl (SNP) moiety is 4-sulfonyl-2-nitrofluorobenzene (SNFB); the chemical agent for modifying an NTAA with a dansyl group is a sulfonyl chloride such as dansyl chloride; the chemical reagent for modifying an NTAA with a 7-methoxy coumarin moiety is 7-methoxycoumarin acetic acid (MCA); the chemical agent for modifying an NTAA with a thioacyl moiety is a thioacylation reagent; the chemical agent for modifying an NTAA with a thioacetyl moiety is a thioacetylation reagent; the chemical agent for modifying an NTAA with an acetyl moiety is an acetylating agent (e.g., acetic anhydride); the chemical agent for modifying an NTAA with a guanyl moiety is a guanidinylating reagent, or the chemical agent for modifying an NTAA with a thiobenzyl moiety is a thiobenzylation reagent.

In a fifth aspect, a method for analyzing a peptide is provided comprising the steps of: (a) providing a peptide and an associated recording tag joined to a solid support; (b) contacting the peptide with a first binding agent capable of binding to the N-terminal amino acid (NTAA) of the peptide, wherein the first binding agent comprises a first coding tag with identifying information regarding the first binding agent; (c) transferring the information of the first coding tag to the recording tag to generate an extended recording tag; and (d) analyzing the extended recording tag.

In certain embodiments, step (b) further comprises contacting the peptide with a second (or higher order) binding agent comprising a second (or higher order) coding tag with identifying information regarding the second (or higher order) binding agent, wherein the second (or higher order) binding agent is capable of binding to a NTAA other than the NTAA of the peptide. In further embodiments, the contacting the peptide with the second (or higher order) binding agent occurs in sequential order following the peptide being contacted with the first binding agent, e.g., the first binding agent and the second (or higher order) binding agent are contacted with the peptide in separate binding cycle reactions. In other embodiments, the contacting the peptide with the second (or higher order) binding agent occurs at the same time as the peptide the being contacted with first binding agent, e.g., as in a single binding cycle reaction comprising the first binding agent and the second (or higher order) binding agent.

In a sixth aspect, a method for analyzing a peptide is provided, comprising the steps of: (a) providing a peptide and an associated recording tag joined to a solid support; (b) contacting the peptide with a first binding agent capable of binding to the N-terminal amino acid (NTAA) of the peptide, wherein the first binding agent comprises a first coding tag with identifying information regarding the first binding agent; (c) transferring the information of the first coding tag to the recording tag to generate a first extended recording tag; (d) removing the NTAA to expose a new NTAA of the peptide; (e) contacting the peptide with a second binding agent capable of binding to the new NTAA, wherein the second binding agent comprises a second coding tag with identifying information regarding the second binding agent; (f) transferring the information of the second coding tag to the first extended recording tag to generate a second extended recording tag; and (g) analyzing the second extended recording tag.

In certain embodiments, the method further comprises between steps (f) and (g) the following steps: (x) repeating steps (d), (e), and (f) one or more times by replacing the second binding agent with a third (or higher order) binding agent capable of binding to the macromolecule, wherein the third (or higher order) binding agent comprises a third (or higher order) coding tag with identifying information regarding the third (or higher order) bind agent; and (y) transferring the information of the third (or higher order) coding tag to the second (or higher order) extended recording tag to generate a third (or higher order) extended recording tag; and wherein the third (or higher order) extended recording tag is analyzed in step (g).

In certain embodiments, the contacting steps (b) and (e) are performed in sequential order, e.g., the first binding agent and the second binding agent are contacted with the peptide in separate binding cycle reactions.

In any of the embodiments provided herein, the methods comprise analyzing a plurality of macromolecules in parallel. In a preferred embodiment, the methods comprise analyzing a plurality of peptides in parallel.

In any of the embodiments provided herein, the step of contacting a macromolecule (or peptide) with a binding agent comprises contacting the macromolecule (or peptide) with a plurality of binding agents.

In any of the embodiments provided herein, the macromolecule may be a protein, polypeptide, or peptide. In further embodiments, the peptide may be obtained by fragmenting a protein or polypeptide from a biological sample.

In any of theembodiments provided herein, the macromolecule may be or comprise a carbohydrate, lipid, nucleic acid, or macrocycle.

In any of the embodiments provided herein, the recording tag may be a DNA molecule, a DNA molecule with modified bases, an RNA molecule, a BNA, molecule, a XNA molecule, an LNA molecule, a PNA molecule, a γPNA molecule (Dragulescu-Andrasi et al., 2006, J. Am. Chem. Soc. 128:10258-10267), a GNA molecule, or any combination thereof.

In any of the embodiments provided herein, the recording tag may comprise a universal priming site. In further embodiments, the universal priming site comprises a priming site for amplification, ligation, sequencing, or a combination thereof.

In any of the embodiments provided herein, the recording tag may comprise a unique molecular identifier, a compartment tag, a partition barcode, sample barcode, a fraction barcode, a spacer sequence, or any combination thereof.

In any of the embodiments provided herein, the coding tag may comprise a unique molecular identifier (UMI), an encoder sequence, a binding cycle specific sequence, a spacer sequence, or any combination thereof.

In any of the embodiments provided herein, the binding cycle specific sequence in the coding tag may be a binding cycle-specific spacer sequence.

In certain embodiments, a binding cycle specific sequence is encoded as a separate barcode from the encoder sequence. In other embodiments, the encoder sequence and binding cycle specific sequence is set forth in a single barcode that is unique for the binding agent and for each cycle of binding.

In certain embodiments, the spacer sequence comprises a common binding cycle sequence that is shared among binding agents from the multiple binding cycles. In other embodiments, the spacer sequence comprises a unique binding cycle sequence that is shared among binding agents from the same binding cycle.

In any of the embodiments provided herein, the recording tag may comprise a barcode.

In any of the embodiments provided herein, the macromolecule and the associated recording tag(s) may be covalently joined to the solid support.

In any of the embodiments provided herein, the solid support may be a bead, a porous bead, a porous matrix, an expandable gel bead or matrix, an array, a glass surface, a silicon surface, a plastic surface, a filter, a membrane, nylon, a silicon wafer chip, a flow through chip, a biochip including signal transducing electronics, a microtiter well, an ELISA plate, a spinning interferometry disc, a nitrocellulose membrane, a nitrocellulose-based polymer surface, a nanoparticle, or a microsphere.

In any of the embodiments provided herein, the solid support may be a polystyrene bead, a polymer bead, an agarose bead, an acrylamide bead, a solid core bead, a porous bead, a paramagnetic bead, glass bead, or a controlled pore bead.

In any of the embodiments provided herein, a plurality of macromolecules and associated recording tags may be joined to a solid support. In further embodiments, the plurality of macromolecules are spaced apart on the solid support at an average distance $\geq$ 50 nm, $\geq$ 100 nm, or $\geq$ 200 nm.

In any of the embodiments provided herein, the binding agent may be a polypeptide or protein. In further embodiments, the binding agent is a modified or variant aminopeptidase, a modified or variant amino acyl tRNA synthetase, a modified or variant anticalin, or a modified or variant ClpS.

In any of the embodiments provided herein, the binding agent may be capable of selectively binding to the macromolecule.

In any of the embodiments provided herein, the coding tag may be a DNA molecule, DNA molecule with modified bases, an RNA molecule, a BNA molecule, an XNA molecule, a LNA molecule, a GNA molecule, a PNA molecule, a γPNA molecule, or a combination thereof.

In any of the embodiments provided herein, the binding agent and the coding tag may be joined by a linker.

In any of the embodiments provided herein, the binding agent and the coding tag may be joined by a SpyTag/SpyCatcher or SnoopTag/SnoopCatcher peptide-protein pair (Zakeri, et al., 2012, Proc Natl Acad Sci U S A 109(12): E690-697; Veggiani et al., 2016, Proc. Natl. Acad. Sci. USA 113:1202-1207, each of which is incorporated by reference in its entirety).

In any of the embodiments provided herein, the transferring of information of the coding tag to the recording tag is mediated by a DNA ligase. Alternatively, the transferring of information of the coding tag to the recording tag is mediated by a DNA polymerase or chemical ligation.

In any of the embodiments provided herein, analyzing the extended recording tag may comprise nucleic acid sequencing. In further embodiments, nucleic acid sequencing is sequencing by synthesis, sequencing by ligation, sequencing by hybridization, polony sequencing, ion semiconductor sequencing, or pyrosequencing. In other embodiments, nucleic acid sequencing is single molecule real-time sequencing, nanopore-based sequencing, nanogap tunneling sequencing, or direct imaging of DNA using advanced microscopy.

In any of the embodiments provided herein, the extended recording tag may be amplified prior to analysis.

In any of the embodiments provided herein, the order of the coding tag information contained on the extended recording tag may provide information regarding the order of binding by the binding agents to the macromolecule and thus, the sequence of analytes detected by the binding agents.

In any of the embodiments provided herein, the frequency of a particular coding tag information (e.g., encoder sequence) contained on the extended recording tag may provide information regarding the frequency of binding by a particular binding agent to the macromolecule and thus, the frequency of the analyte in the macromolecule detected by the binding agent.

In any of the embodiments disclosed herein, multiple macromolecule (e.g., protein) samples, wherein a population of macromolecules within each sample are labeled with recording tags comprising a sample specific barcode, can be pooled. Such a pool of macromolecule samples may be subjected to binding cycles within a single-reaction tube.

In any of the embodiments provided herein, the plurality of extended recording tags representing a plurality of macromolecules may be analyzed in parallel.

In any of the embodiments provided herein, the plurality of extended recording tags representing a plurality of macromolecules may be analyzed in a multiplexed assay.

In any of the embodiments provided herein, the plurality of extended recording tags may undergo a target enrichment assay prior to analysis.

In any of the embodiments provided herein, the plurality of extended recording tags may undergo a subtraction assay prior to analysis.

In any of the embodiments provided herein, the plurality of extended recording tags may undergo a normalization assay to reduce highly abundant species prior to analysis.

In any of the embodiments provided herein, the NTAA may be removed by a modified aminopeptidase, a modified amino acid tRNA synthetase, a mild Edman degradation, an Edmanase enzyme, or anhydrous TFA.

In any of the embodiments provided herein, at least one binding agent may bind to a terminal amino acid residue. In certain embodiments the terminal amino acid residue is an N-terminal amino acid or a C-terminal amino acid.

In any of the embodiments described herein, at least one binding agent may bind to a post-translationally modified amino acid.

Features of the aforementioned embodiments are provided in further detail in the following sections.

IV. Macromolecules

In one aspect, the present disclosure relates to the analysis of macromolecules. A macromolecule is a large molecule composed of smaller subunits. In certain embodiments, a macromolecule is a protein, a protein complex, polypeptide, peptide, nucleic acid molecule, carbohydrate, lipid, macrocycle, or a chimeric macromolecule.

A macromolecule (*e.g.*, protein, polypeptide, peptide) analyzed according the methods disclosed herein may be obtained from a suitable source or sample, including but not limited to: biological samples, such as cells (both primary cells and cultured cell lines), cell lysates or extracts, cell organelles or vesicles, including exosomes, tissues and tissue extracts; biopsy; fecal matter; bodily fluids (such as blood, whole blood, serum, plasma, urine, lymph, bile, cerebrospinal fluid, interstitial fluid, aqueous or vitreous humor, colostrum, sputum, amniotic fluid, saliva, anal and vaginal secretions, perspiration and semen, a transudate, an exudate (*e.g.*, fluid obtained from an abscess or any other site of infection or inflammation) or fluid obtained from a joint (normal joint or a joint affected by disease such as rheumatoid arthritis, osteoarthritis, gout or septic arthritis) of virtually any organism, with mammalian-derived samples, including microbiome-containing samples, being preferred and human-derived samples, including microbiome-containing samples, being particularly preferred; environmental samples (such as air, agricultural, water and soil samples); microbial samples including samples derived from microbial biofilms and/or communities, as well as microbial spores; research samples including extracellular fluids, extracellular supernatants from cell cultures, inclusion bodies in bacteria, cellular compartments including mitochondrial compartments, and cellular periplasm.

In certain embodiments, a macromolecule is a protein, a protein complex, a polypeptide, or peptide. Amino acid sequence information and post-translational modifications of a peptide, polypeptide, or protein are transduced into a nucleic acid encoded library that can be analyzed via next generation sequencing methods. A peptide may comprise L-amino acids, D-amino acids, or both. A peptide, polypeptide, protein, or protein complex may comprise a standard, naturally occurring amino acid, a modified amino acid (e.g., post-translational modification), an amino acid analog, an amino acid mimetic, or any combination thereof. In some embodiments, a peptide, polypeptide, or protein is naturally occurring, synthetically produced, or recombinantly expressed. In any of the aforementioned peptide embodiments, a peptide, polypeptide, protein, or protein complex may further comprise a post-translational modification.

Standard, naturally occurring amino acids include Alanine (A or Ala), Cysteine (C or Cys), Aspartic Acid (D or Asp), Glutamic Acid (E or Glu), Phenylalanine (F or Phe), Glycine (G or Gly), Histidine (H or His), Isoleucine (I or Ile), Lysine (K or Lys), Leucine (L or Leu), Methionine (M or Met), Asparagine (N or Asn), Proline (P or Pro), Glutamine (Q or Gln), Arginine (R or Arg), Serine (S or Ser), Threonine (T or Thr), Valine (V or Val), Tryptophan (W or Trp), and Tyrosine (Y or Tyr). Non-standard amino acids include selenocysteine, pyrrolysine, and N-formylmethionine, β-amino acids, Homo-amino acids, Proline and Pyruvic acid derivatives, 3-substituted Alanine derivatives, Glycine derivatives, Ring-substituted Phenylalanine and Tyrosine Derivatives, Linear core amino acids, and N-methyl amino acids.

A post-translational modification (PTM) of a peptide, polypeptide, or protein may be a covalent modification or enzymatic modification. Examples of post-translation modifications include, but are not limited to, acylation, acetylation, alkylation (including methylation), biotinylation, butyrylation, carbamylation, carbonylation, deamidation, deiminiation, diphthamide formation, disulfide bridge formation, eliminylation, flavin attachment, formylation, gamma-carboxylation, glutamylation, glycylation, glycosylation (e.g., N-linked, O-linked, C-linked, phosphoglycosylation), glypiation, heme C attachment, hydroxylation, hypusine formation, iodination, isoprenylation, lipidation, lipoylation, malonylation, methylation, myristolylation, oxidation, palmitoylation, pegylation, phosphopantetheinylation, phosphorylation, prenylation, propionylation, retinylidene Schiff base formation, S-glutathionylation, S-nitrosylation, S-sulfenylation, selenation, succinylation, sulfination, ubiquitination, and C-terminal amidation. A post-translational modification includes modifications of the amino terminus and/or the carboxyl terminus of a peptide, polypeptide, or protein. Modifications of the terminal amino group include, but are not limited to, des-amino, N-lower alkyl, N-di-lower alkyl, and N-acyl modifications. Modifications of the terminal carboxy group include, but are not limited to, amide, lower alkyl amide, dialkyl amide, and lower alkyl ester modifications (*e.g.*, wherein lower alkyl is C_1-C_4 alkyl). A post-translational modification also includes modifications, such as but not limited to those described above, of amino acids falling between the amino and carboxy termini of a peptide, polypeptide, or protein. Post-translational modification can regulate a protein's "biology" within a cell, e.g., its activity, structure, stability, or localization. Phosphorylation is the most common post-translational modification and plays an important role in regulation of protein, particularly in cell signaling (Prabakaran et al., 2012, Wiley Interdiscip Rev Syst Biol Med 4: 565-583). The addition of sugars to proteins, such as glycosylation, has been shown to promote protein folding, improve stability, and modify regulatory function. The attachment of lipids to proteins enables targeting to the cell membrane.A post-translational modification can also include peptide, polypeptide, or protein modifications to include one or more detectable labels.

In certain embodiments, a peptide, polypeptide, or protein can be fragmented. For example, the fragmented peptide can be obtained by fragmenting a protein from a sample, such as a biological sample. The peptide, polypeptide, or protein can be fragmented by any means known in the art, including fragmentation by a protease or endopeptidase. In some embodiments, fragmentation of a peptide, polypeptide, or protein is targeted by use of a specific protease or endopeptidase. A specific protease or endopeptidase binds and cleaves at a specific consensus sequence (*e.g.*, TEV protease which is specific for ENLYFQ\S consensus sequence). In other embodiments, fragmentation of a peptide, polypeptide, or protein is non-targeted or random by use of a non-specific protease or endopeptidase. A non-specific protease may bind and cleave at a specific amino acid residue rather than a consensus sequence (*e.g.*, proteinase K is a non-specific serine protease). Proteinases and endopeptidases are well known in the art, and examples of such that can be used to cleave a protein or polypeptide into smaller peptide fragments include proteinase K, trypsin, chymotrypsin, pepsin, thermolysin, thrombin, Factor Xa, furin, endopeptidase, papain, pepsin, subtilisin, elastase, enterokinase, Genenase™ I, Endoproteinase LysC, Endoproteinase AspN, Endoproteinase GluC, etc. (Granvogl et al., 2007, Anal Bioanal Chem 389: 991-1002). In certain embodiments, a peptide, polypeptide, or protein is fragmented by proteinase K, or optionally, a thermolabile version of proteinase K to enable rapid inactivation. Proteinase K is quite stable in denaturing reagents, such as urea and SDS, enabling digestion of completely denatured proteins. Protein and polypeptide fragmentation into peptides can be performed before or after attachment of a DNA tag or DNA recording tag.

Chemical reagents can also be used to digest proteins into peptide fragments. A chemical reagent may cleave at a specific amino acid residue (e.g., cyanogen bromide hydrolyzes peptide bonds at the C-terminus of methionine residues). Chemical reagents for fragmenting polypeptides or proteins into smaller peptides include cyanogen bromide (CNBr), hydroxylamine, hydrazine, formic acid, BNPS-skatole [2-(2-nitrophenylsulfenyl)-3-methylindole], iodosobenzoic acid, •NTCB +Ni (2-nitro-5-thiocyanobenzoic acid), etc.

In certain embodiments, following enzymatic or chemical cleavage, the resulting peptide fragments are approximately the same desired length, e.g., from about 10 amino acids to about 70 amino acids, from about 10 amino acids to about 60 amino acids, from about 10 amino acids to about 50 amino acids, about 10 to about 40 amino acids, from about 10 to about 30 amino acids, from about 20 amino acids to about 70 amino acids, from about 20 amino acids to about 60 amino acids, from about 20 amino acids to about 50 amino acids, about 20 to about 40 amino acids, from about 20 to about 30 amino acids, from about 30 amino acids to about 70 amino acids, from about 30 amino acids to about 60 amino acids, from about 30 amino acids to about 50 amino acids, or from about 30 amino acids to about 40 amino acids. A cleavage reaction may be monitored, preferably in real time, by spiking the protein or polypeptide sample with a short test FRET (fluorescence resonance energy transfer) peptide comprising a peptide sequence containing a proteinase or endopeptidase cleavage site. In the intact FRET peptide, a fluorescent group and a quencher group are attached to either end of the peptide sequence containing the cleavage site, and fluorescence resonance energy transfer between the quencher and the fluorophore leads to low fluorescence. Upon cleavage of the test peptide by a protease or endopeptidase, the quencher and fluorophore are separated giving a large increase in fluorescence. A cleavage reaction can be stopped when a certain fluorescence intensity is achieved, allowing a reproducible cleavage end point to be achieved.

A sample of macromolecules (*e.g.*, peptides, polypeptides, or proteins) can undergo protein fractionation methods prior to attachment to a solid support, where proteins or peptides are separated by one or more properties such as cellular location, molecular weight, hydrophobicity, or isoelectric point, or protein enrichment methods. Alternatively, or additionally, protein enrichment methods may be used to select for a specific protein or peptide (*see*, *e.g.*, Whiteaker et al., 2007, Anal. Biochem. 362:44-54, incorporated by reference in its entirety) or to select for a particular post translational modification (*see, e.g.*, Huang et al., 2014. J. Chromatogr. A 1372:1-17, incorporated by reference in its entirety). Alternatively, a particular class or classes of proteins such as immunoglobulins, or immunoglobulin (Ig) isotypes such as IgG, can be affinity enriched or selected for analysis. In the case of immunoglobulin molecules, analysis of the sequence and abundance or frequency of hypervariable sequences involved in affinity binding are of particular interest, particularly as they vary in response to disease progression or correlate with healthy, immune, and/or or disease phenotypes. Overly abundant proteins can also be subtracted from the sample using standard immunoaffinity methods. Depletion of abundant proteins can be useful for plasma samples where over 80% of the protein constituent is albumin and immunoglobulins. Several commercial products are available for depletion of plasma samples of overly abundant proteins, such as PROTIA and PROT20 (Sigma-Aldrich).

In certain embodiments, the macromolecule is comprised of a protein or polypeptide. In one embodiment, the protein or polypeptide is labeled with DNA recording tags through standard amine coupling chemistries (*see, e.g.*, Figures 2B, 2C, 28, 29, 31, 40). The ε-amino group (e.g., of lysine residues) and the N-terminal amino group are particularly susceptible to labeling with amine-reactive coupling agents, depending on the pH of the reaction (Mendoza and Vachet 2009). In a particular embodiment (see, e.g., Figure 2B and Figure 29), the recording tag is comprised of a reactive moiety (e.g., for conjugation to a solid surface, a multifunctional linker, or a macromolecule), a linker, a universal priming sequence, a barcode (*e.g.*, compartment tag, partition barcode, sample barcode, fraction barcode, or any combination thereof), an optional UMI, and a spacer (Sp) sequence for facilitating information transfer to/from a coding tag. In another embodiment, the protein can be first labeled with a universal DNA tag, and the barcode-Sp sequence (representing a sample, a compartment, a physical location on a slide, etc.) are attached to the protein later through and enzymatic or chemical coupling step. (*see, e.g.*, Figures 20, 30, 31, 40). A universal DNA tag comprises a short sequence of nucleotides that are used to label a protein or polypeptide macromolecule and can be used as point of attachment for a barcode (*e.g.*, compartment tag, recording tag, etc.). For example, a recording tag may comprise at its terminus a sequence complementary to the universal DNA tag. In certain embodiments, a universal DNA tag is a universal priming sequence. Upon hybridization of the universal DNA tags on the labeled protein to complementary sequence in recording tags (*e.g.*, bound to beads), the annealed universal DNA tag may be extended via primer extension, transferring the recording tag information to the DNA tagged protein. In a particular embodiment, the protein is labeled with a universal DNA tag prior to proteinase digestion into peptides. The universal DNA tags on the labeled peptides from the digest can then be converted into an informative and effective recording tag.

In certain embodiments, a protein macromolecule can be immobilized to a solid support by an affinity capture reagent (and optionally covalently crosslinked), wherein the recording tag is associated with the affinity capture reagent directly, or alternatively, the protein can be directly immobilized to the solid support with a recording tag (*see, e.g.*, Figure 2C).

V. Solid Support

Macromolecules of the present disclosure are joined to a surface of a solid support (also referred to as "substrate surface"). The solid support can be any porous or non-porous support surface including, but not limited to, a bead, a microbead, an array, a glass surface, a silicon surface, a plastic surface, a filter, a membrane, nylon, a silicon wafer chip, a flow cell, a flow through chip, a biochip including signal transducing electronics, a microtiter well, an ELISA plate, a spinning interferometry disc, a nitrocellulose membrane, a nitrocellulose-based polymer surface, a nanoparticle, or a microsphere. Materials for a solid support include but are not limited to acrylamide, agarose, cellulose, nitrocellulose, glass, gold, quartz, polystyrene, polyethylene vinyl acetate, polypropylene, polymethacrylate, polyethylene, polyethylene oxide, polysilicates, polycarbonates, Teflon, fluorocarbons, nylon, silicon rubber, polyanhydrides, polyglycolic acid, polyactic acid, polyorthoesters, functionalized silane, polypropylfumerate, collagen, glycosaminoglycans, polyamino acids, or any combination thereof. Solid supports further include thin film, membrane, bottles, dishes, fibers, woven fibers, shaped polymers such as tubes, particles, beads, microparticles, or any combination thereof. For example, when solid surface is a bead, the bead can include, but is not limited to, a polystyrene bead, a polymer bead, an agarose bead, an acrylamide bead, a solid core bead, a porous bead, a paramagnetic bead, glass bead, or a controlled pore bead.

In certain embodiments, a solid support is a flow cell. Flow cell configurations may vary among different next generation sequencing platforms. For example, the Illumina flow cell is a planar optically transparent surface similar to a microscope slide, which contains a lawn of oligonucleotide anchors bound to its surface. Template DNA, comprise adapters ligated to the ends that are complimentary to oligonucleotides on the flow cell surface. Adapted single-stranded DNAs are bound to the flow cell and amplified by solid-phase "bridge" PCR prior to sequencing. The 454 flow cell (454 Life Sciences) supports a "picotiter" plate, a fiber optic slide with ~1.6 million 75-picoliter wells. Each individual molecule of sheared template DNA is captured on a separate bead, and each bead is compartmentalized in a private droplet of aqueous PCR reaction mixture within an oil emulsion. Template is clonally amplified on the bead surface by PCR, and the template-loaded beads are then distributed into the wells of the picotiter plate for the sequencing reaction, ideally with one or fewer beads per well. SOLiD (Supported Oligonucleotide Ligation and Detection) instrument from Applied Biosystems, like the 454 system, amplifies template molecules by emulsion PCR. After a step to cull beads that do not contain amplified template, bead-bound template is deposited on the flow cell. A flow cell may also be a simple filter frit, such as a TWIST™ DNA synthesis column (Glen Research).

In certain embodiments, a solid support is a bead, which may refer to an individual bead or a plurality of beads. In some embodiments, the bead is compatible with a selected next generation sequencing platform that will be used for downstream analysis (*e.g.*, SOLiD or 454). In some embodiments, a solid support is an agarose bead, a paramagnetic bead, a polystyrene bead, a polymer bead, an acrylamide bead, a solid core bead, a porous bead, a glass bead, or a controlled pore bead. In further embodiments, a bead may be coated with a binding functionality (*e.g.*, amine group, affinity ligand such as streptavidin for binding to biotin labeled macromolecule, antibody) to facilitate binding to a macromolecule.

Proteins, polypeptides, or peptides can be joined to the solid support, directly or indirectly, by any means known in the art, including covalent and non-covalent interactions, or any combination thereof (*see, e.g.*, Chan et al., 2007, PLoS One 2:e1164; Cazalis et al., Bioconj. Chem. 15:1005-1009; Soellner et al., 2003, J. Am. Chem. Soc. 125:11790-11791; Sun et al., 2006, Bioconjug. Chem. 17-52-57; Decreau et al., 2007, J. Org. Chem. 72:2794-2802; Camarero et al., 2004, J. Am. Chem. Soc. 126:14730-14731; Girish et al., 2005, Bioorg. Med. Chem. Lett. 15:2447-2451; Kalia et al., 2007, Bioconjug. Chem. 18:1064-1069; Watzke et al., 2006, Angew Chem. Int. Ed.

Engl. 45:1408-1412; Parthasarathy et al., 2007, Bioconjugate Chem. 18:469-476; and Bioconjugate Techniques, G. T. Hermanson, Academic Press (2013), and are each hereby incorporated by reference in their entirety). For example, the peptide may be joined to the solid support by a ligation reaction. Alternatively, the solid support can include an agent or coating to facilitate joining, either direct or indirectly, the peptide to the solid support. Any suitable molecule or materials may be employed for this purpose, including proteins, nucleic acids, carbohydrates and small molecules. For example, in one embodiment the agent is an affinity molecule. In another example, the agent is an azide group, which group can react with an alkynyl group in another molecule to facilitate association or binding between the solid support and the other molecule.

Proteins, polypeptides, or peptides can be joined to the solid support using methods referred to as "click chemistry." For this purpose any reaction which is rapid and substantially irreversible can be used to attach proteins, polypeptides, or peptides to the solid support. Exemplary reactions include the copper catalyzed reaction of an azide and alkyne to form a triazole (Huisgen 1, 3-dipolar cycloaddition), strain-promoted azide alkyne cycloaddition (SPAAC), reaction of a diene and dienophile (Diels-Alder), strain-promoted alkyne-nitrone cycloaddition, reaction of a strained alkene with an azide, tetrazine or tetrazole, alkene and azide [3+2] cycloaddition, alkene and tetrazine inverse electron demand Diels-Alder (IEDDA) reaction (e.g., m-tetrazine (mTet) and trans-cyclooctene (TCO)), alkene and tetrazole photoreaction, Staudinger ligation of azides and phosphines, and various displacement reactions, such as displacement of a leaving group by nucleophilic attack on an electrophilic atom (Horisawa 2014, Knall, Hollauf et al. 2014). Exemplary displacement reactions include reaction of an amine with: an activated ester; an N-hydroxysuccinimide ester; an isocyanate; an isothioscyanate or the like.

In some embodiments the macromolecule and solid support are joined by a functional group capable of formation by reaction of two complementary reactive groups, for example a functional group which is the product of one of the foregoing "click" reactions. In various embodiments, functional group can be formed by reaction of an aldehyde, oxime, hydrazone, hydrazide, alkyne, amine, azide, acylazide, acylhalide, nitrile, nitrone, sulfhydryl, disulfide, sulfonyl halide, isothiocyanate, imidoester, activated ester (e.g., N-hydroxysuccinimide ester, pentynoic acid STP ester), ketone, α,β-unsaturated carbonyl, alkene, maleimide, α-haloimide, epoxide, aziridine, tetrazine, tetrazole, phosphine, biotin or thiirane functional group with a complementary reactive group. An exemplary reaction is a reaction of an amine (e.g., primary amine) with an N-hydroxysuccinimide ester or isothiocyanate.

In yet other embodiments, the functional group comprises an alkene, ester, amide, thioester, disulfide, carbocyclic, heterocyclic or heteroaryl group. In further embodiments, the functional group comprises an alkene, ester, amide, thioester, thiourea, disulfide, carbocyclic, heterocyclic or heteroaryl group. In other embodiments, the functional group comprises an amide or thiourea. In some more specific embodiments, functional group is a triazolyl functional group, an amide, or thiourea functional group.

In a preferred embodiment, iEDDA click chemistry is used for immobilizing macromolecules (e.g., proteins, polypeptides, peptides) to a solid support since it is rapid and delivers high yields at low input concentrations. In another preferred embodiment, m-tetrazine rather than tetrazine is used in an iEDDA click chemistry reaction, as m-tetrazine has improved bond stability.

In a preferred embodiment, the substrate surface is functionalized with TCO, and the recording tag-labeled protein, polypeptide, peptide is immobilized to the TCO coated substrate surface via an attached m-tetrazine moiety (Figure 34).

Proteins, polypeptides, or peptides can be immobilized to a surface of a solid support by its C-terminus, N-terminus, or an internal amino acid, for example, via an amine, carboxyl, or sulfydryl group. Standard activated supports used in coupling to amine groups include CNBr-activated, NHS-activated, aldehyde-activated, azlactone-activated, and CDI-activated supports. Standard activated supports used in carboxyl coupling include carbodiimide-activated carboxyl moieties coupling to amine supports. Cysteine coupling can employ maleimide, idoacetyl, and pyridyl disulfide activated supports. An alternative mode of peptide carboxy terminal immobilization uses anhydrotrypsin, a catalytically inert derivative of trypsin that binds peptides containing lysine or arginine residues at their C-termini without cleaving them.

In certain embodiments, a protein, polypeptide, or peptide is immobilized to a solid support via covalent attachment of a solid surface bound linker to a lysine group of the protein, polypeptide, or peptide.

Recording tags can be attached to the protein, polypeptide, or peptides pre- or post-immobilization to the solid support. For example, proteins, polypeptides, or peptides can be first labeled with recording tags and then immobilized to a solid surface via a recording tag comprising at two functional moieties for coupling (see, Figure 28). One functional moiety of the recording tag couples to the protein, and the other functional moiety immobilizes the recording tag-labeled protein to a solid support.

Alternatively, proteins, polypeptides, or peptides are immobilized to a solid support prior to labeling of the proteins, polypeptides or peptides with recording tags. For example, proteins can first be derivitized with reactive groups such as click chemistry moieties. The activated protein molecules can then be attached to a suitable solid support and then labeled with recording tags using the complementary click chemistry moiety. As an example, proteins derivatized with alkyne and mTet moieties may be immobilized to beads derivatized with azide and TCO and attached to recording tags labeled with azide and TCO.

It is understood that the methods provided herein for attaching macromolecules (e.g., proteins, polypeptides, or peptides) to the solid support may also be used to attach recording tags to the solid support or attach recording tags to macromolecules (e.g., proteins polypeptides, or peptides).

In certain embodiments, the surface of a solid support is passivated (blocked) to minimize non-specific absorption to binding agents. A “passivated” surface refers to a surface that has been treated with outer layer of material to minimize non-specific binding of a binding agent. Methods of passivating surfaces include standard methods from the fluorescent single molecule analysis literature, including passivating surfaces with polymer like polyethylene glycol (PEG) (Pan et al., 2015, Phys. Biol. 12:045006), polysiloxane (e.g., Pluronic F-127), star polymers (e.g., star PEG) (Groll et al., 2010, Methods Enzymol. 472:1-18), hydrophobic dichlorodimethylsilane (DDS) + self-assembled Tween-20 (Hua et al., 2014, Nat. Methods 11:1233-1236), and diamond-like carbon (DLC), DLC + PEG (Stavis et al., 2011, Proc. Natl. Acad. Sci. USA 108:983-988). In addition to covalent surface modifications, a number of passivating agents can be employed as well including surfactants like Tween-20, polysiloxane in solution (Pluronic series), poly vinyl alcohol, (PVA), and proteins like BSA and casein. Alternatively, density of proteins, polypeptide, or peptides can be titrated on the surface or within the volume of a solid substrate by spiking a competitor or "dummy" reactive molecule when immobilizing the proteins, polypeptides or peptides to the solid substrate (see, Figure 36A).

In certain embodiments where multiple macromolecules are immobilized on the same solid support, the macromolecules can be spaced appropriately to reduce the occurrence of or prevent a cross-binding or inter-molecular event, e.g., where a binding agent binds to a first macromolecule and its coding tag information is transferred to a recording tag associated with a neighboring macromolecule rather than the recording tag associated with the first macromolecule. To control macromolecule (e.g., protein, polypeptide, or peptide spacing) spacing on the solid support, the density of functional coupling groups (e.g., TCO) may be titrated on the substrate surface (see, Figure 34). In some embodiments, multiple macromolecules are spaced apart on the surface or within the volume (e.g., porous supports) of a solid support at a distance of about 50 nm to about 500 nm, or about 50 nm to about 400 nm, or about 50 nm to about 300 nm, or about 50 nm to about 200 nm, or about 50 nm to about 100 nm. In some embodiments, multiple macromolecules are spaced apart on the surface of a solid support with an average distance of at least 50 nm, at least 60 nm, at least 70 nm, at least 80 nm, at least 90 nm, at least 100 nm, at least 150 nm, at least 200 nm, at least 250 nm, at least 300 nm, at least 350 nm, at least 400 nm, at least 450 nm, or at least 500 nm. In some embodiments, multiple macromolecules are spaced apart on the surface of a solid support with an average distance of at least 50 nm. In some embodiments, macromolecules are spaced apart on the surface or within the volume of a solid support such that, empirically, the relative frequency of inter- to intra-molecular events is $<1:10$; $<1:100$; $<1:1,000$; or $<1:10,000$. A suitable spacing frequency can be determined empirically using a functional assay (see, Example 23), and can be accomplished by dilution and/or by spiking a "dummy" spacer molecule that competes for attachments sites on the substrate surface.

For example, as shown in Figure 34, PEG-5000 (MW ~ 5000) is used to block the interstitial space between peptides on the substrate surface (e.g., bead surface). In addition, the peptide is coupled to a functional moiety that is also attached to a PEG-5000 molecule. In a preferred embodiment, this is accomplished by coupling a mixture of NHS-PEG-5000-TCO + NHS-PEG-5000-Methyl to amine-derivitized beads (see Figure 34). The stoichiometric ratio between the two PEGs (TCO vs. methyl) is titrated to generate an appropriate density of functional coupling moieties (TCO groups) on the substrate surface; the methyl-PEG is inert to coupling. The effective spacing between TCO groups can be calculated by measuring the density of TCO groups on the surface. In certain embodiments, the mean spacing between coupling moieties (e.g., TCO) on the solid surface is at least 50 nm, at least 100 nm, at least 250 nm, or at least 500 nm. After PEG5000-TCO/methyl derivitization of the beads, the excess NH_2 groups on the surface are quenched with a reactive anhydride (e.g. acetic or succinic anhydride).

VI. Recording Tags

At least one recording tag is associated or co-localized directly or indirectly with the macromolecule and joined to the solid support (*see, e.g.*, Figure 5). A recording tag may comprise DNA, RNA, PNA, γPNA, GNA, BNA, XNA, TNA, polynucleotide analogs, or a combination thereof. A recording tag may be single stranded, or partially or completely double stranded. A recording tag may have a blunt end or overhanging end. In certain embodiments, upon binding of a binding agent to a macromolecule, identifying information of the binding agent's coding tag is transferred to the recording tag to generate an extended recording tag. Further extensions to the extended recording tag can be made in subsequent binding cycles.

A recording tag can be joined to the solid support, directly or indirectly (e.g., via a linker), by any means known in the art, including covalent and non-covalent interactions, or any combination thereof. For example, the recording tag may be joined to the solid support by a ligation reaction. Alternatively, the solid support can include an agent or coating to facilitate joining, either direct or indirectly, of the recording tag, to the solid support. Strategies for immobilizing nucleic acid molecules to solid supports (*e.g.*, beads) have been described in U.S. Patent 5,900,481; Steinberg et al. (2004, Biopolymers 73:597-605); Lund et al., 1988 (Nucleic Acids Res. 16: 10861-

10880); and Steinberg et al. (2004, Biopolymers 73:597-605), each of which is incorporated herein by reference in its entirety.

In certain embodiments, the co-localization of a macromolecule (e.g., peptide) and associated recording tag is achieved by conjugating macromolecule and recording tag to a bifunctional linker attached directly to the solid support surface Steinberg et al. (2004, Biopolymers 73:597-605). In further embodiments, a trifunctional moiety is used to derivitize the solid support (*e.g.*, beads), and the resulting bifunctional moiety is coupled to both the macromolecule and recording tag.

Methods and reagents (e.g., click chemistry reagents and photoaffinity labelling reagents) such as those described for attachment of macromolecules and solid supports, may also be used for attachment of recording tags.

In a particular embodiment, a single recording tag is attached to a macromolecule (e.g., peptide), preferably via the attachment to a de-blocked N- or C-terminal amino acid. In another embodiment, multiple recording tags are attached to the macromolecule (e.g., protein, polypeptide, or peptide), preferably to the lysine residues or peptide backbone. In some embodiments, a macromolecule (e.g., protein or polypeptide) labeled with multiple recording tags is fragmented or digested into smaller peptides, with each peptide labeled on average with one recording tag.

In certain embodiments, a recording tag comprises an optional, unique molecular identifier (UMI), which provides a unique identifier tag for each macromolecule (*e.g.*, protein, polypeptide, peptide) to which the UMI is associated with. A UMI can be about 3 to about 40 bases, about 3 to about 30 bases, about 3 to about 20 bases, or about 3 to about 10 bases, or about 3 to about 8 bases. In some embodiments, a UMI is about 3 bases, 4 bases, 5 bases, 6 bases, 7 bases, 8 bases, 9 bases, 10 bases, 11 bases, 12 bases, 13 bases, 14 bases, 15 bases, 16 bases, 17 bases, 18 bases, 19 bases, 20 bases, 25 bases, 30 bases, 35 bases, or 40 bases in length. A UMI can be used to de-convolute sequencing data from a plurality of extended recording tags to identify sequence reads from individual macromolecules. In some embodiments, within a library of macromolecules, each macromolecule is associated with a single recording tag, with each recording tag comprising a unique UMI. In other embodiments, multiple copies of a recording tag are associated with a single macromolecule, with each copy of the recording tag comprising the same UMI. In some embodiments, a UMI has a different base sequence than the spacer or encoder sequences within the binding agents' coding tags to facilitate distinguishing these components during sequence analysis.

In certain embodiments, a recording tag comprises a barcode, e.g., other than the UMI if present. A barcode is a nucleic acid molecule of about 3 to about 30 bases, about 3 to about 25 bases, about 3 to about 20 bases, about 3 to about 10 bases, about 3 to about 10 bases, about 3 to about 8 bases in length. In some embodiments, a barcode is about 3 bases, 4 bases, 5 bases, 6 bases, 7 bases, 8 bases, 9 bases, 10 bases, 11 bases, 12 bases, 13 bases, 14 bases, 15 bases, 20 bases, 25 bases, or 30 bases in length. In one embodiment, a barcode allows for multiplex sequencing of a plurality of samples or libraries. A barcode may be used to identify a partition, a fraction, a compartment, a sample, a spatial location, or library from which the macromolecule (*e.g.*, peptide) derived. Barcodes can be used to de-convolute multiplexed sequence data and identify sequence reads from an individual sample or library. For example, a barcoded bead is useful for methods involving emulsions and partitioning of samples, *e.g.*, for purposes of partitioning the proteome.

A barcode can represent a compartment tag in which a compartment, such as a droplet, microwell, physical region on a solid support, etc. is assigned a unique barcode. The association of a compartment with a specific barcode can be achieved in any number of ways such as by encapsulating a single barcoded bead in a compartment, e.g., by direct merging or adding a barcoded droplet to a compartment, by directly printing or injecting a barcode reagents to a compartment, etc. The barcode reagents within a compartment are used to add compartment-specific barcodes to the macromolecule or fragments thereof within the compartment. Applied to protein partitioning into compartments, the barcodes can be used to map analysed peptides back to their originating protein molecules in the compartment. This can greatly facilitate protein identification. Compartment barcodes can also be used to identify protein complexes.

In other embodiments, multiple compartments that represent a subset of a population of compartments may be assigned a unique barcode representing the subset.

Alternatively, a barcode may be a sample identifying barcode. A sample barcode is useful in the multiplexed analysis of a set of samples in a single reaction vessel or immobilized to a single solid substrate or collection of solid substrates (e.g., a planar slide, population of beads contained in a single tube or vessel, etc.). Macromolecules from many different samples can be labeled with recording tags with sample-specific barcodes, and then all the samples pooled together prior to immobilization to a solid support, cyclic binding, and recording tag analysis. Alternatively, the samples can be kept separate until after creation of a DNA-encoded library, and sample barcodes attached during PCR amplification of the DNA-encoded library, and then mixed together prior to sequencing. This approach could be useful when assaying analytes (e.g., proteins) of different abundance classes. For example, the sample can be split and barcoded, and one portion processed using binding agents to low abundance analytes, and the other portion processed using binding agents to higher abundance analytes. In a particular embodiment, this approach helps to adjust the dynamic range of a particular protein analyte assay to lie within the "sweet spot" of standard expression levels of the protein analyte.

In certain embodiments, peptides, polypeptides, or proteins from multiple different samples are labeled with recording tags containing sample-specific barcodes. The multi-sample barcoded peptides, polypeptides, or proteins can be mixed together prior to a cyclic binding reaction. In this way, a highly-multiplexed alternative to a digital reverse phase protein array (RPPA) is effectively created (Guo, Liu et al. 2012, Assadi, Lamerz et al. 2013, Akbani, Becker et al. 2014, Creighton and Huang 2015). The creation of a digital RPPA-like assay has numerous applications in translational research, biomarker validation, drug discovery, clinical, and precision medicine.

In certain embodiments, a recording tag comprises a universal priming site, *e.g.*, a forward or 5' universal priming site. A universal priming site is a nucleic acid sequence that may be used for priming a library amplification reaction and/or for sequencing. A universal priming site may include, but is not limited to, a priming site for PCR amplification, flow cell adaptor sequences that anneal to complementary oligonucleotides on flow cell surfaces (*e.g.*, Illumina next generation sequencing), a sequencing priming site, or a combination thereof. A universal priming site can be about 10 bases to about 60 bases. In some embodiments, a universal priming site comprises an Illumina P5 primer (5'-AATGATACGGCGACCACCGA-3' – SEQ ID NO:133) or an Illumina P7 primer (5'-CAAGCAGAAGACGGCATACGAGAT – 3' - SEQ ID NO:134).

In certain embodiments, a recording tag comprises a spacer at its terminus, e.g., 3' end. As used herein reference to a spacer sequence in the context of a recording tag includes a spacer sequence that is identical to the spacer sequence associated with its cognate binding agent, or a spacer sequence that is complementary to the spacer sequence associated with its cognate binding agent. The terminal, e.g., 3', spacer on the recording tag permits transfer of identifying information of a cognate binding agent from its coding tag to the recording tag during the first binding cycle (*e.g.*, via annealing of complementary spacer sequences for primer extension or sticky end ligation).

In one embodiment, the spacer sequence is about 1-20 bases in length, about 2-12 bases in length, or 5-10 bases in length. The length of the spacer may depend on factors such as the temperature and reaction conditions of the primer extension reaction for transferring coding tag information to the recording tag.

In a preferred embodiment, the spacer sequence in the recording is designed to have minimal complementarity to other regions in the recording tag; likewise the spacer sequence in the coding tag should have minimal complementarity to other regions in the coding tag. In other words, the spacer sequence of the recording tags and coding tags should have minimal sequence complementarity to components such unique molecular identifiers, barcodes (e.g., compartment, partition, sample, spatial location), universal primer sequences, encoder sequences, cycle specific sequences, etc. present in the recording tags or coding tags.

As described for the binding agent spacers, in some embodiments, the recording tags associated with a library of macromolecules share a common spacer sequence. In other embodiments, the recording tags associated with a library of macromolecules have binding cycle specific spacer sequences that are complementary to the binding cycle specific spacer sequences of their cognate binding agents, which can be useful when using non-concatenated extended recording tags (see Figure 10).

The collection of extended recording tags can be concatenated after the fact (see, e.g., Figure 10). After the binding cycles are complete, the bead solid supports, each bead comprising on average one or fewer than one macromolecule per bead, each macromolecule having a collection of extended recording tags that are co-localized at the site of the macromolecule, are placed in an emulsion. The emulsion is formed such that each droplet, on average, is occupied by at most 1 bead. An optional assembly PCR reaction is performed in-emulsion to amplify the extended recording tags co-localized with the macromolecule on the bead and assemble them in co-linear order by priming between the different cycle specific sequences on the separate extended recording tags (Xiong, Peng et al. 2008). Afterwards the emulsion is broken and the assembled extended recording tags are sequenced.

In another embodiment, the DNA recording tag is comprised of a universal priming sequence (U1), one or more barcode sequences (BCs), and a spacer sequence (Sp1) specific to the first binding cycle. In the first binding cycle, binding agents employ DNA coding tags comprised of an Sp1 complementary spacer, an encoder barcode, and optional cycle barcode, and a second spacer element (Sp2). The utility of using at least two different spacer elements is that the first binding cycle selects one of potentially several DNA recording tags and a single DNA recording tag is extended resulting in a new Sp2 spacer element at the end of the extended DNA recording tag. In the second and subsequent binding cycles, binding agents contain just the Sp2' spacer rather than Sp1'. In this way, only the single extended recording tag from the first cycle is extended in subsequent cycles. In another embodiment, the second and subsequent cycles can employ binding agent specific spacers.

In some embodiments, a recording tag comprises from 5' to 3' direction: a universal forward (or 5') priming sequence, a UMI, and a spacer sequence. In some embodiments, a recording tag comprises from 5' to 3' direction: a universal forward (or 5') priming sequence, an optional UMI, a barcode (e.g., sample barcode, partition barcode, compartment barcode, spatial barcode, or any combination thereof), and a spacer sequence. In some other embodiments, a recording tag comprises from 5' to 3' direction: a universal forward (or 5') priming sequence, a barcode (e.g., sample barcode, partition barcode, compartment barcode, spatial barcode, or any combination thereof), an optional UMI, and a spacer sequence.

Combinatorial approaches may be used to generate UMIs from modified DNA and PNAs. In one example, a UMI may be constructed by "chemical ligating" together sets of short word sequences (4-15mers), which have been designed to be orthogonal to each other (Spiropulos and Heemstra 2012). A DNA template is used to direct the chemical ligation of the "word" polymers. The DNA template is constructed with hybridizing arms that enable assembly of a combinatorial template structure simply by mixing the sub-components together in solution (see, Figure 12C). In certain embodiments, there are no "spacer" sequences in this design. The size of the word space can vary from 10's of words to 10,000's or more words. In certain embodiments, the words are chosen such that they differ from one another to not cross hybridize, yet possess relatively uniform hybridization conditions. In one embodiment, the length of the word will be on the order of 10 bases, with about 1000's words in the subset (this is only 0.1% of the total 10-mer word space ~ 4^{10} = 1 million words). Sets of these words (1000 in subset) can be concatenated together to generate a final combinatorial UMI with complexity = 1000^{n} power. For 4 words concatenated together, this creates a UMI diversity of 10^{12} different elements. These UMI sequences will be appended to the macromolecule (peptides, proteins, etc.) at the single molecule level. In one embodiment, the diversity of UMIs exceeds the number of molecules of macromolecules to which the UMIs are attached. In this way, the UMI uniquely identifies the macromolecule of interest. The use of combinatorial word UMI's facilitates readout on high error rate sequencers, (e.g. nanopore sequencers, nanogap tunneling sequencing, etc.) since single base resolution is not required to read words of multiple bases in length. Combinatorial word approaches can also be used to generate other identity-informative components of recording tags or coding tags, such as compartment tags, partition barcodes, spatial barcodes, sample barcodes, encoder sequences, cycle specific sequences, and barcodes. Methods relating to nanopore sequencing and DNA encoding information with error-tolerant words (codes) are known in the art (see, e.g., Kiah et al., 2015, *Codes for DNA sequence profiles.* IEEE International Symposium on Information Theory (ISIT); Gabrys et al., 2015, *Asymmetric Lee distance codes for DNA-based storage*. IEEE Symposium on Information Theory (ISIT); Laure et al., 2016, *Coding in 2D: Using Intentional Dispersity to Enhance the Information Capacity of Sequence-Coded Polymer Barcodes*. Angew. Chem. Int. Ed. doi:10.1002/anie.201605279; Yazdi et al., 2015, IEEE Transactions on Molecular, Biological and Multi-Scale Communications 1:230-248; and Yazdi et al., 2015, Sci Rep 5:14138, each of which is incorporated by reference in its entirety). Thus, in certain embodiments, an extended recording tag, an extended coding tag, or a di-tag construct in any of the embodiments described herein is comprised of identifying components (e.g., UMI, encoder sequence, barcode, compartment tag, cycle specific sequence, etc.) that are error correcting codes. In some embodiments, the error correcting code is selected from: Hamming code, Lee distance code, asymmetric Lee distance code, Reed-Solomon code, and Levenshtein-Tenengolts code. For nanopore sequencing, the current or ionic flux profiles and asymmetric base calling errors are intrinsic to the type of nanopore and biochemistry employed, and this information can be used to design more robust DNA codes using the aforementioned error correcting approaches. An alternative to employing robust DNA nanopore sequencing barcodes, one can directly use the current or ionic flux signatures of barcode sequences (U.S. Patent No. 7,060,507, incorporated by reference in its entirety), avoiding DNA base calling entirely, and immediately identify the barcode sequence by mapping back to the predicted current/flux signature as described by Laszlo et al. (2014, Nat. Biotechnol. 32:829-833, incorporated by reference in its entirety). In this paper, Laszlo et al. describe the current signatures generated by the biological nanopore, MspA, when passing different word strings through the nanopore, and the ability to map and identify DNA strands by mapping resultant current signatures back to an *in silico* prediction of possible current signatures from a universe of sequences (2014, Nat. Biotechnol. 32:829-833). Similar concepts can be applied to DNA codes and the electrical signal generated by nanogap tunneling current-based DNA sequencing (Ohshiro et al., 2012, Sci Rep 2: 501).

Thus, in certain embodiments, the identifying components of a coding tag, recording tag, or both are capable of generating a unique current or ionic flux or optical signature, wherein the analysis step of any of the methods provided herein comprises detection of the unique current or ionic flux or optical signature in order to identify the identifying components. In some embodiments, the identifying components are selected from an encoder sequence, barcode, UMI, compartment tag, cycle specific sequence, or any combination thereof.

In certain embodiments, all or substantially amount of the macromolecules (e.g., proteins, polypeptides, or peptides) (e.g., at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) within a sample are labeled with a recording tag. Labeling of the macromolecules may occur before or after immobilization of the macromolecules to a solid support.

In other embodiments, a subset of macromolecules (e.g., proteins, polypeptides, or peptides) within a sample are labeled with recording tags. In a particular embodiment, a subset of macromolecules from a sample undergo targeted (analyte specific) labeling with recording tags. Targeted recording tag labeling of proteins may be achieved using target protein-specific binding agents (e.g., antibodies, aptamers, etc.) that are linked a short target-specific DNA capture probe, e.g., analyte-specific barcode, which anneal to complementary target-specific bait sequence, e.g., analyte-specific barcode, in recording tags (see, Figure 28A). The recording tags comprise a reactive moiety for a cognate reactive moiety present on the target protein (e.g., click chemistry labeling, photoaffinity labeling). For example, recording tags may comprise an azide moiety for interacting with alkyne-derivatized proteins, or recording tags may comprise a benzophenone for interacting with native proteins, etc.

(see Figures 28A-B). Upon binding of the target protein by the target protein specific binding agent, the recording tag and target protein are coupled via their corresponding reactive moieties (see, Figure 28B-C). After the target protein is labeled with the recording tag, the target-protein specific binding agent may be removed by digestion of the DNA capture probe linked to the target-protein specific binding agent. For example, the DNA capture probe may be designed to contain uracil bases, which are then targeted for digestion with a uracil-specific excision reagent (e.g., USER™), and the target-protein specific binding agent may be dissocated from the target protein.

In one example, antibodies specific for a set of target proteins can be labeled with a DNA capture probe (e.g., analyte barcode BC_A in Figure 28) that hybridizes with recording tags designed with complementary bait sequence (e.g., analyte barcode BC_A' in Figure 28). Sample-specific labeling of proteins can be achieved by employing DNA-capture probe labeled antibodies hybridizing with complementary bait sequence on recording tags comprising of sample-specific barcodes.

In another example, target protein-specific aptamers are used for targeted recording tag labeling of a subset of proteins within a sample. A target specific-aptamer is linked to a DNA capture probe that anneals with complementary bait sequence in a recording tag. The recording tag comprises a reactive chemical or photo-reactive chemical probes (e.g. benzophenone (BP)) for coupling to the target protein having a corresponding reactive moiety. The aptamer binds to its target protein molecule, bringing the recording tag into close proximity to the target protein, resulting in the coupling of the recording tag to the target protein.

Photoaffinity (PA) protein labeling using photo-reactive chemical probes attached to small molecule protein affinity ligands has been previously described (Park, Koh et al. 2016). Typical photo-reactive chemical probes include probes based on benzophenone (reactive diradical, 365 nm), phenyldiazirine (reactive carbon, 365 nm), and phenylazide (reactive nitrene free radical, 260 nm), activated under irradiation wavelengths as previously described (Smith and Collins 2015). In a preferred embodiment, target proteins within a protein sample are labeled with recording tags comprising sample barcodes using the method disclosed by Li et al.. in which a bait sequence in a benzophenone labeled recording tag is hybridized to a DNA capture probe attached to a cognate binding agent (e.g., nucleic acid aptamer (see Figure 28) (Li, Liu et al. 2013). For photoaffinity labeled protein targets, the use of DNA/RNA aptamers as target protein-specific binding agents are preferred over antibodies since the photoaffinity moiety can self-label the antibody rather than the target protein. In contrast, photoaffinity labeling is less efficient for nucleic acids than proteins, making aptamers a better vehicle for DNA-directed chemical or photo-labeling. Similar to photo-affinity labeling, one can also employ DNA-directed chemical labeling of reactive lysine's (or other moieties) in the proximity of the aptamer binding site in a manner similar to that described by Rosen et al. (Rosen, Kodal et al. 2014, Kodal, Rosen et al. 2016).

In the aforementioned embodiments, other types of linkages besides hybridization can be used to link the target specific binding agent and the recording tag (see, Figure 28A). For example, the two moieties can be covalently linked, using a linker that is designed to be cleaved and release the binding agent once the captured target protein (or other macromolecule) is covalently linked to the recording tag as shown in Figure 28B. A suitable linker can be attached to various positions of the recording tag, such as the 3' end, or within the linker attached to the 5' end of the recording tag.

VII. Binding Agents and Coding Tags

The methods described herein use a binding agent capable of binding to the macromolecule. A binding agent can be any molecule (*e.g.*, peptide, polypeptide, protein, nucleic acid, carbohydrate, small molecule, and the like) capable of binding to a component or feature of a macromolecule. A binding agent can be a naturally occurring, synthetically produced, or recombinantly expressed molecule. A binding agent may bind to a single monomer or subunit of a macromolecule (*e.g.*, a single amino acid of a peptide) or bind to multiple linked subunits of a macromolecule (*e.g.*, dipeptide, tripeptide, or higher order peptide of a longer peptide molecule).

In certain embodiments, a binding agent may be designed to bind covalently. Covalent binding can be designed to be conditional or favored upon binding to the correct moiety. For example, an NTAA and its cognate NTAA-specific binding agent may each be modified with a reactive group such that once the NTAA-specific binding agent is bound to the cognate NTAA, a coupling reaction is carried out to create a covalent linkage between the two. Non-specific binding of the binding agent to other locations that lack the cognate reactive group would not result in covalent attachment. Covalent binding between a binding agent and its target allows for more stringent washing to be used to remove binding agents that are non-specifically bound, thus increasing the specificity of the assay.

In certain embodiments, a binding agent may be a selective binding agent. As used herein, selective binding refers to the ability of the binding agent to preferentially bind to a specific ligand (*e.g.*, amino acid or class of amino acids) relative to binding to a different ligand (*e.g.*, amino acid or class of amino acids). Selectivity is commonly referred to as the equilibrium constant for the reaction of displacement of one ligand by another ligand in a complex with a binding agent. Typically, such selectivity is associated with the spatial geometry of the ligand and/or the manner and degree by which the ligand binds to a binding agent, such as by hydrogen bonding or Van der Waals forces (non-covalent interactions) or by reversible or non-reversible covalent attachment to the binding agent. It should also be understood that selectivity may be relative, and as opposed to absolute, and that different factors can affect the same, including ligand concentration. Thus, in one example, a binding agent selectively binds one of the twenty standard amino acids. In an example of non-selective binding, a binding agent may bind to two or more of the twenty standard amino acids.

In the practice of the methods disclosed herein, the ability of a binding agent to selectively bind a feature or component of a macromolecule need only be sufficient to allow transfer of its coding tag information to the recording tag associated with the macromolecule, transfer of the recording tag information to the coding tag, or transferring of the coding tag information and recording tag information to a di-tag molecule. Thus, selectively need only be relative to the other binding agents to which the macromolecule is exposed. It should also be understood that selectivity of a binding agent need not be absolute to a specific amino acid, but could be selective to a class of amino acids, such as amino acids with nonpolar or non-polar side chains, or with electrically (positively or negatively) charged side chains, or with aromatic side chains, or some specific class or size of side chains, and the like.

In a particular embodiment, the binding agent has a high affinity and high selectivity for the macromolecule of interest. In particular, a high binding affinity with a low off-rate is efficacious for information transfer between the coding tag and recording tag. In certain embodiments, a binding agent has a Kd of < 10 nM, <5 nM, < 1 nM, < 0.5 nM, or < 0.1 nM. In a particular embodiment, the binding agent is added to the macromolecule at a concentration >10X, >100X, or >1000X its Kd to drive binding to completion. A detailed discussion of binding kinetics of an antibody to a single protein molecule is described in Chang et al. (Chang, Rissin et al. 2012).

To increase the affinity of a binding agent to small N-terminal amino acids (NTAAs) of peptides, the NTAA may be modified with an "immunogenic" hapten, such as dinitrophenol (DNP). This can be implemented in a cyclic sequencing approach using Sanger's reagent, dinitrofluorobenzene (DNFB), which attaches a DNP group to the amine group of the NTAA. Commercial anti-DNP antibodies have affinities in the low nM range (~8 nM, LO-DNP-2) (Bilgicer, Thomas et al. 2009); as such it stands to reason that it should be possible to engineer high-affinity NTAA binding agents to a number of NTAAs modified with DNP (via DNFB) and simultaneously achieve good binding selectivity for a particular NTAA. In another example, an NTAA may be modified with sulfonyl nitrophenol (SNP) using 4-sulfonyl-2-nitrofluorobenzene (SNFB). Similar affinity enhancements may also be achieved with alternative NTAA modifiers, such as an acetyl group or an amidinyl (guanidinyl) group.

In certain embodiments, a binding agent may bind to an NTAA, a CTAA, an intervening amino acid, dipeptide (sequence of two amino acids), tripeptide (sequence of three amino acids), or higher order peptide of a peptide molecule. In some embodiments, each binding agent in a library of binding agents selectively binds to a particular amino acid, for example one of the twenty standard naturally occurring amino acids. The standard, naturally-occurring amino acids include Alanine (A or Ala), Cysteine (C or Cys), Aspartic Acid (D or Asp), Glutamic Acid (E or Glu), Phenylalanine (F or Phe), Glycine (G or Gly), Histidine (H or His), Isoleucine (I or Ile), Lysine (K or Lys), Leucine (L or Leu), Methionine (M or Met), Asparagine (N or Asn), Proline (P or Pro), Glutamine (Q or Gln), Arginine (R or Arg), Serine (S or Ser), Threonine (T or Thr), Valine (V or Val), Tryptophan (W or Trp), and Tyrosine (Y or Tyr).

In certain embodiments, a binding agent may bind to a post-translational modification of an amino acid. In some embodiments, a peptide comprises one or more post-translational modifications, which may be the same of different. The NTAA, CTAA, an intervening amino acid, or a combination thereof of a peptide may be post-translationally modified. Post-translational modifications to amino acids include acylation, acetylation, alkylation (including methylation), biotinylation, butyrylation, carbamylation, carbonylation, deamidation, deiminiation, diphthamide formation, disulfide bridge formation, eliminylation, flavin attachment, formylation, gamma-carboxylation, glutamylation, glycylation, glycosylation, glypiation, heme C attachment, hydroxylation, hypusine formation, iodination, isoprenylation, lipidation, lipoylation, malonylation, methylation, myristolylation, oxidation, palmitoylation, pegylation, phosphopantetheinylation, phosphorylation, prenylation, propionylation, retinylidene Schiff base formation, S-glutathionylation, S-nitrosylation, S-sulfenylation, selenation, succinylation, sulfination, ubiquitination, and C-terminal amidation (*see, also,* Seo and Lee, 2004, J. Biochem. Mol. Biol. 37:35-44).

In certain embodiments, a lectin is used as a binding agent for detecting the glycosylation state of a protein, polypeptide, or peptide. Lectins are carbohydrate-binding proteins that can selectively recognize glycan epitopes of free carbohydrates or glycoproteins. A list of lectins recognizing various glycosylation states (e.g., core-fucose, sialic acids, N-acetyl-D-lactosamine, mannose, N-acetyl-glucosamine) include: A, AAA, AAL, ABA, ACA, ACG, ACL, AOL, ASA, BanLec, BC2L-A, BC2LCN, BPA, BPL, Calsepa, CGL2, CNL, Con, ConA, DBA, Discoidin, DSA, ECA, EEL, F17AG, Gal1, Gal1-S, Gal2, Gal3, Gal3C-S, Gal7-S, Gal9, GNA, GRFT, GS-I, GS-II, GSL-I, GSL-II, HHL, HIHA, HPA, I, II, Jacalin, LBA, LCA, LEA, LEL, Lentil, Lotus, LSL-N, LTL, MAA, MAH, MAL_I, Malectin, MOA, MPA, MPL, NPA, Orysata, PA-IIL, PA-IL, PALa, PHA-E, PHA-L, PHA-P, PHAE, PHAL, PNA, PPL, PSA, PSL1a, PTL, PTL-I, PWM, RCA120, RS-Fuc, SAMB, SBA, SJA, SNA, SNA-I, SNA-II, SSA, STL, TJA-I, TJA-II, TxLCI, UDA, UEA-I, UEA-II, VFA, VVA, WFA, WGA (see, Zhang et al., 2016, MABS 8:524-535).

In certain embodiments, a binding agent may bind to a modified or labeled NTAA. A modified or labeled NTAA can be one that is labeled with PITC, 1-fluoro-2,4-dinitrobenzene (Sanger's reagent, DNFB), dansyl chloride (DNS-Cl, or 1-dimethylaminonaphthalene-5-sulfonyl chloride), 4-sulfonyl-2-nitrofluorobenzene (SNFB), an acetylating reagent, a guanidination reagent, a thioacylation reagent, a thioacetylation reagent, or a thiobenzylation reagent.

In certain embodiments, a binding agent can be an aptamer (*e.g.*, peptide aptamer, DNA aptamer, or RNA aptamer), an antibody, an anticalin, an ATP-dependent Clp protease adaptor protein (ClpS), an antibody binding fragment, an antibody mimetic, a peptide, a peptidomimetic, a protein, or a polynucleotide (*e.g.*, DNA, RNA, peptide nucleic acid (PNA), a γPNA, bridged nucleic acid (BNA), xeno nucleic acid (XNA), glycerol nucleic acid (GNA), or threose nucleic acid (TNA), or a variant thereof).

As used herein, the terms antibody and antibodies are used in a broad sense, to include not only intact antibody molecules, for example but not limited to immunoglobulin A, immunoglobulin G, immunoglobulin D, immunoglobulin E, and immunoglobulin M, but also any immunoreactivity component(s) of an antibody molecule that immuno-specifically bind to at least one epitope. An antibody may be naturally occurring, synthetically produced, or recombinantly expressed. An antibody may be a fusion protein. An antibody may be an antibody mimetic. Examples of antibodies include but are not limited to, Fab fragments, Fab' fragments, $F(ab')_2$ fragments, single chain antibody fragments (scFv), miniantibodies, diabodies, crosslinked antibody fragments, Affibody™, nanobodies, single domain antibodies, DVD-Ig molecules, alphabodies, affimers, affitins, cyclotides, molecules, and the like. Immunoreactive products derived using antibody engineering or protein engineering techniques are also expressly within the meaning of the term antibodies. Detailed descriptions of antibody and/or protein engineering, including relevant protocols, can be found in, among other places, J. Maynard and G. Georgiou, 2000, Ann. Rev. Biomed. Eng. 2:339-76; Antibody Engineering, R. Kontermann and S. Dubel, eds., Springer Lab Manual, Springer Verlag (2001); U.S. Patent No. 5,831,012; and S. Paul, Antibody Engineering Protocols, Humana Press (1995).

As with antibodies, nucleic acid and peptide aptamers that specifically recognize a peptide can be produced using known methods. Aptamers bind target molecules in a highly specific, conformation-dependent manner, typically with very high affinity, although aptamers with lower binding affinity can be selected if desired. Aptamers have been shown to distinguish between targets based on very small structural differences such as the presence or absence of a methyl or hydroxyl group and certain aptamers can distinguish between D- and L-enantiomers. Aptamers have been obtained that bind small molecular targets, including drugs, metal ions, and organic dyes, peptides, biotin, and proteins, including but not limited to streptavidin, VEGF, and viral proteins. Aptamers have been shown to retain functional activity after biotinylation, fluorescein labeling, and when attached to glass surfaces and microspheres. (*see*, Jayasena, 1999, Clin Chem 45:1628-50; Kusser2000, J. Biotechnol. 74: 27-39; Colas, 2000, Curr Opin Chem Biol 4:54-9). Aptamers which specifically bind arginine and AMP have been described as well (see, Patel and Suri, 2000, J. Biotech. 74:39-60). Oligonucleotide aptamers that bind to a specific amino acid have been disclosed in Gold et al. (1995, Ann. Rev. Biochem. 64:763-97). RNA aptamers that bind amino acids have also been described (Ames and Breaker, 2011, RNA Biol. 8; 82-89; Mannironi et al., 2000, RNA 6:520-27; Famulok, 1994, J. Am. Chem. Soc. 116:1698-1706).

A binding agent can be made by modifying naturally-occurring or synthetically-produced proteins by genetic engineering to introduce one or more mutations in the amino acid sequence to produce engineered proteins that bind to a specific component or feature of a macromolecule (*e.g.*, NTAA, CTAA, or post-translationally modified amino acid or a peptide). For example, exopeptidases (e.g., aminopeptidases, carboxypeptidases), exoproteases, mutated exoproteases, mutated anticalins, mutated ClpSs, antibodies, or tRNA synthetases can be modified to create a binding agent that selectively binds to a particular NTAA. In another example, carboxypeptidases can be modified to create a binding agent that selectively binds to a particular CTAA. A binding agent can also be designed or modified, and utilized, to specifically bind a modified NTAA or modified CTAA, for example one that has a post-translational modification (*e.g.*, phosphorylated NTAA or phosphorylated CTAA) or one that has been modified with a label (*e.g.*, PTC, 1-fluoro-2,4-dinitrobenzene (using Sanger's reagent, DNFB), dansyl chloride (using DNS-Cl, or 1-dimethylaminonaphthalene-5-sulfonyl chloride), or using a thioacylation reagent, a thioacetylation reagent, an acetylation reagent, an amidination (guanidination) reagent, or a thiobenzylation reagent). Strategies for directed evolution of proteins are known in the art (e.g., reviewed by Yuan et al., 2005, Microbiol. Mol. Biol. Rev. 69:373-392), and include phage display, ribosomal display, mRNA display, CIS display, CAD display, emulsions, cell surface display method, yeast surface display, bacterial surface display, etc.

In some embodiments, a binding agent that selectively binds to a modified NTAA can be utilized. For example, the NTAA may be reacted with phenylisothiocyanate (PITC) to form a phenylthiocarbamoyl-NTAA derivative. In this manner, the binding agent may be fashioned to selectively bind both the phenyl group of the phenylthiocarbamoyl moiety as well as the alpha-carbon R group of the NTAA. Use of PITC in this manner allows for subsequent cleavage of the NTAA by Edman degradation as discussed below. In another embodiment, the NTAA may be reacted with Sanger's reagent (DNFB), to generate a DNP-labeled NTAA (see Figure 3). Optionally, DNFB is used with an ionic liquid such as 1-ethyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]imide ([emim][Tf2N]), in which DNFB is highly soluble. In this manner, the binding agent may be engineered to selectively bind the combination of the DNP and the R group on the NTAA. The addition of the DNP moiety provides a larger "handle" for the interaction of the binding agent with the NTAA, and should lead to a higher affinity interaction. In yet another embodiment, a binding agent may be an aminopeptidase that has been engineered to recognize the DNP-labeled NTAA providing cyclic control of aminopeptidase degradation of the peptide. Once the DNP-labeled NTAA is cleaved, another cycle of DNFB derivitization is performed in order to bind and cleave the newly exposed NTAA. In preferred particular embodiment, the aminopeptidase is a monomeric metallo-protease, such an aminopeptidase activated by zinc (Calcagno and Klein 2016). In another example, a binding agent may selectively bind to an NTAA that is modified with sulfonyl nitrophenol (SNP), e.g., by using 4-sulfonyl-2-nitrofluorobenzene (SNFB). In yet antoehr embodiment, a binding agent may selectively bind to an NTAA that is acetylated or amidinated.

Other reagents that may be used to modify the NTAA include trifluoroethyl isothiocyanate, allyl isothiocyanate, and dimethylaminoazobenzene isothiocyanate.

A binding agent may be engineered for high affinity for a modified NTAA, high specificity for a modified NTAA, or both. In some embodiments, binding agents can be developed through directed evolution of promising affinity scaffolds using phage display.

Engineered aminopeptidase mutants that bind to and cleave individual or small groups of labelled (biotinylated) NTAAs have been described (*see*, PCT Publication No. WO2010/065322, incorporated by reference in its entirety). Aminopeptidases are enzymes that cleave amino acids from the N-terminus of proteins or peptides. Natural aminopeptidases have very limited specificity, and generically cleave N-terminal amino acids in a processive manner, cleaving one amino acid off after another (Kishor et al., 2015, Anal. Biochem. 488:6-8). However, residue specific aminopeptidases have been identified (Eriquez et al., J. Clin. Microbiol. 1980, 12:667-71; Wilce et al., 1998, Proc. Natl. Acad. Sci. USA 95:3472-3477; Liao et al., 2004, Prot. Sci. 13:1802-10). Aminopeptidases may be engineered to specifically bind to 20 different NTAAs representing the standard amino acids that are labeled with a specific moiety (*e.g.*, PTC, DNP, SNP, etc.). Control of the stepwise degradation of the N-terminus of the peptide is achieved by using engineered aminopeptidases that are only active (*e.g.*, binding activity or catalytic activity) in the presence of the label. In another example, Havranak et al. (U.S. Patent Publication 2014/0273004) describes engineering aminoacyl tRNA synthetases (aaRSs) as specific NTAA binders. The amino acid binding pocket of the aaRSs has an intrinsic ability to bind cognate amino acids, but generally exhibits poor binding affinity and specificity. Moreover, these natural amino acid binders don't recognize N-terminal labels. Directed evolution of aaRS scaffolds can be used to generate higher affinity, higher specificity binding agents that recognized the N-terminal amino acids in the context of an N-terminal label.

In another example, highly-selective engineered ClpSs have also been described in the literature. Emili et al. describe the directed evolution of an *E. coli* ClpS protein via phage display, resulting in four different variants with the ability to selectively bind NTAAs for aspartic acid, arginine, tryptophan, and leucine residues (U.S. Patent 9,566,335, incorporated by reference in its entirety).

In a particular embodiment, anticalins are engineered for both high affinity and high specificity to labeled NTAAs (e.g. DNP, SNP, acetylated, etc.). Certain varieties of anticalin scaffolds have suitable shape for binding single amino acids, by virtue of their beta barrel structure. An N-terminal amino acid (either with or without modification) can potentially fit and be recognized in this "beta barrel" bucket. High affinity anticalins with engineered novel binding activities have been described (reviewed by Skerra, 2008, FEBS J. 275: 2677-2683). For example, anticalins with high affinity binding (low nM) to fluorescein and digoxygenin have been engineered (Gebauer and Skerra 2012). Engineering of alternative scaffolds for new binding functions has also been reviewed by Banta et al. (2013, Annu. Rev. Biomed. Eng. 15:93-113).

The functional affinity (avidity) of a given monovalent binding agent may be increased by at least an order of magnitude by using a bivalent or higher order multimer of the monovalent binding agent (Vauquelin and Charlton 2013). Avidity refers to the accumulated strength of multiple, simultaneous, non-covalent binding interactions. An individual binding interaction may be easily dissociated. However, when multiple binding interactions are present at the same time, transient dissociation of a single binding interaction does not allow the binding protein to diffuse away and the binding interaction is likely to be restored. An alternative method for increasing avidity of a binding agent is to include complementary sequences in the coding tag attached to the binding agent and the recording tag associated with the macromolecule.

In some embodiments, a binding agent can be utilized that selectively binds a modified C-terminal amino acid (CTAA). Carboxypeptidases are proteases that cleave terminal amino acids containing a free carboxyl group. A number of carboxypeptidases exhibit amino acid preferences, *e.g.*, carboxypeptidase B preferentially cleaves at basic amino acids, such as arginine and lysine. A carboxypeptidase can be modified to create a binding agent that selectively binds to particular amino acid. In some embodiments, the carboxypeptidase may be engineered to selectively bind both the modification moiety as well as the alpha-carbon R group of the CTAA. Thus, engineered carboxypeptidases may specifically recognize 20 different CTAAs representing the standard amino acids in the context of a C-terminal label. Control of the stepwise degradation from the C-terminus of the peptide is achieved by using engineered carboxypeptidases that are only active (*e.g.*, binding activity or catalytic activity) in the presence of the label. In one example, the CTAA may be modified by a para-Nitroanilide or 7-amino-4-methylcoumarinyl group.

Other potential scaffolds that can be engineered to generate binders for use in the methods described herein include: an anticalin, an amino acid tRNA synthetase (aaRS), ClpS, an Affilin®, an Adnectin™, a T cell receptor, a zinc finger protein, a thioredoxin, GST A1-1, DARPin, an affimer, an affitin, an alphabody, an avimer, a Kunitz domain peptide, a monobody, a single domain antibody, EETI-II, HPSTI, intrabody, lipocalin, PHD-finger, V(NAR) LDTI, evibody, Ig(NAR), knottin, maxibody, neocarzinostatin, pVIII, tendamistat, VLR, protein A scaffold, MTI-II, ecotin, GCN4, Im9, kunitz domain, microbody, PBP, trans-body, tetranectin, WW domain, CBM4-2, DX-88, GFP, iMab, Ldl receptor domain A, Min-23, PDZ-domain, avian pancreatic polypeptide, charybdotoxin/10Fn3, domain antibody (Dab), a2p8 ankyrin repeat, insect defensing A peptide, Designed AR protein, C-type lectin domain, staphylococcal nuclease, Src homology domain 3 (SH3), or Src homology domain 2 (SH2).

A binding agent may be engineered to withstand higher temperatures and mild-denaturing conditions (e.g., presence of urea, guanidinium thiocyanate, ionic solutions, etc.). The use of denaturants helps reduce secondary structures in the surface bound peptides, such as α-helical structures, β-hairpins, β -strands, and other such structures, which may interfere with binding of binding agents to linear peptide epitopes. In one embodiment, an ionic liquid such as 1-ethyl-3-methylimidazolium acetate ([EMIM]+[ACE] is used to reduce peptide secondary structure during binding cycles (Lesch, Heuer et al. 2015).

Any binding agent described also comprises a coding tag containing identifying information regarding the binding agent. A coding tag is a nucleic acid molecule of about 3 bases to about 100 bases that provides unique identifying information for its associated binding agent. A coding tag may comprise about 3 to about 90 bases, about 3 to about 80 bases, about 3 to about 70 bases, about 3 to about 60 bases, about 3 bases to about 50 bases, about 3 bases to about 40 bases, about 3 bases to about 30 bases, about 3 bases to about 20 bases, about 3 bases to about 10 bases, or about 3 bases to about 8 bases. In some embodiments, a coding tag is about 3 bases, 4 bases, 5 bases, 6 bases, 7 bases, 8 bases, 9 bases, 10 bases, 11 bases, 12 bases, 13 bases, 14 bases, 15 bases, 16 bases, 17 bases, 18 bases, 19 bases, 20 bases, 25 bases, 30 bases, 35 bases, 40 bases, 55 bases, 60 bases, 65 bases, 70 bases, 75 bases, 80 bases, 85 bases, 90 bases, 95 bases, or 100 bases in length. A coding tag may be composed of DNA, RNA, polynucleotide analogs, or a combination thereof. Polynucleotide analogs include PNA, γPNA, BNA, GNA, TNA, LNA, morpholino polynucleotides, 2'-O-Methyl polynucleotides, alkyl ribosyl substituted polynucleotides, phosphorothioate polynucleotides, and 7-deaza purine analogs.

A coding tag comprises an encoder sequence that provides identifying information regarding the associated binding agent. An encoder sequence is about 3 bases to about 30 bases, about 3 bases to about 20 bases, about 3 bases to about 10 bases, or about 3 bases to about 8 bases. In some embodiments, an encoder sequence is about 3 bases, 4 bases, 5 bases, 6 bases, 7 bases, 8 bases, 9 bases, 10 bases, 11 bases, 12 bases, 13 bases, 14 bases, 15 bases, 20 bases, 25 bases, or 30 bases in length. The length of the encoder sequence determines the number of unique encoder sequences that can be generated. Shorter encoding sequences generate a smaller number of unique encoding sequences, which may be useful when using a small number of binding agents. Longer encoder sequences may be desirable when analyzing a population of macromolecules. For example, an encoder sequence of 5 bases would have a formula of 5'-NNNNN-3' (SEQ ID NO:135), wherein N may be any naturally occurring nucleotide, or analog. Using the four naturally occurring nucleotides A, T, C, and G, the total number of unique encoder sequences having a length of 5 bases is 1,024. In some embodiments, the total number of unique encoder sequences may be reduced by excluding, for example, encoder sequences in which all the bases are identical, at least three contiguous bases are identical, or both. In a specific embodiment, a set of $\geq$ 50 unique encoder sequences are used for a binding agent library.

In some embodiments, identifying components of a coding tag or recording tag, e.g., the encoder sequence, barcode, UMI, compartment tag, partition barcode, sample barcode, spatial region barcode, cycle specific sequence or any combination thereof, is subject to Hamming distance, Lee distance, asymmetric Lee distance, Reed- Solomon, Levenshtein-Tenengolts, or similar methods for error-correction. Hamming distance refers to the number of positions that are different between two strings of equal length. It measures the minimum number of substitutions required to change one string into the other. Hamming distance may be used to correct errors by selecting encoder sequences that are reasonable distance apart. Thus, in the example where the encoder sequence is 5 base, the number of useable encoder sequences is reduced to 256 unique encoder sequences (Hamming distance of 1 $\rightarrow 4^4$ encoder sequences = 256 encoder sequences). In another embodiment, the encoder sequence, barcode, UMI, compartment tag, cycle specific sequence, or any combination thereof is designed to be easily read out by a cyclic decoding process (Gunderson, 2004, Genome Res. 14:870-7). In another embodiment, the encoder sequence, barcode, UMI, compartment tag, partition barcode, spatial barcode, sample barcode, cycle specific sequence, or any combination thereof is designed to be read out by low accuracy nanopore sequencing, since rather than requiring single base resolution, words of multiple bases (~5-20 bases in length) need to be read. A subset of 15-mer, error-correcting Hamming barcodes that may be used in the methods of the present disclosure are set forth in SEQ ID NOS:1-65 and their corresponding reverse complementary sequences as set forth in SEQ ID NO:66-130.

In some embodiments, each unique binding agent within a library of binding agents has a unique encoder sequence. For example, 20 unique encoder sequences may be used for a library of 20 binding agents that bind to the 20 standard amino acids. Additional coding tag sequences may be used to identify modified amino acids (*e.g*, post-translationally modified amino acids). In another example, 30 unique encoder sequences may be used for a library of 30 binding agents that bind to the 20 standard amino acids and 10 post-translational modified amino acids (*e.g.*, phosphorylated amino acids, acetylated amino acids, methylated amino acids). In other embodiments, two or more different binding agents may share the same encoder sequence. For example, two binding agents that each bind to a different standard amino acid may share the same encoder sequence.

In certain embodiments, a coding tag further comprises a spacer sequence at one end or both ends. A spacer sequence is about 1 base to about 20 bases, about 1 base to about 10 bases, about 5 bases to about 9 bases, or about 4 bases to about 8 bases. In some embodiments, a spacer is about 1 base, 2 bases, 3 bases, 4 bases, 5 bases, 6 bases, 7 bases, 8 bases, 9 bases, 10 bases, 11 bases, 12 bases, 13 bases, 14 bases, 15 bases or 20 bases in length. In some embodiments, a spacer within a coding tag is shorter than the encoder sequence, *e.g.*, at least 1 base, 2, bases, 3 bases, 4 bases, 5 bases, 6, bases, 7 bases, 8 bases, 9 bases, 10 bases, 11 bases, 12 bases, 13 bases, 14 bases, 15 bases, 20 bases, or 25 bases shorter than the encoder sequence. In other embodiments, a spacer within a coding tag is the same length as the encoder sequence. In certain embodiments, the spacer is binding agent specific so that a spacer from a previous binding cycle only interacts with a spacer from the appropriate binding agent in a current binding cycle. An example would be pairs of cognate antibodies containing spacer sequences that only allow information transfer if both antibodies sequentially bind to the macromolecule. A spacer sequence may be used as the primer annealing site for a primer extension reaction, or a splint or sticky end in a ligation reaction. A 5' spacer on a coding tag (see Figure 5A, "*Sp'") may optionally contain pseudo complementary bases to a 3' spacer on the recording tag to increase T_m (Lehoud et al., 2008, Nucleic Acids Res. 36:3409-3419).

In some embodiments, the coding tags within a collection of binding agents share a common spacer sequence used in an assay (*e.g.* the entire library of binding agents used in a multiple binding cycle method possess a common spacer in their coding tags). In another embodiment, the coding tags are comprised of a binding cycle tags, identifying a particular binding cycle. In other embodiments, the coding tags within a library of binding agents have a binding cycle specific spacer sequence. In some embodiments, a coding tag comprises one binding cycle specific spacer sequence. For example, a coding tag for binding agents used in the first binding cycle comprise a "cycle 1" specific spacer sequence, a coding tag for binding agents used in the second binding cycle comprise a "cycle 2" specific spacer sequence, and so on up to "n" binding cycles. In further embodiments, coding tags for binding agents used in the first binding cycle comprise a "cycle 1" specific spacer sequence and a "cycle 2" specific spacer sequence, coding tags for binding agents used in the second binding cycle comprise a "cycle 2" specific spacer sequence and a "cycle 3" specific spacer sequence, and so on up to "n" binding cycles. This embodiment is useful for subsequent PCR assembly of non-concatenated extended recording tags after the binding cycles are completed (see Figure 10). In some embodiments, a spacer sequence comprises a sufficient number of bases to anneal to a complementary spacer sequence in a recording tag or extended recording tag to initiate a primer extension reaction or sticky end ligation reaction.

A cycle specific spacer sequence can also be used to concatenate information of coding tags onto a single recording tag when a population of recording tags is associated with a macromolecule. The first binding cycle transfers information from the coding tag to a randomly-chosen recording tag, and subsequent binding cycles can prime only the extended recording tag using cycle dependent spacer sequences. More specifically, coding tags for binding agents used in the first binding cycle comprise a “cycle 1” specific spacer sequence and a “cycle 2” specific spacer sequence, coding tags for binding agents used in the second binding cycle comprise a “cycle 2” specific spacer sequence and a “cycle 3” specific spacer sequence, and so on up to “n” binding cycles. Coding tags of binding agents from the first binding cycle are capable of annealing to recording tags via complementary cycle 1 specific spacer sequences. Upon transfer of the coding tag information to the recording tag, the cycle 2 specific spacer sequence is positioned at the 3’ terminus of the extended recording tag at the end of binding cycle 1. Coding tags of binding agents from the second binding cycle are capable of annealing to the extended recording tags via complementary cycle 2 specific spacer sequences. Upon transfer of the coding tag information to the extended recording tag, the cycle 3 specific spacer sequence is positioned at the 3’ terminus of the extended recording tag at the end of binding cycle 2, and so on through “n” binding cycles. This embodiment provides that transfer of binding information in a particular binding cycle among multiple binding cycles will only occur on (extended) recording tags that have experienced the previous binding cycles. However, sometimes a binding agent will fail to bind to a cognate macromolecule. Oligonucleotides comprising binding cycle specific spacers after each binding cycle as a “chase” step can be used to keep the binding cycles synchronized even if the event of a binding cycle failure. For example, if a cognate binding agent fails to bind to a macromolecule during binding cycle 1, adding a chase step following binding cycle 1 using oligonucleotides comprising both a cycle 1 specific spacer, a cycle 2 specific spacer, and a “null” encoder sequence. The “null” encoder sequence can be the absence of an encoder sequence or, preferably, a specific barcode that positively identifies a “null” binding cycle. The “null” oligonucleotide is capable of annealing to the recording tag via the cycle 1 specific spacer, and the cycle 2 specific spacer is transferred to the recording tag. Thus, binding agents from binding cycle 2 are capable of annealing to the extended recording tag via the cycle 2 specific spacer despite the failed binding cycle 1 event.

The "null" oligonucleotide marks binding cycle 1 as a failed binding event within the extended recording tag.

In preferred embodiment, binding cycle-specific encoder sequences are used in coding tags. Binding cycle-specific encoder sequences may be accomplished either via the use of completely unique analyte (e.g., NTAA)-binding cycle encoder barcodes or through a combinatoric use of an analyte (e.g., NTAA) encoder sequence joined to a cycle-specific barcode (see Figure 35). The advantage of using a combinatoric approach is that fewer total barcodes need to be designed. For a set of 20 analyte binding agents used across 10 cycles, only 20 analyte encoder sequence barcodes and 10 binding cycle specific barcodes need to be designed. In contrast, if the binding cycle is embedded directly in the binding agent encoder sequence, then a total of 200 independent encoder barcodes may need to be designed. An advantage of embedding binding cycle information directly in the encoder sequence is that the total length of the coding tag can be minimized when employing error-correcting barcodes on a nanopore readout. The use of error-tolerant barcodes allows highly accurate barcode identification using sequencing platforms and approaches that are more error-prone, but have other advantages such as rapid speed of analysis, lower cost, and/or more portable instrumentation. One such example is a nanopore-based sequencing readout.

In some embodiments, a coding tag comprises a cleavable or nickable DNA strand within the second (3') spacer sequence proximal to the binding agent (see, Figure 32). For example, the 3' spacer may have one or more uracil bases that can be nicked by uracil-specific excision reagent (USER). USER generates a single nucleotide gap at the location of the uracil. In another example, the 3' spacer may comprise a recognition sequence for a nicking endonuclease that hydrolyzes only one strand of a duplex. Preferably, the enzyme used for cleaving or nicking the 3' spacer sequence acts only on one DNA strand (the 3' spacer of the coding tag), such that the other strand within the duplex belonging to the (extended) recording tag is left intact. These embodiments is particularly useful in assays analysing proteins in their native conformation, as it allows the non-denaturing removal of the binding agent from the (extended) recording tag after primer extension has occurred and leaves a single stranded DNA spacer sequence on the extended recording tag available for subsequent binding cycles.

The coding tags may also be designed to contain palindromic sequences. Inclusion of a palindromic sequence into a coding tag allows a nascent, growing, extended recording tag to fold upon itself as coding tag information is transferred. The extended recording tag is folded into a more compact structure, effectively decreasing undesired inter-molecular binding and primer extension events.

In some embodiments, a coding tag comprises analyte-specific spacer that is capable of priming extension only on recording tags previously extended with binding agents recognizing the same analyte. An extended recording tag can be built up from a series of binding events using coding tags comprising analyte-specific spacers and encoder sequences. In one embodiment, a first binding event employs a binding agent with a coding tag comprised of a generic 3' spacer primer sequence and an analyte-specific spacer sequence at the 5' terminus for use in the next binding cycle; subsequent binding cycles then use binding agents with encoded analyte-specific 3' spacer sequences. This design results in amplifiable library elements being created only from a correct series of cognate binding events. Off-target and cross-reactive binding interactions will lead to a non-amplifiable extended recording tag. In one example, a pair of cognate binding agents to a particular macromolecule analyte is used in two binding cycles to identify the analyte. The first cognate binding agent contains a coding tag comprised of a generic spacer 3' sequence for priming extension on the generic spacer sequence of the recording tag, and an encoded analyte-specific spacer at the 5' end, which will be used in the next binding cycle. For matched cognate binding agent pairs, the 3' analyte-specific spacer of the second binding agent is matched to the 5' analyte-specific spacer of the first binding agent. In this way, only correct binding of the cognate pair of binding agents will result in an amplifiable extended recording tag. Cross-reactive binding agents will not be able to prime extension on the recording tag, and no amplifiable extended recording tag product generated. This approach greatly enhances the specificity of the methods disclosed herein. The same principle can be applied to triplet binding agent sets, in which 3 cycles of binding are employed. In a first binding cycle, a generic 3' Sp sequence on the recording tag interacts with a generic spacer on a binding agent coding tag. Primer extension transfers coding tag information, including an analyte specific 5' spacer, to the recording tag. Subsequent binding cycles employ analyte specific spacers on the binding agents' coding tags.

In certain embodiments, a coding tag may further comprise a unique molecular identifier for the binding agent to which the coding tag is linked. A UMI for the binding agent may be useful in embodiments utilizing extended coding tags or di-tag molecules for sequencing readouts, which in combination with the encoder sequence provides information regarding the identity of the binding agent and number of unique binding events for a macromolecule.

In another embodiment, a coding tag includes a randomized sequence (a set of N's, where N= a random selection from A, C, G, T, or a random selection from a set of words). After a series of "n" binding cycles and transfer of coding tag information to the (extended) recording tag, the final extended recording tag product will be composed of a series of these randomized sequences, which collectively form a "composite" unique molecule identifier (UMI) for the final extended recording tag. If for instance each coding tag contains an (NN) sequence (4*4=16 possible sequences), after 10 sequencing cycles, a combinatoric set of 10 distributed 2-mers is formed creating a total diversity of $16^{10} \sim 10^{12}$ possible composite UMI sequences for the extended recording tag products. Given that a peptide sequencing experiment uses $\sim 10^9$ molecules, this diversity is more than sufficient to create an effective set of UMIs for a sequencing experiment. Increased diversity can be achieved by simply using a longer randomized region (NNN, NNNN, etc.) within the coding tag.

A coding tag may include a terminator nucleotide incorporated at the 3' end of the 3' spacer sequence. After a binding agent binds to a macromolecule and their corresponding coding tag and recording tags anneal via complementary spacer sequences, it is possible for primer extension to transfer information from the coding tag to the recording tag, or to transfer information from the recording tag to the coding tag. Addition of a terminator nucleotide on the 3' end of the coding tag prevents transfer of recording tag information to the coding tag. It is understood that for embodiments described herein involving generation of extended coding tags, it may be preferable to include a terminator nucleotide at the 3' end of the recording tag to prevent transfer of coding tag information to the recording tag.

A coding tag may be a single stranded molecule, a double stranded molecule, or a partially double stranded. A coding tag may comprise blunt ends, overhanging ends, or one of each. In some embodiments, a coding tag is partially double stranded, which prevents annealing of the coding tag to internal encoder and spacer sequences in a growing extended recording tag.

A coding tag is joined to a binding agent directly or indirectly, by any means known in the art, including covalent and non-covalent interactions. In some embodiments, a coding tag may be joined to binding agent enzymatically or chemically. In some embodiments, a coding tag may be joined to a binding agent via ligation. In other embodiments, a coding tag is joined to a binding agent via affinity binding pairs (*e.g.*, biotin and streptavidin).

In some embodiments, a binding agent is joined to a coding tag via SpyCatcher-SpyTag interaction (see, Figure 43B). The SpyTag peptide forms an irreversible covalent bond to the SpyCatcher protein via a spontaneous isopeptide linkage, thereby offering a genetically encoded way to create peptide interactions that resist force and harsh conditions (Zakeri et al., 2012, Proc. Natl. Acad. Sci. 109:E690-697; Li et al., 2014, J. Mol. Biol. 426:309-317). A binding agent may be expressed as a fusion protein comprising the SpyCatcher protein. In some embodiments, the SpyCatcher protein is appended on the N-terminus or C-terminus of the binding agent. The SpyTag peptide can be coupled to the coding tag using standard conjugation chemistries (Bioconjugate Techniques, G. T. Hermanson, Academic Press (2013)).

In other embodiments, a binding agent is joined to a coding tag via SnoopTag-SnoopCatcher peptide-protein interaction. The SnoopTag peptide forms an isopeptide bond with the SnoopCatcher protein (Veggiani et al., Proc. Natl. Acad. Sci. USA, 2016, 113:1202-1207). A binding agent may be expressed as a fusion protein comprising the SnoopCatcher protein. In some embodiments, the SnoopCatcher protein is appended on the N-terminus or C-terminus of the binding agent. The SnoopTag peptide can be coupled to the coding tag using standard conjugation chemistries.

In yet other embodiments, a binding agent is joined to a coding tag via the HaloTag® protein fusion tag and its chemical ligand. HaloTag is a modified haloalkane dehalogenase designed to covalently bind to synthetic ligands (HaloTag ligands) (Los et al., 2008, ACS Chem. Biol. 3:373-382). The synthetic ligands comprise a chloroalkane linker attached to a variety of useful molecules. A covalent bond forms between the HaloTag and the chloroalkane linker that is highly specific, occurs rapidly under physiological conditions, and is essentially irreversible.

In certain embodiments, a macromolecule is also contacted with a non-cognate binding agent. As used herein, a non-cognate binding agent is referring to a binding agent that is selective for a different macromolecule feature or component than the particular macromolecule being considered. For example, if the *n* NTAA is phenylalanine, and the peptide is contacted with three binding agents selective for phenylalanine, tyrosine, and asparagine, respectively, the binding agent selective for phenylalanine would be first binding agent capable of selectively binding to the n^{th} NTAA (*i.e.,* phenylalanine), while the other two binding agents would be non-cognate binding agents for that peptide (since they are selective for NTAAs other than phenylalanine). The tyrosine and asparagine binding agents may, however, be cognate binding agents for other peptides in the sample. If the *n* NTAA (phenylalanine) was then cleaved from the peptide, thereby converting the *n-1* amino acid of the peptide to the *n-1* NTAA (*e.g.,* tyrosine), and the peptide was then contacted with the same three binding agents, the binding agent selective for tyrosine would be second binding agent capable of selectively binding to the *n-1* NTAA (*i.e.,* tyrosine), while the other two binding agents would be non-cognate binding agents (since they are selective for NTAAs other than tyrosine).

Thus, it should be understood that whether an agent is a binding agent or a non-cognate binding agent will depend on the nature of the particular macromolecule feature or component currently available for binding. Also, if multiple macromolecules are analyzed in a multiplexed reaction, a binding agent for one macromolecule may be a non-cognate binding agent for another, and vice versa. According, it should be understood that the following description concerning binding agents is applicable to any type of binding agent described herein (*i.e.*, both cognate and non-cognate binding agents).

VIII. Cyclic Transfer of Coding Tag Information to Recording Tags

In the methods described herein, upon binding of a binding agent to a macromolecule, identifying information of its linked coding tag is transferred to a recording tag associated with the macromolecule, thereby generating an "extended recording tag." An extended recording tag may comprise information from a binding agent's coding tag representing each binding cycle performed. However, an extended recording tag may also experience a "missed" binding cycle, e.g., because a binding agent fails to bind to the macromolecule, because the coding tag was missing, damaged, or defective, because the primer extension reaction failed. Even if a binding event occurs, transfer of information from the coding tag to the recording tag may be incomplete or less than 100% accurate, e.g., because a coding tag was damaged or defective, because errors were introduced in the primer extension reaction). Thus, an extended recording tag may represent 100%, or up to 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 65%, 55%, 50%, 45%, 40%, 35%, 30% of binding events that have occurred on its associated macromolecule. Moreover, the coding tag information present in the extended recording tag may have at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% identity the corresponding coding tags.

In certain embodiments, an extended recording tag may comprise information from multiple coding tags representing multiple, successive binding events. In these embodiments, a single, concatenated extended recording tag can be representative of a single macromolecule (*see*, Figure 2A). As referred to herein, transfer of coding tag information to a recording tag also includes transfer to an extended recording tag as would occur in methods involving multiple, successive binding events.

In certain embodiments, the binding event information is transferred from a coding tag to a recording tag in a cyclic fashion (see Figures 2A and 2C). Cross-reactive binding events can be informatically filtered out after sequencing by requiring that at least two different coding tags, identifying two or more independent binding events, map to the same class of binding agents (cognate to a particular protein). An optional sample or compartment barcode can be included in the recording tag, as well an optional UMI sequence. The coding tag can also contain an optional UMI sequence along with the encoder and spacer sequences. Universal priming sequences (U1 and U2) may also be included in extended recording tags for amplification and NGS sequencing (see Figure 2A).

Coding tag information associated with a specific binding agent may be transferred to a recording tag using a variety of methods. In certain embodiments, information of a coding tag is transferred to a recording tag via primer extension (Chan, McGregor et al. 2015). A spacer sequence on the 3'-terminus of a recording tag or an extended recording tag anneals with complementary spacer sequence on the 3' terminus of a coding tag and a polymerase (*e.g.*, strand-displacing polymerase) extends the recording tag sequence, using the annealed coding tag as a template (*see*, Figures 5-7). In some embodiments, oligonucleotides complementary to coding tag encoder sequence and 5' spacer can be pre-annealed to the coding tags to prevent hybridization of the coding tag to internal encoder and spacer sequences present in an extended recording tag. The 3' terminal spacer, on the coding tag, remaining single stranded, preferably binds to the terminal 3' spacer on the recording tag. In other embodiments, a nascent recording tag can be coated with a single stranded binding protein to prevent annealing of the coding tag to internal sites. Alternatively, the nascent recording tag can also be coated with RecA (or related homologues such as uvsX) to facilitate invasion of the 3' terminus into a completely double stranded coding tag (Bell et al., 2012, Nature 491:274-278). This configuration prevents the double stranded coding tag from interacting with internal recording tag elements, yet is susceptible to strand invasion by the RecA coated 3' tail of the extended recording tag (Bell, et al., 2015, Elife 4:

e08646). The presence of a single-stranded binding protein can facilitate the strand displacement reaction.

In a preferred embodiment, a DNA polymerase that is used for primer extension possesses strand-displacement activity and has limited or is devoid of 3'-5 exonuclease activity. Several of many examples of such polymerases include Klenow exo- (Klenow fragment of DNA Pol 1), T4 DNA polymerase exo-, T7 DNA polymerase exo (Sequenase 2.0), Pfu exo-, Vent exo-, Deep Vent exo-, Bst DNA polymerase large fragment exo-, Bca Pol, 9°N Pol, and Phi29 Pol exo-. In a preferred embodiment, the DNA polymerase is active at room temperature and up to 45°C. In another embodiment, a "warm start" version of a thermophilic polymerase is employed such that the polymerase is activated and is used at about 40°C-50°C. An exemplary warm start polymerase is Bst 2.0 Warm Start DNA Polymerase (New England Biolabs).

Additives useful in strand-displacement replication include any of a number of single-stranded DNA binding proteins (SSB proteins) of bacterial, viral, or eukaryotic origin, such as SSB protein of E. coli, phage T4 gene 32 product, phage T7 gene 2.5 protein, phage Pf3 SSB, replication protein A RPA32 and RPA14 subunits (Wold, 1997); other DNA binding proteins, such as adenovirus DNA-binding protein, herpes simplex protein ICP8, BMRF1 polymerase accessory subunit, herpes virus UL29 SSB-like protein; any of a number of replication complex proteins known to participate in DNA replication, such as phage T7 helicase/primase, phage T4 gene 41 helicase, E. coli Rep helicase, E. coli recBCD helicase, recA, E. coli and eukaryotic topoisomerases (Champoux, 2001).

Mis-priming or self-priming events, such as when the terminal spacer sequence of the recoding tag primes extension self-extension may be minimized by inclusion of single stranded binding proteins (T4 gene 32, E. coli SSB, etc.), DMSO (1-10%), formamide (1-10%), BSA(10-100 ug/ml), TMACl (1-5 mM), ammonium sulfate (10-50 mM), betaine (1-3 M), glycerol (5-40%), or ethylene glycol (5-40%), in the primer extension reaction.

Most type A polymerases are devoid of 3' exonuclease activity (endogenous or engineered removal), such as Klenow exo-, T7 DNA polymerase exo-

(Sequenase 2.0), and Taq polymerase catalyzes non-templated addition of a nucleotide, preferably an adenosine base (to lesser degree a G base, dependent on sequence context) to the 3' blunt end of a duplex amplification product. For Taq polymerase, a 3' pyrimidine (C>T) minimizes non-templated adenosine addition, whereas a 3' purine nucleotide (G>A) favours non-templated adenosine addition. In embodiments using Taq polymerase for primer extension, placement of a thymidine base in the coding tag between the spacer sequence distal from the binding agent and the adjacent barcode sequence (e.g., encoder sequence or cycle specific sequence) accommodates the sporadic inclusion of a non-templated adenosine nucleotide on the 3' terminus of the spacer sequence of the recording tag. (Figure 43A). In this manner, the extended recording tag (with or without a non-templated adenosine base) can anneal to the coding tag and undergo primer extension.

Alternatively, addition of non-templated base can be reduced by employing a mutant polymerase (mesophilic or thermophilic) in which non-templated terminal transferase activity has been greatly reduced by one or more point mutations, especially in the O-helix region (see U.S. Patent 7,501,237) (Yang, Astatke et al. 2002). Pfu exo-, which is 3' exonuclease deficient and has strand-displacing ability, also does not have non-templated terminal transferase activity.

In another embodiment, optimal polymerase extension buffers are comprised of 40-120 mM buffering agent such as Tris-Acetate, Tris-HCl, HEPES, etc. at a pH of 6-9.

Self-priming/mis-priming events initiated by self-annealing of the terminal spacer sequence of the extended recording tag with internal regions of the extended recording tag may be minimized by including pseudo-complementary bases in the recording/extended recording tag (Lahoud, Timoshchuk et al. 2008), (Hoshika, Chen et al. 2010). Pseudo-complementary bases show significantly reduced hybridization affinities for the formation of duplexes with each other due the presence of chemical modification. However, many pseudo-complementary modified bases can form strong base pairs with natural DNA or RNA sequences. In certain embodiments, the coding tag spacer sequence is comprised of multiple A and T bases, and commercially available pseudo-complementary bases 2-aminoadenine and 2-thiothymine are incorporated in the recording tag using phosphoramidite oligonucleotide synthesis. Additional pseudocomplementary bases can be incorporated into the extended recording tag during primer extension by adding pseudo-complementary nucleotides to the reaction (Gamper, Arar et al. 2006).

To minimize non-specific interaction of the coding tag labeled binding agents in solution with the recording tags of immobilized proteins, competitor (also referred to as blocking) oligonucleotides complementary to recording tag spacer sequences are added to binding reactions to minimize non-specific interaction s (Figure 32A-D). Blocking oligonucleotides are relatively short. Excess competitor oligonucleotides are washed from the binding reaction prior to primer extension, which effectively dissociates the annealed competitor oligonucleotides from the recording tags, especially when exposed to slightly elevated temperatures (e.g., 30-50 °C). Blocking oligonucleotides may comprise a terminator nucleotide at its 3' end to prevent primer extension.

In certain embodiments, the annealing of the spacer sequence on the recording tag to the complementary spacer sequence on the coding tag is metastable under the primer extension reaction conditions (i.e., the annealing Tm is similar to the reaction temperature). This allows the spacer sequence of the coding tag to displace any blocking oligonucleotide annealed to the spacer sequence of the recording tag.

Coding tag information associated with a specific binding agent may also be transferred to a recording tag via ligation (see, e.g., Figures 6 and 7). Ligation may be a blunt end ligation or sticky end ligation. Ligation may be an enzymatic ligation reaction. Examples of ligases include, but are not limited to T4 DNA ligase, T7 DNA ligase, T3 DNA ligase, Taq DNA ligase, E. coli DNA ligase, 9°N DNA ligase, Electroligase®. Alternatively, a ligation may be a chemical ligation reaction (see Figure 7). In the illustration, a spacer-less ligation is accomplished by using hybridization of a "recording helper" sequence with an arm on the coding tag. The annealed complement sequences are chemically ligated using standard chemical ligation or "click chemistry" (Gunderson, Huang et al. 1998, Peng, Li et al. 2010, El-Sagheer, Cheong et al. 2011, El-Sagheer, Sanzone et al. 2011, Sharma, Kent et al. 2012, Roloff and Seitz 2013, Litovchick, Clark et al. 2014, Roloff, Ficht et al. 2014).

In another embodiment, transfer of PNAs can be accomplished with chemical ligation using published techniques. The structure of PNA is such that it has a 5' N-terminal amine group and an unreactive 3' C-terminal amide. Chemical ligation of PNA requires that the termini be modified to be chemically active. This is typically done by derivitizing the 5' N-terminus with a cysteinyl moiety and the 3' C-terminus with a thioester moiety. Such modified PNAs easily couple using standard native chemical ligation conditions (Roloff et al., 2013, Bioorgan. Med. Chem. 21:3458-3464).

In some embodiments, coding tag information can be transferred using topoisomerase. Topoisomerase can be used be used to ligate a topo-charged 3' phosphate on the recording tag to the 5' end of the coding tag, or complement thereof (Shuman et al., 1994, J. Biol. Chem. 269:32678-32684).

As described herein, a binding agent may bind to a post-translationally modified amino acid. Thus, in certain embodiments involving peptide macromolecules, an extended recording tag comprises coding tag information relating to amino acid sequence and post-translational modifications. In some embodiments, detection of internal post-translationally modified amino acids (e.g., phosphorylation, glycosylation, succinylation, ubiquitination, S-Nitrosylation, methylation, N-acetylation, lipidation, etc.) is be accomplished prior to detection and cleavage of terminal amino acids (*e.g.*, NTAA or CTAA). In one example, a peptide is contacted with binding agents for PTM modifications, and associated coding tag information are transferred to the recording tag as described above (see Figure 8A). Once the detection and transfer of coding tag information relating to amino acid modifications is complete, the PTM modifying groups can be removed before detection and transfer of coding tag information for the primary amino acid sequence using N-terminal or C-terminal degradation methods. Thus, resulting extended recording tags indicate the presence of post-translational modifications in a peptide sequence, though not the sequential order, along with primary amino acid sequence information (see Figure 8B).

In some embodiments, detection of internal post-translationally modified amino acids may occur concurrently with detection of primary amino acid sequence. In one example, an NTAA (or CTAA) is contacted with a binding agent specific for a post-translationally modified amino acid, either alone or as part of a library of binding agents (e.g., library composed of binding agents for the 20 standard amino acids and selected post-translational modified amino acids). Successive cycles of terminal amino acid cleavage and contact with a binding agent (or library of binding agents) follow. Thus, resulting extended recording tags indicate the presence and order of post-translational modifications in the context of a primary amino acid sequence.

In certain embodiments, an ensemble of recording tags may be employed per macromolecule to improve the overall robustness and efficiency of coding tag information transfer (see, e.g., Figure 9). The use of an ensemble of recording tags associated with a given macromolecule rather than a single recording tag improves the efficiency of library construction due to potentially higher coupling yields of coding tags to recording tags, and higher overall yield of libraries. The yield of a single concatenated extended recording tag is directly dependent on the stepwise yield of concatenation, whereas the use of multiple recording tags capable of accepting coding tag information does not suffer the exponential loss of concatenation.

An example of such an embodiment is shown in Figures 9 and 10. In Figure 9A and 10A, multiple recording tags are associated with a single macromolecule (by spatial co-localization or confinement of a single macromolecule to a single bead) on a solid support. Binding agents are exposed to the solid support in cyclical fashion and their corresponding coding tag transfers information to one of the co-localized multiple recording tags in each cycle. In the example shown in Figure 9A, the binding cycle information is encoded into the spacer present on the coding tag. For each binding cycle, the set of binding agents is marked with a designated cycle-specific spacer sequence (Figure 9A and 9B). For example, in the case of NTAA binding agents, the binding agents to the same amino acid residue are be labelled with different coding tags or comprise cycle-specific information in the spacer sequence to denote both the binding agent identity and cycle number.

As illustrated in Figure 9A, in a first cycle of binding (Cycle 1), a plurality of NTAA binding agents is contacted with the macromolecule. The binding agents used in Cycle 1 possess a common spacer sequence that is complementary to the spacer sequence of the recording tag. The binding agents used in Cycle 1 also possess a 3'-spacer sequence comprising Cycle 1 specific sequence. During binding Cycle 1, a first NTAA binding agent binds to the free terminus of the macromolecule, the complementary sequences of the common spacer sequence in the first coding tag and recording tag anneal, and the information of a first coding tag is transferred to a cognate recording tag via primer extension from the common spacer sequence. Following removal of the NTAA to expose a new NTAA, binding Cycle 2 contacts a plurality of NTAA binding agents that possess a common spacer sequence that is complementary to the spacer sequence of a recording tag. The binding agents used in Cycle 2 also possess a 3'-spacer sequence comprising Cycle 2 specific sequence. A second NTAA binding agent binds to the NTAA of the macromolecule, and the information of a second coding tag is transferred to a recording tag via primer extension. These cycles are repeated up to "n" binding cycles, generating a plurality of extended recording tags co-localized with the single macromolecule, wherein each extended recording tag possesses coding tag information from one binding cycle. Because each set of binding agents used in each successive binding cycle possess cycle specific spacer sequences in the coding tags, binding cycle information can be associated with binding agent information in the resulting extended recording tags

In an alternative embodiment, multiple recording tags are associated with a single macromolecule on a solid support (*e.g.*, bead) as in Figure 9A, but in this case binding agents used in a particular binding cycle have coding tags flanked by a cycle-specific spacer for the current binding cycle and a cycle specific spacer for the next binding cycle (Figures 10A and 10B). The reason for this design is to support a final assembly PCR step (Figure 10C) to convert the population of extended recording tags into a single co-linear, extended recording tag. A library of single, co-linear extended recording tag can be subjected to enrichment, subtraction and/or normalization methods prior to sequencing. In the first binding cycle (Cycle 1), upon binding of a first binding agent, the information of a coding tag comprising a Cycle 1 specific spacer (C'1) is transferred to a recording tag comprising a complementary Cycle 1 specific spacer (C1) at its terminus. In the second binding cycle (Cycle 2), upon binding of a second binding agent, the information of a coding tag comprising a Cycle 2 specific spacer (C'2) is transferred to a different recording tag comprising a complementary Cycle 2 specific spacer (C2) at its terminus. This process continues until the n^{th} binding cycle. In some embodiments, the n^{th} coding tag in the extended recording tag is capped with a universal reverse priming sequence, *e.g.*, the universal reverse priming sequence can be incorporated as part of the n^{th} coding tag design or the universal reverse priming sequence can be added in a subsequent reaction after the n^{th} binding cycle, such as an amplification reaction using a tailed primer. In some embodiments, at each binding cycle a macromolecule is exposed to a collection of binding agents joined to coding tags comprising identifying information regarding their corresponding binding agents and binding cycle information (Figure 9 and Figure 10). In a particular embodiment, following completion of the n^{th} binding cycle, the bead substrates coated with extended recording tags are placed in an oil emulsion such that on average there is fewer than or approximately equal to 1 bead/droplet. Assembly PCR is then used to amplify the extended recording tags from the beads, and the multitude of separate recording tags are assembled collinear order by priming via the cycle specific spacer sequences within the separate extended recording tags (Figure 10C) (Xiong et al., 2008, FEMS Microbiol. Rev.32:522-540). Alternatively, instead of using cycle-specific spacer with the binding agents' coding tags, a cycle specific spacer can be added separately to the extended recording tag during or after each binding cycle. One advantage of using a population of extended recording tags, which collectively represent a single macromolecule vs. a single concatenated extended recording tag representing a single macromolecule is that a higher concentration of recording tags can increase efficiency of transfer of the coding tag information. Moreover, a binding cycle can be repeated several times to ensure completion of cognate binding events. Furthermore, surface amplification of extended recording tags may be able to provide redundancy of information transfer (see Figure 4B). If coding tag information is not always transferred, it should in most cases still be possible to use the incomplete collection of coding tag information to identify macromolecules that have very high information content, such as proteins. Even a short peptide can embody a very large number of possible protein sequences. For example, a 10-mer peptide has 20^{10} possible sequences. Therefore, partial or incomplete sequence that may contain deletions and/or ambiguities can often still be mapped uniquely.

In some embodiments, in which proteins in their native conformation are being queried, the cyclic binding assays are performed with binding agents harbouring coding tags comprised of a cleavable or nickable DNA strand within the spacer element proximal to the binding agent (Figure 32). For example, the spacer proximal to the binding agent may have one or more uracil bases that can be nicked by uracil-specific excision reagent (USER). In another example, the spacer proximal to the binding agent may comprise a recognition sequence for a nicking endonuclease that hydrolyzes only one strand of a duplex. This design allows the non-denaturing removal of the binding agent from the extended recording tag and creates a free single stranded DNA spacer element for subsequent immunoassay cycles. In a preferred embodiment, a uracil base is incorporated into the coding tag to permit enzymatic USER removal of the binding agent after the primer extension step (Figures 32E-F). After USER excision of uracils, the binding agent and truncated coding tag can be removed under a variety of mild conditions including high salt (4M NaCl, 25% formamide) and mild heat to disrupt the protein-binding agent interaction. The other truncated coding tag DNA stub remaining annealed on the recording tag (Figure 32F) readily dissociates at slightly elevated temperatures.

Coding tags comprised of a cleavable or nickable DNA strand within the spacer element proximal to the binding agent also allows for a single homogeneous assay for transferring of coding tag information from multiple bound binding agents (see Figure 33). In a preferred embodiment, the coding tag proximal to the binding agent comprises a nicking endonuclease sequence motif, which is recognized and nicked by a nicking endonuclease at a defined sequence motif in the context of dsDNA. After binding of multiple binding agents, a combined polymerase extension (devoid of strand-displacement activity) + nicking endonuclease reagent mix is used to generate repeated transfers of coding tags to the proximal recording tag or extended recording tag. After each transfer step, the resulting extended recording tag-coding tag duplex is nicked by the nicking endonuclease releasing the truncated spacer attached to the binding agent and exposing the extended recording tag 3' spacer sequence, which is capable of annealing to the coding tags of additional proximal bound binding agents (Figures 33B-D). The placement of the nicking motif in the coding tag spacer sequence is designed to create a metastable hybrid, which can easily be exchanged with a non-cleaved coding tag spacer sequence. In this way, if two or more binding agents simultaneously bind the same protein molecule, binding information via concatenation of coding tag information from multiply bound binding agents onto the recording tag occurs in a single reaction mix without any cyclic reagent exchanges (Figures 33C-D). This embodiment is particularly useful for the next generation protein assay (NGPA), especially with polyclonal antibodies (or mixed population of monoclonal antibody) to multivalent epitopes on a protein.

For embodiments involving analysis of denatured proteins, polypeptides, and peptides, the bound binding agent and annealed coding tag can be removed following primer extension by using highly denaturing conditions (e.g., 0.1-0.2 N NaOH, 6M Urea, 2.4 M guanidinium isothiocyanate, 95% formamide, etc.).

IX. Cyclic Transfer of Recording Tag Information to Coding Tags or Di-Tag Constructs

In another aspect, rather than writing information from the coding tag to the recording tag following binding of a binding agent to a macromolecule, information may be transferred from the recording tag comprising an optional UMI sequence (*e.g.* identifying a particular peptide or protein molecule) and at least one barcode (e.g., a compartment tag, partition barcode, sample barcode, spatial location barcode, etc.), to the coding tag, thereby generating an extended coding tag (see Figure 11A). In certain embodiments, the binding agents and associated extended coding tags are collected following each binding cycle and, optionally, prior to Edman degradation chemistry steps. In certain embodiments, the coding tags comprise a binding cycle specific tag. After completion of all the binding cycles, such as detection of NTAAs in cyclic Edman degradation, the complete collection of extended coding tags can be amplified and sequenced, and information on the peptide determined from the association between UMI (peptide identity), encoder sequence (NTAA binding agent), compartment tag (single cell or subset of proteome), binding cycle specific sequence (cycle number), or any combination thereof. Library elements with the same compartment tag/UMI sequence map back to the same cell, subset of proteome, molecule, etc. and the peptide sequence can be reconstructed. This embodiment may be useful in cases where the recording tag sustains too much damage during the Edman degradation process.

Provided herein are methods for analyzing a plurality of macromolecules, comprising: (a) providing a plurality of macromolecules and associated recording tags joined to a solid support; (b) contacting the plurality of macromolecules with a plurality of binding agents capable of binding to the plurality of macromolecules, wherein each binding agent comprises a coding tag with identifying information regarding the binding agent; (c) (i) transferring the information of the macromolecule associated recording tags to the coding tags of the binding agents that are bound to the macromolecules to generate extended coding tags (see Figure 11A); or (ii) transferring the information of macromolecule associated recording tags and coding tags of the binding agents that are bound to the macromolecules to a di-tag construct (see Figure 11B); (d) collecting the extended coding tags or di-tag constructs; (e) optionally repeating steps (b) – (d) for one or more binding cycles; (f) analyzing the collection of extended coding tags or di-tag constructs.

In certain embodiments, the information transfer from the recording tag to the coding tag can be accomplished using a primer extension step where the 3' terminus of recording tag is optionally blocked to prevent primer extension of the recording tag (see, e.g., Figure 11A). The resulting extended coding tag and associated binding agent can be collected after each binding event and completion of information transfer. In an example illustrated in Figure 11B, the recording tag is comprised of a universal priming site (U2'), a barcode (e.g., compartment tag "CT"), an optional UMI sequence, and a common spacer sequence (Sp1). In certain embodiments, the barcode is a compartment tag representing an individual compartment, and the UMI can be used to map sequence reads back to a particular protein or peptide molecule being queried. As illustrated in the example in Figure 11B, the coding tag is comprised of a common spacer sequence (Sp2'), a binding agent encoder sequence, and universal priming site (U3). Prior to the introduction of the coding tag-labeled binding agent, an oligonucleotide (U2) that is complementary to the U2' universal priming site of the recording tag and comprises a universal priming sequence U1 and a cycle specific tag, is annealed to the recording tag U2'. Additionally, an adapter sequence, Sp1'-Sp2, is annealed to the recording tag Sp1. This adapter sequence also capable of interacting with the Sp2' sequence on the coding tag, bringing the recording tag and coding tag in proximity to each other. A gap-fill extension ligation assay is performed either prior to or after the binding event. If the gap fill is performed before the binding cycle, a post-binding cycle primer extension step is used to complete di-tag formation. After collection of di-tags across a number of binding cycles, the collection of di-tags is sequenced, and mapped back to the originating peptide molecule via the UMI sequence. It is understood that to maximize efficacy, the diversity of the UMI sequences must exceed the diversity of the number of single molecules tagged by the UMI.

In certain embodiments, the macromolecule is a protein or a peptide. The peptide may be obtained by fragmenting a protein from a biological sample.

The recording tag may be a DNA molecule, RNA molecule, PNA molecule, BNA molecule, XNA molecule, LNA molecule a γPNA molecule, or a combination thereof. The recording tag comprises a UMI identifying the macromolecule (*e.g.*, peptide) to which it is associated. In certain embodiments, the recording tag further comprises a compartment tag. The recording tag may also comprise a universal priming site, which may be used for downstream amplification. In certain embodiments, the recording tag comprises a spacer at its 3' terminus. A spacer may be complementary to a spacer in the coding tag. The 3'-terminus of the recording tag may be blocked (*e.g.*, photo-labile 3' blocking group) to prevent extension of the recording tag by a polymerase, facilitating transfer of information of the macromolecule associated recording tag to the coding tag or transfer of information of the macromolecule associated recording tag and coding tag to a di-tag construct.

The coding tag comprises an encoder sequence identifying the binding agent to which the coding agent is linked. In certain embodiments, the coding tag further comprises a unique molecular identifier (UMI) for each binding agent to which the coding tag is linked. The coding tag may comprise a universal priming site, which may be used for downstream amplification. The coding tag may comprise a spacer at its 3'-terminus. The spacer may be complementary to the spacer in the recording tag and can be used to initiate a primer extension reaction to transfer recording tag information to the coding tag. The coding tag may also comprise a binding cycle specific sequence, for identifying the binding cycle from which an extended coding tag or di-tag originated.

Transfer of information of the recording tag to the coding tag may be effected by primer extension or ligation. Transfer of information of the recording tag and coding tag to a di-tag construct may be generated a gap fill reaction, primer extension reaction, or both.

A di-tag molecule comprises functional components similar to that of an extended recording tag. A di-tag molecule may comprise a universal priming site derived from the recording tag, a barcode (e.g., compartment tag) derived from the recording tag, an optional unique molecular identifier (UMI) derived from the recording tag, an optional spacer derived from the recording tag, an encoder sequence derived from the coding tag, an optional unique molecular identifier derived from the coding tag, a binding cycle specific sequence, an optional spacer derived from the coding tag, and a universal priming site derived from the coding tag.

In certain embodiments, the recording tag can be generated using combinatorial concatenation of barcode encoding words. The use of combinatorial encoding words provides a method by which annealing and chemical ligation can be used to transfer information from a PNA recording tag to a coding tag or di-tag construct (see, e.g., Figures 12A-D). In certain embodiments where the methods of analyzing a peptide disclosed herein involve cleavage of a terminal amino acid via an

Edman degradation, it may be desirable employ recording tags resistant to the harsh conditions of Edman degradation, such as PNA. One harsh step in the Edman degradation protocol is anhydrous TFA treatment to cleave the N-terminal amino acid. This step will typically destroy DNA. PNA, in contrast to DNA, is highly-resistant to acid hydrolysis. The challenge with PNA is that enzymatic methods of information transfer become more difficult, *i.e.*, information transfer via chemical ligation is a preferred mode. In Figure 11B, recording tag and coding tag information are written using an enzymatic gap-fill extension ligation step, but this is not currently feasibly with PNA template, unless a polymerase is developed that uses PNA. The writing of the barcode and UMI from the PNA recording tag to a coding tag is problematic due to the requirement of chemical ligation, products which are not easily amplified. Methods of chemical ligation have been extensively described in the literature (Gunderson et al. 1998, Genome Res. 8:1142-1153; Peng et al., 2010, Eur. J. Org. Chem. 4194-4197; El-Sagheer et al., 2011, Org. Biomol. Chem. 9:232-235; El-Sagheer et al., 2011, Proc. Natl. Acad. Sci. USA 108:11338-11343; Litovchick et al., 2014, Artif. DNA PNA XNA 5: e27896; Roloff et al., 2014, Methods Mol. Biol. 1050:131-141).

To create combinatorial PNA barcodes and UMI sequences, a set of PNA words from an n-mer library can be combinatorially ligated. If each PNA word derives from a space of 1,000 words, then four combined sequences generate a coding space of $1{,}000^4 = 10^{12}$ codes. In this way, from a starting set of 4,000 different DNA template sequences, over 10^{12} PNA codes can be generated (Figure 12A). A smaller or larger coding space can be generated by adjusting the number of concatenated words, or adjusting the number of elementary words. As such, the information transfer using DNA sequences hybridized to the PNA recording tag can be completed using DNA word assembly hybridization and chemical ligation (see Figure 12B). After assembly of the DNA words on the PNA template and chemical ligation of the DNA words, the resulting intermediate can be used to transfer information to/from the coding tag (see Figure 12C and Figure 12D).

In certain embodiments, the macromolecule and associated recording tag are covalently joined to the solid support. The solid support may be a bead, an array, a glass surface, a silicon surface, a plastic surface, a filter, a membrane, nylon, a silicon wafer chip, a flow through chip, a biochip including signal transducing electronics, a microtiter well, an ELISA plate, a spinning interferometry disc, a nitrocellulose membrane, a nitrocellulose-based polymer surface, a nanoparticle, or a microsphere. The solid support may be a polystyrene bead, a polymer bead, an agarose bead, an acrylamide bead, a solid core bead, a porous bead, a paramagnetic bead, a glass bead, or a controlled pore bead.

In certain embodiments, the binding agent is a protein or a polypeptide. In some embodiments, the binding agent is a modified or variant aminopeptidase, a modified or variant amino acyl tRNA synthetase, a modified or variant anticalin, a modified or variant ClpS, or a modified or variant antibody or binding fragment thereof. In certain embodiments, the binding agent binds to a single amino acid residue, a di-peptide, a tri-peptide, or a post-translational modification of the peptide. In some embodiments, the binding agent binds to an N-terminal amino acid residue, a C-terminal amino acid residue, or an internal amino acid residue. In some embodiments, the binding agent binds to an N-terminal peptide, a C-terminal peptide, or an internal peptide. In some embodiments, the binding agent is a site-specific covalent label of an amino acid of post-translational modification of a peptide.

In certain embodiments, following contacting the plurality of macromolecules with a plurality of binding agents in step (b), complexes comprising the macromolecule and associated binding agents are dissociated from the solid support and partitioned into an emulsion of droplets or microfluidic droplets. In some embodiments, each microfluidic droplet comprises at most one complex comprising the macromolecule and the binding agents.

In certain embodiments, the recording tag is amplified prior to generating an extended coding tag or di-tag construct. In embodiments where complexes comprising the macromolecule and associated binding agents are partitioned into droplets or microfluidic droplets such that there is at most one complex per droplet, amplification of recording tags provides additional recording tags as templates for transferring information to coding tags or di-tag constructs (see Figure 13 and Figure

14). Emulsion fusion PCR may be used to transfer the recording tag information to the coding tag or to create a population of di-tag constructs.

The collection of extended coding tags or di-tag constructs that are generated may be amplified prior to analysis. Analysis of the collection of extended coding tags or di-tag constructs may comprise a nucleic acid sequencing method. The sequencing by synthesis, sequencing by ligation, sequencing by hybridization, polony sequencing, ion semiconductor sequencing, or pyrosequencing. The nucleic acid sequencing method may be single molecule real-time sequencing, nanopore-based sequencing, or direct imaging of DNA using advanced microscopy.

Edman degradation and methods that chemically label N-terminal amines such as PITC, Sanger's agent (DNFB), SNFB, acetylation reagents, amidination (guanidination) reagents, etc. can also modify internal amino acids and the exocyclic amines on standard nucleic acid or PNA bases such as adenine, guanine, and cytosine. In a certain embodiments, the peptide's ε-amines of lysine residues are blocked with an acid anhydride, a guandination agent, or similar blocking reagent, prior to sequencing. Although exocyclic amines of DNA bases are much less reactive the primary N-terminal amine of peptides, controlling the reactivity of amine reactive agents toward N-terminal amines reducing non-target activity toward internal amino acids and exocyclic amines on DNA bases is important to the sequencing assay. The selectivity of the modification reaction can be modulated by adjusting reaction conditions such as pH, solvent (aqueous vs. organic, aprotic, non-polar, polar aprotic, ionic liquids, etc.), bases and catalysts, co-solvents, temperature, and time. In addition, reactivity of exocyclic amines on DNA bases is modulated by whether the DNA is in ssDNA or dsDNA form. To minimize modification, prior to NTAA chemical modification, the recording tag can be hybridized with complementary DNA probes: P1', {Sample BCs}', {Sp-BC}', etc. In another embodiment, the use of nucleic acids having protected exocyclic amines can also be used (Ohkubo, Kasuya et al. 2008). In yet another embodiment, "less reactive" amine labeling compounds, such as SNFB, mitigates off-target labeling of internal amino acids and exocylic amines on DNA (Carty and Hirs 1968). SNFB is less reactive than DNFB due to the fact that the para sulfonyl group is more electron withdrawing the para nitro group, leading to less active fluorine substitution with SNFB than DNFB.

Titration of coupling conditions and coupling reagents to optimize NTAA α-amine modification and minimize off-target amino acid modification or DNA modification is possible through careful selection of chemistry and reaction conditions (concentrations, temperature, time, pH, solvent type, etc.). For instance, DNFB is known to react with secondary amines more readily in aprotic solvents such as acetonitrile versus in water. Mild modification of the exocyclic amines may still allow a complementary probe to hybridize the sequence but would likely disrupt polymerase-based primer extension. It is also possible to protect the exocylic amine while still allowing hydrogen bonding. This was described in a recent publication in which protected bases are still capable of hybridizing to targets of interest (Ohkubo, Kasuya et al. 2008). In one embodiment, an engineered polymerase is used to incorporate nucleotides with protected bases during extension of the recording tag on a DNA coding tag template. In another embodiment, an engineered polymerase is used to incorporate nucleotides on a recording tag PNA template (w/ or w/o protected bases) during extension of the coding tag on the PNA recording tag template. In another embodiment, the information can be transferred from the recording tag to the coding tag by annealing an exogenous oligonucleotide to the PNA recording tag. Specificity of hybridization can be facilitated by choosing UMIs which are distinct in sequence space, such as designs based on assembly of n-mer words (Gerry, Witowski et al. 1999).

While Edman-like N-terminal peptide degradation sequencing can be used to determine the linear amino acid sequence of the peptide, an alternative embodiment can be used to perform partial compositional analysis of the peptide with methods utilizing extended recording tags, extended coding tags, and di-tags. Binding agents or chemical labels can be used to identify both N-terminal and internal amino acids or amino acid modifications on a peptide. Chemical agents can covalently modify amino acids (e.g., label) in a site-specific manner (Sletten and Bertozzi 2009, Basle, Joubert et al. 2010) (Spicer and Davis 2014). A coding tag can be attached to a chemical labeling agent that targets a single amino acid, to facilitate encoding and subsequent identification of site-specific labeled amino acids (see, Figure 13).

Peptide compositional analysis does not require cyclic degradation of the peptide, and thus circumvents issues of exposing DNA containing tags to harsh Edman chemistry. In a cyclic binding mode, one can also employ extended coding tags or di-tags to provide compositional information (amino acids or dipeptide/tripeptide information), PTM information, and primary amino acid sequence. In one embodiment, this composition information can be read out using an extended coding tag or di-tag approach described herein. If combined with UMI and compartment tag information, the collection of extended coding tags or di-tags provides compositional information on the peptides and their originating compartmental protein or proteins. The collection of extended coding tags or di-tags mapping back to the same compartment tag (and ostensibly originating protein molecule) is a powerful tool to map peptides with partial composition information. Rather than mapping back to the entire proteome, the collection of compartment tagged peptides is mapped back to a limited subset of protein molecules, greatly increasing the uniqueness of mapping.

Binding agents used herein may recognize a single amino acid, dipeptide, tripeptide, or even longer peptide sequence motifs. Tessler (2011, Digital Protein Analysis: Technologies for Protein Diagnostics and Proteomics through Single Molecule Detection. Ph.D., Washington University in St. Louis) demonstrated that relatively selective dipeptide antibodies can be generated for a subset of charged dipeptide epitopes (Tessler 2011). The application of directed evolution to alternate protein scaffolds (e.g., aaRSs, anticalins, ClpSs, etc.) and aptamers may be used to expand the set of dipeptide/tripeptide binding agents. The information from dipeptide/tripeptide compositional analysis coupled with mapping back to a single protein molecule may be sufficient to uniquely identify and quantitate each protein molecule. At a maximum, there are a total of 400 possible dipeptide combinations. However, a subset of the most frequent and most antigenic (charged, hydrophilic, hydrophobic) dipeptide should suffice to which to generate binding agents. This number may constitute a set of 40-100 different binding agents. For a set of 40 different binding agents, the average 10-mer peptide has about an 80% chance of being bound by at least one binding agent. Combining this information with all the peptides deriving from the same protein molecule may allow identification of the protein molecule. All this information about a peptide and its originating protein can be combined to give more accurate and precise protein sequence characterization.

A recent digital protein characterization assay has been proposed that uses partial peptide sequence information (Swaminathan et al., 2015, PLoS Comput. Biol. 11:e1004080) (Yao, Docter et al. 2015). Namely, the approach employs fluorescent labeling of amino acids which are easily labeled using standard chemistry such as cysteine, lysine, arginine, tyrosine, aspartate/glutamate (Basle, Joubert et al. 2010). The challenge with partial peptide sequence information is that the mapping back to the proteome is a one-to-many association, with no unique protein identified. This one-to-many mapping problem can be solved by reducing the entire proteome space to limited subset of protein molecules to which the peptide is mapped back. In essence, a single partial peptide sequence may map back to 100's or 1000's of different protein sequences, however if it is known that a set of several peptides (for example, 10 peptides originating from a digest of a single protein molecule) all map back to a single protein molecule contained in the subset of protein molecules within a compartment, then it is easier to deduce the identity of the protein molecule. For instance, an intersection of the peptide proteome maps for all peptides originating from the same molecule greatly restricts the set of possible protein identities (see Figure 15).

In particular, mappability of a partial peptide sequence or composition is significantly enhanced by making innovative use of compartmental tags and UMIs. Namely, the proteome is initially partitioned into barcoded compartments, wherein the compartmental barcode is also attached to a UMI sequence. The compartment barcode is a sequence unique to the compartment, and the UMI is a sequence unique to each barcoded molecule within the compartment (see Figure 16). In one embodiment, this partitioning is accomplished using methods similar to those disclosed in PCT Publication WO2016/061517, which is incorporated by reference in its entirety, by direct interaction of a DNA tag labeled polypeptide with the surface of a bead via hybridization to DNA compartment barcodes attached to the bead (see Figure 31). A primer extension step transfers information from the bead-linked compartment barcode to the DNA tag on the polypeptide (Figure 20). In another embodiment, this partitioning is accomplished by co-encapsulating UMI containing, barcoded beads and protein molecules into droplets of an emulsion. In addition, the droplet optionally contains a protease that digests the protein into peptides. A number of proteases can be used to digest the reporter tagged polypeptides (Switzar, Giera et al. 2013). Co-encapsulation of enzymatic ligases, such as butelase I, with proteases may will call for modification to the enzyme, such as pegylation, to make it resistant to protease digestion (Frokjaer and Otzen 2005, Kang, Wang et al. 2010). After digestion, the peptides are ligated to the barcode-UMI tags. In the preferred embodiment, the barcode-UMI tags are retained on the bead to facilitate downstream biochemical manipulations (see Figure 13).

After barcode-UMI ligation to the peptides, the emulsion is broken and the beads harvested. The barcoded peptides can be characterized by their primary amino acid sequence, or their amino acid composition. Both types of information about the peptide can be used to map it back to a subset of the proteome. In general, sequence information maps back to a much smaller subset of the proteome than compositional information. Nonetheless, by combining information from multiple peptides (sequence or composition) with the same compartment barcode, it is possible to uniquely identify the protein or proteins from which the peptides originate. In this way, the entire proteome can be characterized and quantitated. Primary sequence information on the peptides can be derived by performing a peptide sequencing reaction with extended recording tag creation of a DNA Encoded Library (DEL) representing the peptide sequence. In the preferred embodiment, the recording tag is comprised of a compartmental barcode and UMI sequence. This information is used along with the primary or PTM amino acid information transferred from the coding tags to generate the final mapped peptide information.

An alternative to peptide sequence information is to generate peptide amino acid or dipeptide/tripeptide compositional information linked to compartmental barcodes and UMIs. This is accomplished by subjecting the beads with UMI-barcoded peptides to an amino acid labeling step, in which select amino acids (internal) on each peptide are site-specifically labeled with a DNA tag comprising amino acid code information and another amino acid UMI (AA UMI) (see, Figure 13). The amino acids (AAs) most tractable to chemical labeling are lysines, arginines, cysteines, tyrosines, tryptophans, and aspartates/glutamates, but it may also be feasible to develop labeling schemes for the other AAs as well (Mendoza and Vachet, 2009). A given peptide may contain several AAs of the same type. The presence of multiple amino acids of the same type can be distinguished by virtue of the attached AA UMI label. Each labeling molecule has a different UMI within the DNA tag enabling counting of amino acids. An alternative to chemical labeling is to "label" the AAs with binding agents. For instance, a tyrosine-specific antibody labeled with a coding tag comprising AA code information and an AA UMI could be used mark all the tyrosines of the peptides. The caveat with this approach is the steric hindrance encountered with large bulky antibodies, ideally smaller scFvs, anticalins, or ClpS variants would be used for this purpose.

In one embodiment, after tagging the AAs, information is transferred between the recording tag and multiple coding tags associated with bound or covalently coupled binding agents on the peptide by compartmentalizing the peptide complexes such that a single peptide is contained per droplet and performing an emulsion fusion PCR to construct a set of extended coding tags or di-tags characterizing the amino acid composition of the compartmentalized peptide. After sequencing the di-tags, information on peptides with the same barcodes can be mapped back to a single protein molecule.

In a particular embodiment, the tagged peptide complexes are disassociated from the bead (see Figure 13), partitioned into small mini-compartments (e.g., micro-emulsion) such that on average only a single labeled/bound binding agent peptide complex resides in a given compartment. In a particular embodiment, this compartmentalization is accomplished through generation of micro-emulsion droplets (Shim, Ranasinghe et al. 2013, Shembekar, Chaipan et al. 2016). In addition to the peptide complex, PCR reagents are also co-encapsulated in the droplets along with three primers (U1, Sp, and $U2_{tr}$). After droplet formation, a few cycles of emulsion PCR are performed (~5-10 cycles) at higher annealing temperature such than only U1 and Sp anneal and amplify the recording tag product (see Figure 13). After this initial 5-10 cycles of PCR, the annealing temperature is reduced such that $U2_{tr}$ and the Sp_{tr} on the amino acid code tags participate in the amplification, and another ~10 rounds are performed. The three-primer emulsion PCR effectively combines the peptide UMI-barcode with all the AA code tags generating a di-tag library representation of the peptide and its amino acid composition. Other modalities of performing the three primer PCR and concatenation of the tags can also be employed. Another embodiment is the use of a 3' blocked U2 primer activated by photo-deblocking, or addition of an oil soluble reductant to initiate 3' deblocking of a labile blocked 3' nucleotide. Post-emulsion PCR, another round of PCR can be performed with common primers to format the library elements for NGS sequencing.

In this way, the different sequence components of the library elements are used for counting and classification purposes. For a given peptide (identified by the compartment barcode-UMI combination), there are many library elements, each with an identifying AA code tag and AA UMI (see Figure 13). The AA code and associated UMI is used to count the occurrences of a given amino acid type in a given peptide. Thus the peptide (perhaps a GluC, LysC, or Endo AsnN digest) is characterized by its amino acid composition (e.g., 2 Cys,1 Lys, 1 Arg, 2 Tyr, etc.) without regard to spatial ordering. This nonetheless provides a sufficient signature to map the peptide to a subset of the proteome, and when used in combination with the other peptides derived from the same protein molecule, to uniquely identify and quantitate the protein.

X. Terminal Amino Acid (TAA) Labelling Methods

In certain embodiments, a terminal amino acid (*e.g.*, NTAA or CTAA) of a peptide is modified or labeled prior to contacting the peptide with a binding agent in the methods described herein.

In some embodiments, the NTAA is reacted with phenylisothiocyanate (PITC) to generate a phenylthiocarbamoyl (PTC)-NTAA derivative. Edman degradation typically uses phenyl isothiocyanate (PITC) to label the N-terminus. PITC has two properties well suited for the methods disclosed herein: (1) PITC labels the N-terminus amine group with high efficiency; and (2) the resultant PTC derivitized NTAA undergoes self-isomerization, upon acid treatment, resulting in cleaving of the amino acid from the remaining peptide.

Other reagents that may be used to label the NTAA include: 4-sulfophenyl isothiocyanate, 3-pyridyl isothiocyante (PYITC), 2-piperidinoethyl isothiocyanate (PEITC), 3-(4-morpholino) propyl isothiocyanate (MPITC), 3-(diethylamino)propyl isothiocyanate (DEPTIC) (Wang et al., 2009, Anal Chem 81: 1893-1900), (1-fluoro-2,4-dinitrobenzene (Sanger's reagent, DNFB), dansyl chloride (DNS-Cl, or 1-dimethylaminonaphthalene-5-sulfonyl chloride), 4-sulfonyl-2-nitrofluorobenzene (SNFB), acetylation reagents, amidination (guanidination) reagents, 2-carboxy-4,6-dinitrochlorobenzene, 7-methoxycoumarin acetic acid, a thioacylation reagent, a thioacetylation reagent, and a thiobenzylation reagent. If the NTAA is blocked to labelling, there are a number of approaches to unblock the terminus, such as removing N-acetyl blocks with acyl peptide hydrolase (APH) (Farries, Harris et al., 1991, Eur. J. Biochem. 196:679-685). Methods of unblocking the N-terminus of a peptide are known in the art (*see, e.g.*, Krishna et al., 1991, Anal. Biochem. 199:45-50; Leone et al., 2011, Curr. Protoc. Protein Sci., Chapter 11:Unit11.7; Fowler et al., 2001, Curr. Protoc. Protein Sci., Chapter 11: Unit 11.7, each of which is hereby incorporated by reference in its entirety).

Dansyl chloride reacts with the free amine group of a peptide to yield a dansyl derivative of the NTAA. DNFB and SNFB react the α-amine groups of a peptide to produce DNP-NTAA, and SNP-NTAA, respectively. Additionally, both DNFB and SNFB also react with the with ε-amine of lysine residues. DNFB also reacts with tyrosine and histidine amino acid residues. SNFB has better selectivity for amine groups than DNFB, and is preferred for NTAA modification(Carty and Hirs 1968). In certain embodiments, lysine ε-amines are pre-blocked with an organic anhydride prior to polypeptide protease digestion into peptides.

Another useful NTAA modifier is an acetyl group since a known enzyme exists to remove acetylated NTAAs, namely acyl peptide hydrolases (APH) which cleaves the N-terminal acetylated amino acid, effectively shortening the peptide by a single amino acid {Chang, 2015 #373;Friedmann, 2013 #374}. The NTAA can be chemically acetylated with acetic anhydride or enzymatically acetylated with N-terminal acetyltransferases (NAT) {Chang, 2015 #373; Friedmann, 2013 #374}. Yet another useful NTAA modifier is an amidinyl (guanidinyl) moiety since a proven cleavage chemistry of the amidinated NTAA is known in the literature, namely mild incubation of the N-terminal amidinated peptide with 0.5-2% NaOH results in cleavage of the N-terminal amino acid {Hamada, 2016 #383}. This effectively provides a mild Edman-like chemical N-terminal degradation peptide sequencing process. Moroever, certain amidination (guanidination) reagents and the downstream NaOH cleavage are quite compatible with DNA encoding.

The presence of the DNP/SNP, acetyl, or amidinyl (guanidinyl) group on the NTAA may provide a better handle for interaction with an engineered binding agent. A number of commercial DNP antibodies exist with low nM affinities. Other methods of labeling the NTAA include labeling with trypligase (Liebscher et al., 2014, Angew Chem Int Ed Engl 53:3024-3028) and amino acyl transferase (Wagner, et al., 2011, J Am Chem Soc 133:15139-15147).

Isothiocyates, in the presence of ionic liquids, have been shown to have enhanced reactivity to primary amines. Ionic liquids are excellent solvents (and serve as a catalyst) in organic chemical reactions and can enhance the reaction of isothiocyanates with amines to form thioureas. An example is the use of the ionic liquid 1-butyl-3-methyl-imidazolium tetraflouoraborate [Bmim][BF4] for rapid and efficient labeling of aromatic and aliphatic amines by phenyl isothiocyanate (PITC) (Le, Chen et al. 2005). Edman degradation involves the reaction of isothiocyanates, such at PITC, with the amino N-terminus of peptides. As such, in one embodiment ionic liquids are used to improve the efficiency of the Edman degradation process by providing milder labeling and degradation conditions. For instance, the use of 5% (vol./vol.) PITC in ionic liquid [Bmim][BF4] at 25 °C for 10 min. is more efficient than labeling under standard Edman PITC derivatization conditions which employ 5% (vol./vol.) PITC in a solution containing pyridine, ethanol, and ddH2O (1:1:1 vol./vol./vol.) at 55 °C for 60 min (Wang, Fang et al. 2009). In a preferred embodiment, internal lysine, tyrosine, histidine, and cysteine amino acids are blocked within the polypeptide prior to fragmentation into peptides. In this way, only the peptide α-amine group of the NTAA is accessible for modification during the peptide sequencing reaction. This is particularly relevant when using DNFB (Sanger' reagent) and dansyl chloride.

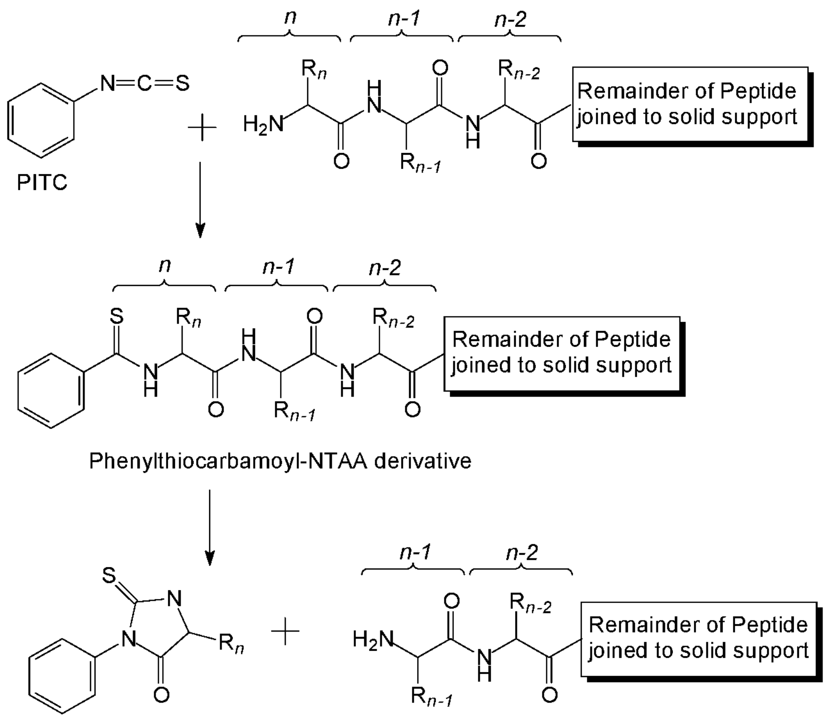
In certain embodiments, the NTAA have been blocked prior to the NTAA labelling step (particularly the original N-terminus of the protein). If so, there are a number of approaches to unblock the N-terminus, such as removing N-acetyl blocks with acyl peptide hydrolase (APH) (Farries, Harris et al. 1991). A number of other methods of unblocking the N-terminus of a peptide are known in the art (*see, e.g.*, Krishna et al., 1991, Anal. Biochem. 199:45-50; Leone et al., 2011, Curr. Protoc. Protein Sci., Chapter 11:Unit11.7; Fowler et al., 2001, Curr. Protoc. Protein Sci., Chapter 11: Unit 11.7, each of which is hereby incorporated by reference in its entirety).

The CTAA can be modified with a number of different carboxyl-reactive reagents as described by Hermanson (Hermanson 2013). In another example, the CTAA is modified with a mixed anhydride and an isothiocyanate to generate a thiohydantoin ((Liu and Liang 2001) and U.S. Patent No. 5,049,507). The thiohydantoin modified peptide can be cleaved at elevated temperature in base to expose the penultimate CTAA, effectively generating a C-terminal based peptide degradation sequencing approach (Liu and Liang 2001). Other modifications that can be made to the CTAA include addition of a para-nitroanilide group and addition of 7-amino-4-methylcoumarinyl group.

XI. Terminal Amino Acid Cleavage Methods

In certain embodiments relating to analyzing peptides, following binding of a terminal amino acid (N-terminal or C-terminal) by a binding agent and transfer of coding tag information to a recording tag, transfer of recording tag information to a coding tag, transfer of recording tag information and coding tag information to a di-tag construct, the terminal amino acid is removed or cleaved from the peptide to expose a new terminal amino acid. In some embodiments, the terminal amino acid is an NTAA. In other embodiments, the terminal amino acid is a CTAA.

Cleavage of a terminal amino acid can be accomplished by any number of known techniques, including chemical cleavage and enzymatic cleavage. An example of chemical cleavage is Edman degradation. During Edman degradation of the peptide the *n* NTAA is reacted with phenyl isothiocyanate (PITC) under mildly alkaline conditions to form the phenylthiocarbamoyl-NTAA derivative. Next, under acidic conditions, the phenylthiocarbamoyl-NTAA derivative is cleaved generating a free thiazolinone derivative, and thereby converting the *n-1* amino acid of the peptide to an N-terminal amino acid (*n-1* NTAA). The steps in this process are illustrated below:



Typical Edman Degradation, as described above requires deployment of harsh high temperature chemical conditions (*e.g.*, anhydrous TFA) for long incubation times. These conditions are generally not compatible with nucleic acid encoding of macromolecules.

To convert chemical Edman Degradation to a nucleic acid encoding-friendly approach, the harsh chemical steps are replaced with mild chemical degradation or efficient enzymatic steps. In one embodiment, chemical Edman degradation can be employed using milder conditions than original described. Several milder cleavage conditions for Edman degradation have been described in the literature, including replacing anhydrous TFA with triethylamine acetate in acetonitrile (see, e.g., Barrett, 1985, Tetrahedron Lett. 26:4375-4378, incorporated by reference in its entirety). Cleavage of the NTAA may also be accomplished using thioacylation degradation, which uses milder cleavage conditions as compared to Edman degradation (*see*, U.S. Patent 4,863,870).

In another embodiment, cleavage by anhydrous TFA may be replaced with an "Edmanase", an engineered enzyme that catalyzes the removal of the PITC-derivatized N-terminal amino acid via nucleophilic attack of the thiourea sulfur atom on the carbonyl group of the scissile peptide bond under mild conditions (*see*, U.S. Patent Publication US2014/0273004, incorporated by reference in its entirety). Edmanase was made by modifying *cruzain*, a cysteine protease from *Trypanosoma cruzi* (Borgo, 2014). A C25G mutation removes the catalytic cysteine residue while three mutations (G65S, A138C, L160Y) were selected to create steric fit with the phenyl moiety of the Edman reagent (PITC).

Enzymatic cleavage of a NTAA may also be accomplished by an aminopeptidase. Aminopeptidases naturally occur as monomeric and multimeric enzymes, and may be metal or ATP-dependent. Natural aminopeptidases have very limited specificity, and generically cleave N-terminal amino acids in a processive manner, cleaving one amino acid off after another. For the methods described here, aminopeptidases may be engineered to possess specific binding or catalytic activity to the NTAA only when modified with an N-terminal label. For example, an aminopeptidase may be engineered such than it only cleaves an N-terminal amino acid if it is modified by a group such as DNP/SNP, PTC, dansyl chloride, acetyl, amidinyl, etc. In this way, the aminopeptidase cleaves only a single amino acid at a time from the N-terminus, and allows control of the degradation cycle. In some embodiments, the modified aminopeptidase is non-selective as to amino acid residue identity while being selective for the N-terminal label. In other embodiments, the modified aminopeptidase is selective for both amino acid residue identity and the N-terminal label. An example of a model of modifying the specificity of enzymatic NTAA degradation is illustrated by Borgo and Havranek, where through structure-function aided design, a methionine aminopeptidase was converted into a leucine aminopeptidase (Borgo and Havranek 2014). A similar approach can be taken with a modified NTAA, such as DNP/SNP-modified NTAAs, wherein an aminopeptidase is engineered (using both structural-function based-design and directed evolution) to cleave only an N-terminal amino acid having a DNP/SNP group present. Engineered aminopeptidase mutants that bind to and cleave individual or small groups of labelled (biotinylated) NTAAs have been described (*see*, PCT Publication No. WO2010/065322).

In certain embodiments, a compact monomeric metalloenzymatic aminopeptidase is engineered to recognize and cleave DNP-labeled NTAAs. The use of a monomeric metallo-aminopeptidase has two key advantages: 1) compact monomeric proteins are much easier to display and screen using phage display; 2) a metallo-aminopeptidase has the unique advantage in that its activity can be turned on/off at will by adding or removing the appropriate metal cation. Exemplary aminopeptidases include the M28 family of aminopeptidases, such as Streptomyces sp. KK506 (SKAP) (Yoo, Ahn et al. 2010), Streptomyces griseus (SGAP), Vibrio proteolyticus (VPAP), (Spungin and Blumberg 1989, Ben-Meir, Spungin et al. 1993). These enzymes are stable, robust, and active at room temperature and pH 8.0, and thus compatible with mild conditions preferred for peptide analysis.

In another embodiment, cyclic cleavage is attained by engineering the aminopeptidase to be active only in the presence of the N-terminal amino acid label. Moreover, the aminopeptidase may be engineered to be non-specific, such that it does not selectively recognize one particular amino acid over another, but rather just recognizes the labeled N-terminus. In a preferred embodiment, a metallopeptidase monomeric aminopeptidase (e.g. *Vibro* leucine aminopeptidase) (Hernandez-Moreno, Villasenor et al. 2014), is engineered to cleave only modified NTAAs (e.g., PTC, DNP, SNP, acetylated, acylated, etc.)

In yet another embodiment, cyclic cleavage is attained by using an engineered acylpeptide hydrolase (APH) to cleave an acetylated NTAA. APH is a serine peptidase that is capable of catalyzing the removal of Nα-acetylated amino acids from blocked peptides, and is a key regulator of N-terminally acetylated proteins in eukaryal, bacterial and archaeal cells. In certain embodiments, the APH is a dimeric and has only exopeptidase activity (Gogliettino, Balestrieri et al. 2012, Gogliettino, Riccio et al. 2014). The engineered APH may have higher affinity and less selectivity than endogenous or wild type APHs.

In yet another embodiment, amidination (guanidinylation) of the NTAA is employed to enable mild cleavage of the labeled NTAA using NaOH (Hamada, 2016, incorporated by reference in its entirety). A number of amidination (guanidinylation) reagents are known in the art including: S-methylisothiurea, 3,5-dimethylpyrazole-1-carboxamidine, S-ethylthiouronium bromide, S-ethylthiouronium chloride, O-methylisourea, O-methylisouronium sulfate, O-methylisourea hydrogen sulfate, 2-methyl-1-nitroisourea, aminoiminomethanesulfonic acid, cyanamide, cyanoguanide, dicyandiamide, 3,5-dimethyl-1-guanylpyrazole nitrate and 3,5-dimethyl pyrazole, N,N′-bis(ortho-chloro-Cbz)-S- methylisothiourea and N,N′-bis(ortho-bromo-Cbz)-S-methylisothiourea (Katritzky, 2005, incorporated by reference in its entirety).

An example of a NTAA labeling, binding, and degradation workflow is as follows (see Figure 41 and 42): a large collection of recording tag labeled peptides (e.g., 50 million - 1 billion) from a proteolytic digest are immobilized randomly on a single molecule sequencing substrate (e.g., porous beads) at an appropriate intramolecular spacing. In a cyclic manner, the N-terminal amino acid (NTAA) of each peptide are modified with a small chemical moiety (e.g., DNP, SNP, acetyl) to provide cyclic control of the NTAA degradation process, and enhance binding affinity by a cognate binding agent. The modified N-terminal amino acid (e.g., DNP-NTAA, SNP-NTAA, acetyl-NTAA) of each immobilized peptide is bound by the cognate NTAA binding agent, and information from the coding tag associated with the bound NTAA binding agent is transferred to the recording tag associated with the immobilized peptide. After NTAA recognition, binding, and transfer of coding tag information to the recording tag, the labelled NTAA is removed by exposure to an engineered aminopeptidase (e.g., for DNP-NTAA or SNP-NTAA) or engineered APH (e.g., for acetyl-NTAA), that is capable of NTAA cleavage only in the presence of the label. Other NTAA labels (e.g., PITC) could also be employed with a suitably engineered aminopeptidase. In a particular embodiment, a single engineered aminopeptidase or APH universally cleaves all possible NTAAs (including post-translational modification variants) that possess the N-terminal amino acid label. In another particular embodiment, two, three, four, or more engineered aminopeptidases or APHs are used to cleave the repertoire of labeled NTAAs.

Aminopeptidases with activity to DNP or SNP labeled NTAAs may be selected using a screen combining tight-binding selection on the apo-enzyme (inactive in absence of metal cofactor) followed by a functional catalytic selection step, like the approach described by Ponsard et al. in engineering the metallo-beta-lactamase enzyme for benzylpenicillin (Ponsard, Galleni et al. 2001, Fernandez-Gacio, Uguen et al. 2003). This two-step selection is involves using a metallo-AP activated by addition of Zn2+ ions. After tight binding selection to an immobilized peptide substrate, Zn2+ is introduced, and catalytically active phage capable of hydrolyzing the NTAA labeled with DNP or SNP leads to release of the bound phage into the supernatant. Repeated selection rounds are performed to enrich for active APs for DNP or SNP labeled NTAA cleavage.

In any of the embodiments provided herein, recruitment of an NTAA cleavage reagent to the NTAA may be enhanced via a chimeric cleavage enzyme and chimeric NTAA modifier, wherein the chimeric cleavage enzyme and chimeric NTAA modifier each comprise a moiety capable of a tight binding reaction with each other (e.g., biotin-streptavidin) (see, Figure 39). For example, an NTAA may be modified with biotin-PITC, and a chimeric cleavage enzyme (streptavidin-Edmanase) is recruited to the modified NTAA via the streptavidin-biotin interaction, improving the affinity and efficiency of the cleavage enzyme. The modified NTAA is cleaved and diffuses away from the peptide along with the associated cleavage enzyme. In the example of a chimeric Edmanase, this approach effectively increases the affinity K_D from μM to sub-picomolar. A similar cleavage enhancement can also be realized via tethering using a DNA tag on the cleavage agent interacting with the recording tag (see Figure 44).

As an alternative to NTAA cleavage, a dipeptidyl amino peptidase (DAP) can be used to cleave the last two N-terminal amino acids from the peptide. In certain embodiments, a single NTAA can be cleaved (see Figure 45): Figure 45 depicts an approach to N-terminal degradation in which N-terminal ligation of a butelase I peptide substrate attaches a TEV endopeptidase substrate to the N-terminal of the peptide. After attachment, TEV endopeptidase cleaves the newly ligated peptide from the query peptide (peptide undergoing sequencing) leaving a single asparagine (N) attached to the NTAA. Incubation with DAP, which cleaves two amino acids from the N-terminus, results in a net removal of the original NTAA. This whole process can be cycled in the N-terminal degradation process.

For embodiments relating to CTAA binding agents, methods of cleaving CTAA from peptides are also known in the art. For example, U.S. Patent 6,046,053 discloses a method of reacting the peptide or protein with an alkyl acid anhydride to convert the carboxy-terminal into oxazolone, liberating the C-terminal amino acid by reaction with acid and alcohol or with ester. Enzymatic cleavage of a CTAA may also be accomplished by a carboxypeptidase. Several carboxypeptidases exhibit amino acid preferences, *e.g.*, carboxypeptidase B preferentially cleaves at basic amino acids, such as arginine and lysine. As described above, carboxypeptidases may also be modified in the same fashion as aminopeptidases to engineer carboxypeptidases that specifically bind to CTAAs having a C-terminal label. In this way, the carboxypeptidase cleaves only a single amino acid at a time from the C-terminus, and allows control of the degradation cycle. In some embodiments, the modified carboxypeptidase is non-selective as to amino acid residue identity while being selective for the C-terminal label.

In other embodiments, the modified carboxypeptidase is selective for both amino acid residue identity and the C-terminal label.

XII. Processing and Analysis of Extended Recording Tags, Extended Coding Tags, or Di-Tags

Extended recording tag, extended coding tag, and di-tag libraries representing the macromolecule(s) of interest can be processed and analysed using a variety of nucleic acid sequencing methods. Examples of sequencing methods include, but are not limited to, chain termination sequencing (Sanger sequencing); next generation sequencing methods, such as sequencing by synthesis, sequencing by ligation, sequencing by hybridization, polony sequencing, ion semiconductor sequencing, and pyrosequencing; and third generation sequencing methods, such as single molecule real time sequencing, nanopore-based sequencing, duplex interrupted sequencing, and direct imaging of DNA using advanced microscopy.

A library of extended recording tags, extended coding tags, or di-tags may be amplified in a variety of ways. A library of extended recording tags, extended coding tags, or di-tags may undergo exponential amplification, *e.g.*, via PCR or emulsion PCR. Emulsion PCR is known to produce more uniform amplification (Hori, Fukano et al. 2007). Alternatively, a library of extended recording tags, extended coding tags, or di-tags may undergo linear amplification, *e.g.*, via *in vitro* transcription of template DNA using T7 RNA polymerase. The library of extended recording tags, extended coding tags, or di-tags can be amplified using primers compatible with the universal forward priming site and universal reverse priming site contained therein. A library of extended recording tags, extended coding tags, or di-tags can also be amplified using tailed primers to add sequence to either the 5'-end, 3'-end or both ends of the extended recording tags, extended coding tags, or di-tags. Sequences that can be added to the termini of the extended recording tags, extended coding tags, or di-tags include library specific index sequences to allow multiplexing of multiple libraries in a single sequencing run, adaptor sequences, read primer sequences, or any other sequences for making the library of extended recording tags, extended coding tags, or di-tags compatible for a sequencing platform. An example of a library amplification in preparation for next generation sequencing is as follows: a 20 μl PCR reaction volume is set up using an extended recording tag library eluted from ~1 mg of beads (~ 10 ng), 200 uM dNTP, 1 μM of each forward and reverse amplification primers, 0.5 μl (1U) of Phusion Hot Start enzyme (New England Biolabs) and subjected to the following cycling conditions: 98° C for 30 sec followed by 20 cycles of 98° C for 10 sec, 60° C for 30 sec, 72° C for 30 sec, followed by 72° C for 7 min, then hold at 4° C.

In certain embodiments, either before, during or following amplification, the library of extended recording tags, extended coding tags, or di-tags can undergo target enrichment. Target enrichment can be used to selectively capture or amplify extended recording tags representing macromolecules of interest from a library of extended recording tags, extended coding tags, or di-tags before sequencing. Target enrichment for protein sequence is challenging because of the high cost and difficulty in producing highly-specific binding agents for target proteins. Antibodies are notoriously non-specific and difficult to scale production across thousands of proteins. The methods of the present disclosure circumvent this problem by converting the protein code into a nucleic acid code which can then make use of a wide range of targeted DNA enrichment strategies available for DNA libraries. Peptides of interest can be enriched in a sample by enriching their corresponding extended recording tags. Methods of targeted enrichment are known in the art, and include hybrid capture assays, PCR-based assays such as TruSeq custom Amplicon (Illumina), padlock probes (also referred to as molecular inversion probes), and the like (*see*, Mamanova et al., 2010, Nature Methods 7: 111-118; Bodi et al., J. Biomol. Tech. 2013, 24:73-86; Ballester et al., 2016, Expert Review of Molecular Diagnostics 357-372; Mertes et al., 2011, Brief Funct. Genomics 10:374-386; Nilsson et al., 1994, Science 265:2085-8; each of which are incorporated herein by reference in their entirety).

In one embodiment, a library of extended recording tags, extended coding tags, or di-tags is enriched via a hybrid capture-based assay (see, e.g., Figure 17A and Figure 17B). In a hybrid-capture based assay, the library of extended recording tags, extended coding tags, or di-tags is hybridized to target-specific oligonucleotides or "bait oligonucleotide" that are labelled with an affinity tag (*e.g.*, biotin). Extended recording tags, extended coding tags, or di-tags hybridized to the target-specific oligonucleotides are "pulled down" via their affinity tags using an affinity ligand (*e.g.*, streptavidin coated beads), and background (non-specific) extended recording tags are washed away (*see*, e.g., Figure 17). The enriched extended recording tags, extended coding tags, or di-tags are then obtained for positive enrichment (*e.g.*, eluted from the beads).

For bait oligonucleotides synthesized by array-based "in situ" oligonucleotide synthesis and subsequent amplification of oligonucleotide pools, competing baits can be engineered into the pool by employing several sets of universal primers within a given oligonucleotide array. For each type of universal primer, the ratio of biotinylated primer to non-biotinylated primer controls the enrichment ratio. The use of several primer types enables several enrichment ratios to be designed into the final oligonucleotide bait pool.

A bait oligonucleotide can be designed to be complementary to an extended recording tag, extended coding tag, or di-tag representing a macromolecule of interest. The degree of complementarity of a bait oligonucleotide to the spacer sequence in the extended recording tag, extended coding tag, or di-tag can be from 0% to 100%, and any integer in between. This parameter can be easily optimized by a few enrichment experiments. In some embodiments, the length of the spacer relative to the encoder sequence is minimized in the coding tag design or the spacers are designed such that they unavailable for hybridization to the bait sequences. One approach is to use spacers that form a secondary structure in the presence of a cofactor. An example of such a secondary structure is a G-quadruplex, which is a structure formed by two or more guanine quartets stacked on top of each other (Bochman, Paeschke et al. 2012). A guanine quartet is a square planar structure formed by four guanine bases that associate through Hoogsteen hydrogen bonding. The G-quadruplex structure is stabilized in the presence of a cation, *e.g.*, K+ ions vs. Li+ ions.

To minimize the number of bait oligonucleotides employed, a set of relatively unique peptides from each protein can be bioinformatically identified, and only those bait oligonucleotides complementary to the corresponding extended recording tag library representations of the peptides of interest are used in the hybrid capture assay. Sequential rounds or enrichment can also be carried out, with the same or different bait sets.

To enrich the entire length of a macromolecule (*e.g.*, protein or polypeptide) in a library of extended recording tags, extended coding tags, or di-tags representing fragments thereof (*e.g.*, peptides), "tiled" bait oligonucleotides can be designed across the entire nucleic acid representation of the protein.

In another embodiment, primer extension and ligation-based mediated amplification enrichment (AmpliSeq, PCR, TruSeq TSCA, etc.) can be used to select and module fraction enriched of library elements representing a subset of macromolecules. Competing oligos can also be employed to tune the degree of primer extension, ligation, or amplification. In the simplest implementation, this can be accomplished by having a mix of target specific primers comprising a universal primer tail and competing primers lacking a 5' universal primer tail. After an initial primer extension, only primers with the 5' universal primer sequence can be amplified. The ratio of primer with and without the universal primer sequence controls the fraction of target amplified. In other embodiments, the inclusion of hybridizing but non-extending primers can be used to modulate the fraction of library elements undergoing primer extension, ligation, or amplification.

Targeted enrichment methods can also be used in a negative selection mode to selectively remove extended recording tags, extended coding tags, or di-tags from a library before sequencing. Thus, in the example described above using biotinylated bait oligonucleotides and streptavidin coated beads, the supernatant is retained for sequencing while the bait-oligonucleotide:extended recording tag, extended coding tag, or di-tag hybrids bound to the beads are not analysed. Examples of undesirable extended recording tags, extended coding tags, or di-tags that can be removed are those representing over abundant macromolecule species, *e.g.*, for proteins, albumin, immunoglobulins, etc.

A competitor oligonucleotide bait, hybridizing to the target but lacking a biotin moiety, can also be used in the hybrid capture step to modulate the fraction of any particular locus enriched. The competitor oligonucleotide bait competes for hybridization to the target with the standard biotinylated bait effectively modulating the fraction of target pulled down during enrichment (Figure 17). The ten orders dynamic range of protein expression can be compressed by several orders using this competitive suppression approach, especially for the overly abundant species such as albumin. Thus, the fraction of library elements captured for a given locus relative to standard hybrid capture can be modulated from 100% down to 0% enrichment.

Additionally, library normalization techniques can be used to remove overly abundant species from the extended recording tag, extended coding tag, or di-tag library. This approach works best for defined length libraries originating from peptides generated by site-specific protease digestion such as trypsin, LysC, GluC, etc. In one example, normalization can be accomplished by denaturing a double-stranded library and allowing the library elements to re-anneal. The abundant library elements re-anneal more quickly than less abundant elements due to the second-order rate constant of bimolecular hybridization kinetics (Bochman, Paeschke et al. 2012). The ssDNA library elements can be separated from the abundant dsDNA library elements using methods known in the art, such as chromatography on hydroxyapatite columns (VanderNoot, et al., 2012, Biotechniques 53:373-380) or treatment of the library with a duplex-specific nuclease (DSN) from Kamchatka crab (Shagin et al., 2002, Genome Res. 12:1935-42) which destroys the dsDNA library elements.

Any combination of fractionation, enrichment, and subtraction methods, of the macromolecules before attachment to the solid support and/or of the resulting extended recording tag library can economize sequencing reads and improve measurement of low abundance species.

In some embodiments, a library of extended recording tags, extended coding tags, or di-tags is concatenated by ligation or end-complementary PCR to create a long DNA molecule comprising multiple different extended recorder tags, extended coding tags, or di-tags, respectively (Du et al., 2003, BioTechniques 35:66-72; Muecke et al., 2008, Structure 16:837-841; U.S. Patent No. 5,834,252, each of which is incorporated by reference in its entirety). This embodiment is preferable for nanopore sequencing in which long strands of DNA are analyzed by the nanopore sequencing device.

In some embodiments, direct single molecule analysis is performed on an extended recording tag, extended coding tag, or di-tag (*see, e.g.*, Harris et al., 2008, Science 320:106-109). The extended recording tags, extended coding tags, or di-tags can be analysed directly on the solid support, such as a flow cell or beads that are compatible for loading onto a flow cell surface (optionally microcell patterned), wherein the flow cell or beads can integrate with a single molecule sequencer or a single molecule decoding instrument. For single molecule decoding, hybridization of several rounds of pooled fluorescently-labelled of decoding oligonucleotides (Gunderson et al., 2004, Genome Res. 14:970-7) can be used to ascertain both the identity and order of the coding tags within the extended recording tag. To deconvolute the binding order of the coding tags, the binding agents may be labelled with cycle-specific coding tags as described above (see also, Gunderson et al., 2004, Genome Res. 14:970-7). Cycle-specific coding tags will work for both a single, concatenated extended recording tag representing a single macromolecule, or for a collection of extended recording tags representing a single macromolecule.

Following sequencing of the extended reporter tag, extended coding tag, or di-tag libraries, the resulting sequences can be collapsed by their UMIs and then associated to their corresponding macromolecules (*e.g.*, peptides, proteins, protein complex) and aligned to the totality of the macromolecule type in the cell (*e.g.*, proteome for peptide, polypeptide, protein macromolecules). Resulting sequences can also be collapsed by their compartment tags and associated to their corresponding compartmental proteome, which in a particular embodiment contains only a single or a very limited number of protein molecules. Both protein identification and quantification can easily be derived from this digital peptide information.

In some embodiments, the coding tag sequence can be optimized for the particular sequencing analysis platform. In a particular embodiment, the sequencing platform is nanopore sequencing. In some embodiments, the sequencing platform has a per base error rate of > 5%, > 10%, >15%, > 20%, > 25%, or > 30%. For example, if the extended recording tag is to be analyzed using a nanopore sequencing instrument, the barcode sequences (e.g., encoder sequences) can be designed to be optimally electrically distinguishable in transit through a nanopore. Peptide sequencing according to the methods described herein may be well-suited for nanopore sequencing, given that the single base accuracy for nanopore sequencing is still rather low (75%-85%), but determination of the "encoder sequence" should be much more accurate (> 99%). Moreover, a technique called duplex interrupted nanopore sequencing (DI) can be employed with nanopore strand sequencing without the need for a molecular motor, greatly simplifying the system design (Derrington, Butler et al. 2010). Readout of the extended recording tag via DI nanopore sequencing requires that the spacer elements in the concatenated extended recording tag library be annealed with complementary oligonucleotides. The oligonucleotides used herein may comprise LNAs, or other modified nucleic acids or analogs to increase the effective Tm of the resultant duplexes. As the single-stranded extended recording tag decorated with these duplex spacer regions is passed through the pore, the double strand region will become transiently stalled at the constriction zone enabling a current readout of about three bases adjacent to the duplex region. In a particular embodiment for DI nanopore sequencing, the encoder sequence is designed in such a way that the three bases adjacent to the spacer element create maximally electrically distinguishable nanopore signals (Derrington et al., 2010, Proc. Natl. Acad. Sci. USA 107:16060-5). As an alternative to motor-free DI sequencing, the spacer element can be designed to adopt a secondary structure such as a G-quartet, which will transiently stall the extended recording tag, extended coding tag, or di-tag as it passes through the nanopore enabling readout of the adjacent encoder sequence (Shim, Tan et al. 2009, Zhang, Zhang et al. 2016). After proceeding past the stall, the next spacer will again create a transient stall, enabling readout of the next encoder sequence, and so forth.

The methods disclosed herein can be used for analysis, including detection, quantitation and/or sequencing, of a plurality of macromolecules (e.g., peptides) simultaneously (multiplexing). Multiplexing as used herein refers to analysis of a plurality of macromolecules in the same assay. The plurality of macromolecules can be derived from the same sample or different samples. The plurality of macromolecules can be derived from the same subject or different subjects. The plurality of macromolecules that are analyzed can be different macromolecules (e.g., peptides), or the same macromolecule (e.g., peptide) derived from different samples. A plurality of macromolecules includes 2 or more macromolecules, 5 or more macromolecules, 10 or more macromolecules, 50 or more macromolecules, 100 or more macromolecules, 500 or more macromolecules, 1000 or more macromolecules, 5,000 or more macromolecules, 10,000 or more macromolecules, 50,000 or more macromolecules, 100,000 or more macromolecules, 500,000 or more macromolecules, or 1,000,000 or more macromolecules.

Sample multiplexing can be achieved by upfront barcoding of recording tag labeled macromolecule samples. Each barcode represents a different sample, and samples can be pooled prior to cyclic binding assays or sequence analysis. In this way, many barcode-labeled samples can be simultaneously processed in a single tube. This approach is a significant improvement on immunoassays conducted on reverse phase protein arrays (RPPA) (Akbani, Becker et al. 2014, Creighton and Huang 2015, Nishizuka and Mills 2016). In this way, the present disclosure essentially provides a highly digital sample and analyte multiplexed alternative to the RPPA assay with a simple workflow.

XIII. Macromolecule Characterization via Cyclic Rounds of NTAA Recognition, Recording Tag Extension, and NTAA Cleavage

In certain embodiments, the methods for analyzing a macromolecule provided in the present disclosure comprise multiple binding cycles, where the macromolecule is contacted with a plurality of binding agents, and successive binding of binding agents transfers historical binding information in the form of a nucleic acid based coding tag to at least one recording tag associated with the macromolecule. In this way, a historical record containing information about multiple binding events is generated in a nucleic acid format.

In embodiments relating to methods of analyzing peptide macromolecules using an N-terminal degradation based approach (see, Figure 3, Figure 4, Figure 41, and Figure 42), following contacting and binding of a first binding agent to an *n* NTAA of a peptide of *n* amino acids and transfer of the first binding agent's coding tag information to a recording tag associated with the peptide, thereby generating a first order extended recording tag, the *n* NTAA is cleaved as described herein. Cleavage of the *n* NTAA converts the *n-1* amino acid of the peptide to an N-terminal amino acid, which is referred to herein as an *n-1* NTAA. As described herein, the *n* NTAA may optionally be labeled with a moiety (*e.g.*, PTC, DNP, SNP, acetyl, amidinyl, etc.), which is particularly useful in conjunction with cleavage enzymes that are engineered to bind to a labeled form of NTAA. If the *n* NTAA was labeled, the *n-1* NTAA is then labeled with the same moiety. A second binding agent is contacted with the peptide and binds to the *n-1* NTAA, and the second binding agent's coding tag information is transferred to the first order extended recording tag thereby generating a second order extended recording tag (*e.g.*, for generating a concatenated n^{th} order extended recording tag representing the peptide), or to a different recording tag (*e.g.*, for generating multiple extended recording tags, which collectively represent the peptide). Cleavage of the *n-1* NTAA converts the *n-2* amino acid of the peptide to an N-terminal amino acid, which is referred to herein as *n-2* NTAA. Additional binding, transfer, cleavage, and optionally NTAA labeling, can occur as described above up to *n* amino acids to generate an n^{th} order extended recording tag or *n* separate extended recording tags, which collectively represent the peptide. As used herein, an *n* "order" when used in reference to a binding agent, coding tag, or extended recording tag, refers to the *n* binding cycle, wherein the binding agent and its associated coding tag is used or the *n* binding cycle where the extended recording tag is created.

In some embodiments, contacting of the first binding agent and second binding agent to the macromolecule, and optionally any further binding agents (*e.g.*, third binding agent, fourth binding agent, fifth binding agent, and so on), are performed at the same time. For example, the first binding agent and second binding agent, and optionally any further order binding agents, can be pooled together, for example to form a library of binding agents. In another example, the first binding agent and second binding agent, and optionally any further order binding agents, rather than being pooled together, are added simultaneously to the macromolecule. In one embodiment, a library of binding agents comprises at least 20 binding agents that selectively bind to the 20 standard, naturally occurring amino acids.

In other embodiments, the first binding agent and second binding agent, and optionally any further order binding agents, are each contacted with the macromolecule in separate binding cycles, added in sequential order. In certain embodiments, the use of multiple binding agents at the same time is preferred, because the parallel approach saves time and because the binding agents are in competition, which reduces non-specific binding by non-cognate binding agents to a site that is bound by a cognate binding agent.

The length of the final extended recording tags generated by the methods described herein is dependent upon multiple factors, including the length of the coding tag (*e.g.*, encoder sequence and spacer), the length of the recording tag (*e.g.*, unique molecular identifier, spacer, universal priming site, bar code), the number of binding cycles performed, and whether coding tags from each binding cycle are transferred to the same extended recording tag or to multiple extended recording tags. In an example for a concatenated extended recording tag representing a peptide and produced by an Edman degradation like cleavage method, if the coding tag has an encoder sequence of 5 bases that is flanked on each side by a spacer of 5 bases, the coding tag information on the final extended recording tag, which represents the peptide's binding agent history, is 10 bases x number of Edman Degradation cycles. For a 20-cycle run, the extended recording is at least 200 bases (not including the initial recording tag sequence). This length is compatible with standard next generation sequencing instruments.

After the final binding cycle and transfer of the final binding agent's coding tag information to the extended recording tag, the recorder tag can be capped by addition of a universal reverse priming site via ligation, primer extension or other methods known in the art. In some embodiments, the universal forward priming site in the recording tag is compatible with the universal reverse priming site that is appended to the final extended recording tag. In some embodiments, a universal reverse priming site is an Illumina P7 primer (5'-CAAGCAGAAGACGGCATACGAGAT – 3' - SEQ ID NO:134) or an Illumina P5 primer (5'-AATGATACGGCGACCACCGA-3' – SEQ ID NO133). The sense or antisense P7 may be appended, depending on strand sense of the recording tag. An extended recording tag library can be cleaved or amplified directly from the solid support (*e.g.*, beads) and used in traditional next generation sequencing assays and protocols.

In some embodiments, a primer extension reaction is performed on a library of single stranded extended recording tags to copy complementary strands thereof.

The NGPS peptide sequencing assay comprises several chemical and enzymatic steps in a cyclical progression. The fact that NGPS sequencing is single molecule confers several key advantages to the process. The first key advantage of single molecule assay is the robustness to inefficiencies in the various cyclical chemical/enzymatic steps. This is enabled through the use of cycle-specific barcodes present in the coding tag sequence.

Using cycle-specific coding tags, we track information from each cycle. Since this is a single molecule sequencing approach, even 70% efficiency at each binding/transfer cycle in the sequencing process is more than sufficient to generate mappable sequence information. As an example, a ten-base peptide sequence "CPVQLWVDST" (SEQ ID NO:169) might be read as "CPXQXWXDXT" (SEQ ID NO:170) on our sequence platform (where X = any amino acid; the presence an amino acid is inferred by cycle number tracking). This partial amino acid sequence read is more than sufficient to uniquely map it back to the human p53 protein using BLASTP. As such, none of our processes have to be perfect to be robust. Moreover, when cycle-specific barcodes are combined with our partitioning concepts, absolute identification of the protein can be accomplished with only a few amino acids identified out of 10 positions since we know what set of peptides map to the original protein molecule (via compartment barcodes).

XIV. Protein normalization via fractionation, compartmentalization, and limited binding capacity resins.

One of the key challenges with proteomics analysis is addressing the large dynamic range in protein abundance within a sample. Proteins span greater than 10 orders of dynamic range within plasma (even "Top 20" depleted plasma). In certain embodiments, subtraction of certain protein species (e.g., highly abundant proteins) from the sample is performed prior to analysis. This can be accomplished, for example, using commercially available protein depletion reagents such as Sigma's PROT20 immuno-depletion kit, which deplete the top 20 plasma proteins. Additionally, it would be useful to have an approach that greatly reduced the dynamic range even further to a manageable 3-4 orders. In certain embodiments, a protein sample dynamic range can be modulated by fractionating the protein sample using standard fractionation methods, including electrophoresis and liquid chromatography (Zhou, Ning et al. 2012), or partitioning the fractions into compartments (e.g., droplets) loaded with limited capacity protein binding beads/resin (e.g. hydroxylated silica particles) (McCormick 1989) and eluting bound protein. Excess protein in each compartmentalized fraction is washed away.

Examples of electrophoretic methods include capillary electrophoresis (CE), capillary isoelectric focusing (CIEF), capillary isotachophoresis (CITP), free flow electrophoresis, gel-eluted liquid fraction entrapment electrophoresis (GELFrEE). Examples of liquid chromatography protein separation methods include reverse phase (RP), ion exchange (IE), size exclusion (SE), hydrophilic interaction, etc. Examples of compartment partitions include emulsions, droplets, microwells, physically separated regions on a flat substrate, etc. Exemplary protein binding beads/resins include silica nanoparticles derivitized with phenol groups or hydroxyl groups (e.g., StrataClean Resin from Agilent Technologies, RapidClean from LabTech, etc.). By limiting the binding capacity of the beads/resin, highly-abundant proteins eluting in a given fraction will only be partially bound to the beads, and excess proteins removed.

XV. Partitioning of Proteome of a Single Cell or Molecular Subsampling

In another aspect, the present disclosure provides methods for massively-parallel analysis of proteins in a sample using barcoding and partitioning techniques. Current approaches to protein analysis involve fragmentation of protein macromolecules into shorter peptide molecules suitable for peptide sequencing. Information obtained using such approaches is therefore limited by the fragmentation step and excludes, *e.g.*, long range continuity information of a protein, including post-translational modifications, protein-protein interactions occurring in each sample, the composition of a protein population present in a sample, or the origin of the protein macromolecule, such as from a particular cell or population of cells. Long range information of post-translation modifications within a protein molecule (e.g., proteoform characterization) provides a more complete picture of biology, and long range information on what peptides belong to what protein molecule provides a more robust mapping of peptide sequence to underlying protein sequence (see Figure 15A). This is especially relevant when the peptide sequencing technology only provides incomplete amino acid sequence information, such as information from only 5 amino acid types. By using the partitioning methods disclosed herein, combined with information from a number of peptides originating from the same protein molecule, the identity of the protein molecule (e.g. proteoform) can be more accurately assessed. Association of compartment tags with proteins and peptides derived from same compartment(s) facilitates reconstruction of molecular and cellular information. In typical proteome analysis, cells are lysed and proteins digested into short peptides, disrupting global information on which proteins derive from which cell or cell type, and which peptides derive from which protein or protein complex. This global information is important to understanding the biology and biochemistry within cells and tissues.

Partitioning refers to the random assignment of a unique barcode to a subpopulation of macromolecules from a population of macromolecules within a sample. Partitioning may be achieved by distributing macromolecules into compartments. A partition may be comprised of the macromolecules within a single compartment or the macromolecules within multiple compartments from a population of compartments.

A subset of macromolecules or a subset of a protein sample that has been separated into or on the same physical compartment or group of compartments from a plurality (e.g., millions to billions) of compartments are identified by a unique compartment tag. Thus, a compartment tag can be used to distinguish constituents derived from one or more compartments having the same compartment tag from those in another compartment (or group of compartments) having a different compartment tag, even after the constituents are pooled together.

The present disclosure provides methods of enhancing protein analysis by partitioning a complex proteome sample (e.g., a plurality of protein complexes, proteins, or polypeptides) or complex cellular sample into a plurality of compartments, wherein each compartment comprises a plurality of compartment tags that are the same within an individual compartment (save for an optional UMI sequence) and are different from the compartment tags of other compartments (see, Figure 18-20). The compartments optionally comprise a solid support (*e.g.*, bead) to which the plurality of compartment tags are joined thereto. The plurality of protein complexes, proteins, or polypeptides are fragmented into a plurality of peptides, which are then contacted to the plurality of compartment tags under conditions sufficient to permit annealing or joining of the plurality of peptides with the plurality of compartment tags within the plurality of compartments, thereby generating a plurality of compartment tagged peptides. Alternatively, the plurality of protein complexes, proteins, or polypeptides are joined to a plurality of compartment tags under conditions sufficient to permit annealing or joining of the plurality of protein complexes, proteins or polypeptides with the plurality of compartment tags within a plurality of compartments, thereby generating a plurality of compartment tagged protein complexes, proteins, polypeptides. The compartment tagged protein complexes, proteins, or polypeptides are then collected from the plurality of compartments and optionally fragmented into a plurality of compartment tagged peptides. One or more compartment tagged peptides are analyzed according to any of the methods described herein.

In certain embodiments, compartment tag information is transferred to a recording tag associated with a macromolecule (e.g., peptide) via primer extension (Figure 5) or ligation (Figure 6).

In some embodiments, the compartment tags are free in solution within the compartments. In other embodiments, the compartment tags are joined directly to the surface of the compartment (*e.g.*, well bottom of microtiter or picotiter plate) or a bead or bead within a compartment.

A compartment can be an aqueous compartment (*e.g.*, microfluidic droplet) or a solid compartment. A solid compartment includes, for example, a nanoparticle, a microsphere, a microtiter or picotiter well or a separated region on an array, a glass surface, a silicon surface, a plastic surface, a filter, a membrane, nylon, a silicon wafer chip, a flow cell, a flow through chip, a biochip including signal transducing electronics, an ELISA plate, a spinning interferometry disc, a nitrocellulose membrane, or a nitrocellulose-based polymer surface. In certain embodiments, each compartment contains, on average, a single cell.

A solid support can be any support surface including, but not limited to, a bead, a microbead, an array, a glass surface, a silicon surface, a plastic surface, a filter, a membrane, nylon, a silicon wafer chip, a flow cell, a flow through chip, a biochip including signal transducing electronics, a microtiter well, an ELISA plate, a spinning interferometry disc, a nitrocellulose membrane, a nitrocellulose-based polymer surface, a nanoparticle, or a microsphere. Materials for a solid support include but are not limited to acrylamide, agarose, cellulose, nitrocellulose, glass, gold, quartz, polystyrene, polyethylene vinyl acetate, polypropylene, polymethacrylate, polyethylene, polyethylene oxide, polysilicates, polycarbonates, Teflon, fluorocarbons, nylon, silicon rubber, polyanhydrides, polyglycolic acid, polyactic acid, polyorthoesters, functionalized silane, polypropylfumerate, collagen, glycosaminoglycans, polyamino acids, or any combination thereof. In certain embodiments, a solid support is a bead, for example, a polystyrene bead, a polymer bead, an agarose bead, an acrylamide bead, a solid core bead, a porous bead, a paramagnetic bead, glass bead, or a controlled pore bead.

Various methods of partitioning samples into compartments with compartment tagged beads is reviewed in Shembekar et al., (Shembekar, Chaipan et al. 2016). In one example, the proteome is partitioned into droplets via an emulsion to enable global information on protein molecules and protein complexes to be recorded using the methods disclosed herein (see, e.g., Figure 18 and Figure 19). In certain embodiments, the proteome is partitioned in compartments (*e.g.*, droplets) along with compartment tagged beads, an activate-able protease (directly or indirectly via heat, light, etc.), and a peptide ligase engineered to be protease-resistant (e.g., modified lysines, pegylation, etc.). In certain embodiments, the proteome can be treated with a denaturant to assess the peptide constituents of a protein or polypeptide. If information regarding the native state of a protein is desired, an interacting protein complex can be partitioned into compartments for subsequent analysis of the peptides derived therefrom.

A compartment tag comprises a barcode, which is optionally flanked by a spacer or universal primer sequence on one or both sides. The primer sequence can be complementary to the 3' sequence of a recording tag, thereby enabling transfer of compartment tag information to the recording tag via a primer extension reaction (*see*, Figures 22A-B). The barcode can be comprised of a single stranded nucleic acid molecule attached to a solid support or compartment or its complementary sequence hybridized to solid support or compartment, or both strands (see, e.g., Figure 16). A compartment tag can comprise a functional moiety, for example attached to the spacer, for coupling to a peptide. In one example, a functional moiety (*e.g.*, aldehyde) is one that is capable of reacting with the N-terminal amino acid residue on the plurality of peptides. In another example, the functional moiety is capable of reacting with an internal amino acid residue (e.g., lysine or lysine labeled with a "click" reactive moiety) on the plurality of peptides. In another embodiment, the functional moiety may simply be a complementary DNA sequence capable of hybridizing to a DNA tag-labeled protein. Alternatively, a compartment tag can be a chimeric molecule, further comprising a peptide comprising a recognition sequence for a protein ligase (*e.g.*, butelase I or homolog thereof) to allow ligation of the compartment tag to a peptide of interest (*see*, Figure 22A). A compartment tag can be a component within a larger nucleic acid molecule, which optionally further comprises a unique molecular identifier for providing identifying information on the peptide that is joined thereto, a spacer sequence, a universal priming site, or any combination thereof. This UMI sequence generally differs among a population of compartment tags within a compartment. In certain embodiments, a compartment tag is a component within a recording tag, such that the same tag that is used for providing individual compartment information is also used to record individual peptide information for the peptide attached thereto.

In certain embodiments, compartment tags can be formed by printing, spotting, ink-jetting the compartment tags into the compartment. In certain embodiments, a plurality of compartment tagged beads is formed, wherein one barcode type is present per bead, via split-and-pool oligonucleotide ligation or synthesis as described by Klein et al., 2015, Cell 161:1187-1201; Macosko et al., 2015, Cell 161:1202-1214; and Fan et al., 2015, Science 347:1258367. Compartment tagged beads can also be formed by individual synthesis or immobilization. In certain embodiments, the compartment tagged beads further comprise bifunctional recording tags, in which one portion comprises the compartment tag comprising a recording tag, and the other portion comprises a functional moiety to which the digested peptides can be coupled (Figure 19 and Figure 20).

In certain embodiments, the plurality of proteins or polypeptides within the plurality of compartments is fragmented into a plurality of peptides with a protease. A protease can be a metalloprotease. In certain embodiments, the activity of the metalloprotease is modulated by photo-activated release of metallic cations. Examples of endopeptidases that can be used include: trypsin, chymotrypsin, elastase, thermolysin, pepsin, clostripan, glutamyl endopeptidase (GluC), endopeptidase ArgC, peptidyl-asp metallo-endopeptidase (AspN), endopeptidase LysC and endopeptidase LysN. Their mode of activation varies depending on buffer and divalent cation requirements. Optionally, following sufficient digestion of the proteins or polypeptides into peptide fragments, the protease is inactivated (*e.g.*, heat, fluoro-oil or silicone oil soluble inhibitor, such as a divalent cation chelation agent).

In certain embodiments of peptide barcoding with compartment tags, a protein molecule (optionally, denatured polypeptide) is labeled with DNA tags by conjugation of the DNA tags to ε-amine moieties of the protein's lysine groups or indirectly via click chemistry attachment to a protein/polypeptide pre-labeled with a reactive click moiety such as alkyne (see Figure 2B and Figure 20A). The DNA tag-labeled polypeptides are then partitioned into compartments comprising compartment tags (*e.g.*, DNA barcodes bound to beads contained within droplets) (see Figure 20B), wherein a compartment tag contains a barcode that identifies each compartment. In one embodiment, a single protein/polypeptide molecule is co-encapsulated with a single species of DNA barcodes associated with a bead (see Figure 20B). In another embodiment, the compartment can constitute the surface of a bead with attached compartment (bead) tags similar to that described in PCT Publication WO2016/061517 (incorporated by reference in its entirety), except as applied to proteins rather than DNA. The compartment tag can comprise a barcode (BC) sequence, a universal priming site (U1'), a UMI sequence, and a spacer sequence (Sp). In one embodiment, concomitant with or after partitioning, the compartment tags are cleaved from the bead and hybridize to the DNA tags attached to the polypeptide, for example via the complementary U1 and U1' sequences on the DNA tag and compartment tag, respectively. For partitioning on beads, the DNA tag-labeled protein can be directly hybridized to the compartment tags on the bead surface (see, Figure 20C). After this hybridization step, the polypeptides with hybridized DNA tags are extracted from the compartments (*e.g.*, emulsion "cracked", or compartment tags cleaved from bead), and a polymerase-based primer extension step is used to write the barcode and UMI information to the DNA tags on the polypeptide to yield a compartment barcoded recording tag (see, Figure 20D). A LysC protease digestion may be used to cleave the polypeptide into constituent peptides labeled at their C-terminal lysine with a recording tag containing universal priming sequences, a compartment tag, and a UMI (see, Figure 20E). In one embodiment, the LysC protease is engineered to tolerate DNA-tagged lysine residues. The resultant recording tag labeled peptides are immobilized to a solid substrate (*e.g.*, bead) at an appropriate density to minimize intermolecular interactions between recording tagged peptides (see, Figures 20E and 20F).

Attachment of the peptide to the compartment tag (or vice versa) can be directly to an immobilized compartment tag, or to its complementary sequence (if double stranded). Alternatively, the compartment tag can be detached from the solid support or surface of the compartment, and the peptide and solution phase compartment tag joined within the compartment. In one embodiment, the functional moiety on the compartment tag (e.g., on the terminus of oligonucleotide) is an aldehyde which is coupled directly to the amine N-terminus of the peptide through a Schiff base (see Figure 16). In another embodiment, the compartment tag is constructed as a nucleic acid-peptide chimeric molecule comprising peptide motif (n-X…XXCGSHV-c) for a protein ligase. The nucleic acid-peptide compartment tag construct is conjugated to digested peptides using a peptide ligase, such as butelase I or a homolog thereof. Butelase I, and other asparaginyl endopeptidase (AEP) homologues, can be used to ligate the C-terminus of the oligonucleotide-peptide compartment tag construct to the N-terminus of the digested peptides (Nguyen, Wang et al. 2014, Nguyen, Cao et al. 2015). This reaction is fast and highly efficient. The resultant compartment tagged peptides can be subsequently immobilized to a solid support for nucleic-acid peptide analysis as described herein.

In certain embodiments, compartment tags that are joined to a solid support or surface of a compartment are released prior to joining the compartment tags with the plurality of fragmented peptides (*see* Figure 18). In some embodiments, following collection of the compartment tagged peptides from the plurality of compartments, the compartment tagged peptides are joined to a solid support in association with recording tags. Compartment tag information can then be transferred from the compartment tag on the compartment tagged peptide to the associated recording tag (*e.g.*, via a primer extension reaction primed from complementary spacer sequences within the recording tab and compartment tag). In some embodiments, the compartment tags are then removed from the compartment tagged peptides prior to peptide analysis according to the methods described herein. In further embodiments, the sequence specific protease (*e.g.*, Endo AspN) that is initially used to digest the plurality of proteins is also used to remove the compartment tag from the N terminus of the peptide after transfer of the compartment tag information to the associated recording tag (*see* Figure 22B).

Approaches for compartmental-based partitioning include droplet formation through microfluidic devices using T-junctions and flow focusing, emulsion generation using agitation or extrusion through a membrane with small holes (*e.g.*, track etch membrane), etc. (*see,* Figure 21). A challenge with compartmentalization is addressing the interior of the compartment. In certain embodiments, it may be difficult to conduct a series of different biochemical steps within a compartment since exchanging fluid components is challenging. As previously described, one can modify a limited feature of the droplet interior, such as pH, chelating agent, reducing agents, etc. by addition of the reagent to the fluoro-oil of the emulsion. However, the number of compounds that have solubility in both aqueous and organic phases is limited. One approach is to limit the reaction in the compartment to essentially the transfer of the barcode to the molecule of interest.

After labeling of the proteins/peptides with recording tags comprised of compartment tags (barcodes), the protein/peptides are immobilized on a solid-support at a suitable density to favor intramolecular transfer of information from the coding tag of a bound cognate binding agent to the corresponding recording tag/tags attached to the bound peptide or protein molecule. Intermolecular information transfer is minimized by controlling the intermolecular spacing of molecules on the surface of the solid-support.

In certain embodiments, the compartment tags need not be unique for each compartment in a population of compartments. A subset of compartments (two, three, four, or more) in a population of compartments may share the same compartment tag. For instance, each compartment may be comprised of a population of bead surfaces which act to capture a subpopulation of macromolecules from a sample (many molecules are captured per bead). Moreover, the beads comprise compartment barcodes which can be attached to the captured macromolecules. Each bead has only a single compartment barcode sequence, but this compartment barcode may be replicated on other beads with in the compartment (many beads mapping to the same barcode). There can be (although not required) a many-to-one mapping between physical compartments and compartment barcodes, moreover, there can be (although not required) a many-to-one mapping between macromolecules within a compartment. A partition barcode is defined as an assignment of a unique barcode to a subsampling of macromolecules from a population of macromolecules within a sample. This partition barcode may be comprised of identical compartment barcodes arising from the partitioning of macromolecules within compartments labeled with the same barcode. The use of physical compartments effectively subsamples the original sample to provide assignment of partition barcodes. For instance, a set of beads labeled with 10,000 different compartment barcodes is provided. Furthermore, suppose in a given assay, that a population of 1 million beads are used in the assay. On average, there are 100 beads per compartment barcode (Poisson distribution). Further suppose that the beads capture an aggregate of 10 million macromolecules. On average, there are 10 macromolecules per bead, with 100 compartments per compartment barcode, there are effectively 1000 macromolecules per partition barcode (comprised of 100 compartment barcodes for 100 distinct physical compartments).

In another embodiment, single molecule partitioning and partition barcoding of polypeptides is accomplished by labeling polypeptides (chemically or enzymatically) with an amplifiable DNA UMI tag (e.g., recording tag) at the N or C terminus, or both (see Figure 37). DNA tags are attached to the body of the polypeptide (internal amino acids) via non-specific photo-labeling or specific chemical attachment to reactive amino acids such as lysines as illustrated in Figure 2B. Information from the recording tag attached to the terminus of the peptide is transferred to the DNA tags via an enzymatic emulsion PCR (Williams, Peisajovich et al. 2006, Schutze, Rubelt et al. 2011) or emulsion in vitro transcription/reverse transcription (IVT/RT) step. In the preferred embodiment, a nanoemulsion is employed such that, on average, there is fewer than a single polypeptide per emulsion droplet with size from 50 nm-1000 nm (Nishikawa, Sunami et al. 2012, Gupta, Eral et al. 2016). Additionally, all the components of PCR are included in the aqueous emulsion mix including primers, dNTPs, Mg2+, polymerase, and PCR buffer. If IVT/RT is used, then the recording tag is designed with a T7/SP6 RNA polymerase promoter sequence to generate transcripts that hybridize to the DNA tags attached to the body of the polypeptide (Ryckelynck, Baudrey et al. 2015). A reverse transcriptase (RT) copies the information from the hybridized RNA molecule to the DNA tag. In this way, emulsion PCR or IVT/RT can be used to effectively transfer information from the terminus recording tag to multiple DNA tags attached to the body of the polypeptide.

Encapsulation of cellular contents via gelation in beads is a useful approach to single cell analysis (Tamminen and Virta 2015, Spencer, Tamminen et al. 2016). Barcoding single cell droplets enables all components from a single cell to be labeled with the same identifier (Klein, Mazutis et al. 2015, Gunderson, Steemers et al. 2016, Zilionis, Nainys et al. 2017). Compartment barcoding can be accomplished in a number of ways including direct incorporation of unique barcodes into each droplet by droplet joining (Raindance), by introduction of a barcoded beads into droplets (10X Genomics), or by combinatorial barcoding of components of the droplet post encapsulation and gelation using and split-pool combinatorial barcoding as described by Gunderson et al. (Gunderson, Steemers et al. 2016) and PCT Publication WO2016/130704, incorporated by reference in its entirety. A similar combinatorial labeling scheme can also be applied to nuclei as described by Adey et al. (Vitak, Torkenczy et al. 2017).

The above droplet barcoding approaches have been used for DNA analysis but not for protein analysis. Adapting the above droplet barcoding platforms to work with proteins requires several innovative steps. The first is that barcodes are primarily comprised of DNA sequences, and this DNA sequence information needs to be conferred to the protein analyte. In the case of a DNA analyte, it is relatively straightforward to transfer DNA information onto a DNA analyte. In contrast, transferring DNA information onto proteins is more challenging, particularly when the proteins are denatured and digested into peptides for downstream analysis. This requires that each peptide be labeled with a compartment barcode. The challenge is that once the cell is encapsulated into a droplet, it is difficult to denature the proteins, protease digest the resultant polypeptides, and simultaneously label the peptides with DNA barcodes. Encapsulation of cells in polymer forming droplets and their polymerization (gelation) into porous beads, which can be brought up into an aqueous buffer, provides a vehicle to perform multiple different reaction steps, unlike cells in droplets (Tamminen and Virta 2015, Spencer, Tamminen et al. 2016) (Gunderson, Steemers et al. 2016). Preferably, the encapsulated proteins are crosslinked to the gel matrix to prevent their subsequent diffusion from the gel beads. This gel bead format allows the entrapped proteins within the gel to be denatured chemically or enzymatically, labeled with DNA tags, protease digested, and subjected to a number of other interventions. Figure 38 depicts exemplary encapsulation and lysis of a single cell in a gel matrix.

XVI. Tissue and Single Cell Spatial Proteomics

Another use of barcodes is the spatial segmentation of a tissue on the surface an array of spatially distributed DNA barcode sequences. If tissue proteins are labelled with DNA recording tags comprising barcodes reflecting the spatial position of the protein within the cellular tissue mounted on the array surface, then the spatial distribution of protein analytes within the tissue slice can later be reconstructed after sequence analysis, much as is done for spatial transcriptomics as described by Stahl et al. (2016, Science 353(6294):78-82) and Crosetto et al. (Corsetto, Bienko et al., 2015). The attachment of spatial barcodes can be accomplished by releasing array-bound barcodes from the array and diffusing them into the tissue section, or alternatively, the proteins in the tissue section can be labeled with DNA recording tags, and then the proteins digested with a protease to release labeled peptides that can diffuse and hybridize to spatial barcodes on the array. The barcode information can then be transferred (enzymatically or chemically) to the recording tags attached to the peptides.

Spatial barcoding of the proteins within a tissue can be accomplished by placing a fixed/permeabilized tissue slice, chemically labelled with DNA recording tags, on a spatially encoded DNA array, wherein each feature on the array has a spatially identifiable barcode (*see*, Figure 23). To attach an array barcode to the DNA tag, the tissue slice can be digested with a protease, releasing DNA tag labelled peptides, which can diffuse and hybridize to proximal array features adjacent to the tissue slice. The array barcode information can be transferred to the DNA tag using chemical/enzymatic ligation or polymerase extension. Alternatively, rather than allowing the labelled peptides to diffuse to the array surface, the barcodes sequences on the array can be cleaved and allowed to diffuse into proximal areas on the tissue slice and hybridize to DNA tag-labelled proteins therein. Once again, the barcoding information can be transferred by chemical/enzymatic ligation or polymerase extension. In this second case, protease digestion can be performed following transfer of barcode information. The result of either approach is a collection of recording tag-labelled protein or peptides, wherein the recording tag comprises a barcode harbouring 2-D spatial information of the protein/peptides's location within the originating tissue. Moreover, the spatial distribution of post-translational modifications can be characterized. This approach provides a sensitive and highly-multiplexed *in situ* digital immunohistochemistry assay, and should form the basis of modern molecular pathology leading to much more accurate diagnosis and prognosis.

In another embodiment, spatial barcoding can be used within a cell to identify the protein constituents/PTMs within the cellular organelles and cellular compartments (Christoforou et al., 2016, Nat. Commun. 7:8992, incorporated by reference in its entirety). A number of approaches can be used to provide intracellular spatial barcodes, which can be attached to proximal proteins. In one embodiment, cells or tissue can be sub-cellular fractionated into constituent organelles, and the different protein organelle fractions barcoded. Other methods of spatial cellular labelling are described in the review by Marx, 2015, Nat Methods 12:815-819, incorporated by reference in its entirety; similar approaches can be used herein.

The following examples are provided for the purpose of illustration, and not limitation.

EXAMPLES

EXAMPLE 1: DIGESTION OF PROTEIN SAMPLE WITH PROTEINASE K

A library of peptides is prepared from a protein sample by digestion with a protease such as trypsin, Proteinase K, etc. Trypsin cleaves preferably at the C-terminal side of positively charged amino acids like lysine and arginine, whereas Proteinase K cleaves non-selectively across the protein. As such, Proteinase K digestions require careful titration using a preferred enzyme-to-polypeptide ratio to provide sufficient proteolysis to generate short peptides (~ 30 amino acids), but not over-digest the sample. In general, a titration of the functional activity needs to be performed for a given Proteinase K lot. In this example, a protein sample is digested with proteinase K, for 1 h at 37°C at a 1:10-1:100 (w/w) enzyme:protein ratio in 1X PBS/1 mM EDTA/0.5 mM $CaCl_2$/0.5% SDS (pH 8.0). After incubation, PMSF is added to a 5 mM final concentration to inhibit further digestion.

The specific activity of Proteinase K can be measured by incubating the "chemical substrate" benzoyl arginine -p-nitroanilide with Proteinase K and measuring the development of the yellow colored p-nitroaniline product that absorbs at ~ 410 nm. Enzyme activity is measured in units, where one unit equals 1 μmole of p-nitroanilide produced /min, and specific activity is measured in units of enzyme activity/mg total protein. The specific activity is then calculated by dividing the enzyme activity by the total amount of protein in the solution.

EXAMPLE 2: SAMPLE PREP USING SP3 ON BEAD PROTEASE DIGESTION AND LABELING

Proteins are extracted and denatured using an SP3 sample prep protocol as described by Hughes et al. (2014, Mol Syst Biol 10:757). After extraction, the protein mix (and beads) is solubilized in 50 mM borate buffer (pH 8.0) w/ 1 mM EDTA supplemented with 0.02% SDS at 37° C for 1 hr. After protein solubilization, disulfide bonds are reduced by adding DTT to a final concentration of 5 mM, and incubating the sample at 50 °C for 10 min. The cysteines are alkylated by addition of iodoacetamide to a final concentration of 10 mM and incubated in the dark at room temperature for 20 min. The reaction is diluted two-fold in 50 mM borate buffer, and Glu-C or Lys-C is added in a final proteinase:protein ratio of 1:50 (w/w). The sample is incubated at 37 °C o/n (~16 hrs.) to complete digestion. After sample digestion as described by Hughes et al. (*supra*), the peptides are bound to the beads by adding 100% acetonitrile to a final concentration of 95% acetonitrile and washed with acetonitrile in an 8 min. incubation. After washing, peptides are eluted off the beads in 10 µl of 2% DMSO by a 5 min. pipette mixing step.

EXAMPLE 3: COUPLING OF THE RECORDING TAG TO THE PEPTIDE

A DNA recording tag is coupled to a peptide in several ways (see, Aslam et al., 1998, Bioconjugation: Protein Coupling Techniques for the Biomedical Sciences, Macmillan Reference LTD; Hermanson GT, 1996, Bioconjugate Techniques, Academic Press Inc., 1996). In one approach, an oligonucleotide recording tag is constructed with a 5' amine that couples to the C-terminus of the peptide using carbdiimide chemistry, and an internal strained alkyne, DBCO-dT (Glen Research, VA), that couples to azide beads using click chemistry. The recording tag is coupled to the peptide in solution using large molar excess of recording tag to drive the carbodiimide coupling to completion, and limit peptide-peptide coupling. Alternatively, the oligonucleotide is constructed with a 5' strained alkyne (DBCO-dT), and is coupled to an azide-derivitized peptide (via azide-PEG-amine and carbodiimide coupling to C-terminus of peptide), and the coupled to aldehyde-reactive HyNic hydrazine beads. The recording tag oligonucleotide can easily be labeled with an internal aldehyde formylindole (Trilink) group for this purpose. Alternatively, rather than coupling to the C-terminal amine, the recording tags can instead be coupled to internal lysine residues (preferably after a Lys-C digest, or alternatively a Glu-C digest). In one approach, this can be accomplished by activating the lysine amine with an NHS-azide (or NHS-PEG-azide) group and then coupling to a 5' amine-labeled recording tag. In another approach, a 5' amine-labeled recording tag can be reacted with excess NHS homo-bifunctional cross-linking reagents, such as DSS, to create a 5' NHS activated recording tag. This 5' NHS activated recording tag can be directly coupled to the ε-amino group of the lysine residues of the peptide.

EXAMPLE 4: SITE-SPECIFIC LABELING OF AMINO ACIDS ON A PEPTIDE

Five different examples of amino acids on proteins or peptides that can be modified directly with activated DNA tags (using activation with heterobifunctional amino acid site-specific reagents) or indirectly via click chemistry heterobifunctional reagent that site-specifically labels amino acids with a click moiety that is later used to attach a cognate click moiety on the DNA tag (Lundblad 2014). A typical protein input comprises 1 μg protein in 50 μl appropriate aqueous buffer containing 0.1% RapiGest™ SF surfactant, and 5 mM TCEP. RapiGest™ SD is useful as an acid degradable surfactant for denaturing proteins into polypeptides for improving labeling or digestion. The following amino acid labeling strategies can be used: cysteines using maleimide chemistry --- 200 μM Sulfo-SMCC-activated DNA tags are used to site-specifically label cysteines in 100 mM MES buffer (pH 6.5) + 1% TX-100 for 1 hr.; lysines using NHS chemistry --- 200 μM DSS or BS^3-activated DNA tags are used to site-specifically label lysine on solution phase proteins or the bead-bound peptides in borate buffer (50 mM, pH 8.5) + 1% TX-100 for 1 hr. at room temp; tyrosine is modified with 4-Phenyl-3H-1,2,4-triazoline-3,5(4H)-diones (PTAD) or diazonium chemistry --- for diazonium chemistry, DNA Tags are activated with EDC and 4-carboxylbenzene diazonium tetrafluoroborate (Aikon International, China). The diazo linkage with tyrosine is created by incubating the protein or bead-bound peptides with 200 μM diazonium-derivitized DNA tags in borate buffer (50 mM, pH 8.5) + 1% TX-100 for 1 h on ice (Nguyen, Cao et al. 2015). Aspartate/glutamate is modified using EDC chemistry --- an amine-labeled DNA tag is incubated with the bead-bound peptides and 100 mM EDC/50 mM imidazole in pH 6.5 MES for 1 hr. at room temperature (Basle et al., 2010, Chem. Biol. 17:213-227). After labeling, excess activated DNA tags are removed using protein binding elution from C4 resin ZipTips (Millipore). The eluted proteins are brought up 50 μl 1X PBS buffer.

EXAMPLE 5: IMMOBILIZING STRAINED ALKYNE RECORDING TAG-LABELED PEPTIDES TO AZIDE-ACTIVATED BEADS

Azide-derivitized Dynabeads® M-270 beads are generated by reacting commercially-available amine Dynabeads® M-270 with an azide PEG NHS ester heterobifunctional linker (JenKem Technology, TX). Moreover, the surface density of azide can be titrated by mixing in methoxy or hydroxyl PEG NHS ester in the appropriate ratio. For a given peptide sample, 1-2 mg azide-derivitized Dynabeads® M-270 beads (~1.3 x 10^8 beads) is diluted in 100 µl borate buffer (50 mM sodium borate, pH 8.5), 1 ng recording tag-peptide is added, and incubated for 1 hr. at 23-37 °C. Wash 3X with 200 µl borate buffer.

EXAMPLE 6: CREATING FORMYLINDOLE REACTIVE HYNIC BEADS

HyNic derivitization of amine beads creates formylindole reactive beads. An aliquot of 20 mg Dynabeads® M-270 Amine beads (2.8 µm) beads are suspended in 200 ul borate buffer. After a brief sonication, 1-2 mg Sulfo-S-HyNic (succinimidyl 6-hydrazinonicotinate acetone hydrazone, SANH) (Catalog # S-1002, Solulink, San Diego) is added and the reaction mixture is shaken for 1 hr. at room temperature. The beads are then washed 2X with borate buffer, and 1X with citrate buffer (200 mM sodium citrate). The beads are suspended in a final concentration of 10 mg/ml in citrate buffer.

EXAMPLE 7: IMMOBILIZING RECORDING TAG FORMLINDOLE-LABELED PEPTIDES TO ACTIVATED BEADS

An aliquot of 1-2 mg HyNic activated Dynabeads® M-270 beads (~1.3 x 10^8 beads) are diluted in 100 µl citrate buffer supplemented with 50 mM aniline, ~1 ng recording tag peptide conjugate is added and incubated for 1 hr. at 37 °C. The beads are washed 3X with 200 µl citrate buffer, and re-suspended in 100 µl borate buffer.

EXAMPLE 8: OLIGONUCLEOTIDE MODEL SYSTEM - RECORDING OF BINDING AGENT HISTORY BY TRANSFER OF IDENTIFYING INFORMATION OF CODING TAG TO RECORDING TAG IN CYCLIC FASHION

For nucleic acid coding tags and recording tags, information can be transferred from the coding tag on the bound binding agent to the proximal recording tag by ligation or primer extension using standard nucleic acid enzymology. This can be demonstrated with a simple model system consisting of an oligonucleotide with the 5' portion representing the binding agent target, and the 3' portion representing the recording tag. The oligonucleotide can be immobilized at an internal site using click chemistry through a dT-alkyne modification (DBCO-dT, Glen Research). In the example shown in Figure 24A, the immobilized oligonucleotide (AB target) contains two target binding regions, labeled A and B, to which cognate oligonucleotide "binding agents" can bind, the A oligo and the B oligo. The A oligo and B oligonucleotides are linked to coding tags (differing in sequence and length) which interact with the recording tag through a common spacer (Sp) to initiate primer extension (or ligation). The length of Sp should be kept short (e.g., 6-9 bases) to minimize non-specific interaction during binding agent binding. In this particular example, the length of the coding tag is designed to easily distinguish by gel analysis an "A" oligo binding event (10 base encoder sequence) from a "B" oligo binding event (20 base encoder sequence).

Simple analysis on a PAGE gel enables measurement of the efficiency of A or B coding tag transfer, and allows easy optimization of experimental parameters. In addition to the AB target sequence, a similar oligonucleotide CD target sequence is employed (see, Figure 24B), except C and D are different hybridization sequences non-interacting with A and B. Furthermore, C and D contain coding tags of differing sequences and lengths, comprising a 30 base DNA code and 40 base DNA code, respectively. The purpose of the second target sequence, CD, is to assess cross interaction between the AB and CD target molecules. Given specific hybridization, the extended recording tag for the CD target should not contain A or B coding tag information unless intermolecular crossing occurs between the A or B coding tags connected to oligos bound to the AB target. Likewise, the extended recording tag for the AB target should contain no C or D coding tag information. In the situation where the AB and CD targets are in close physical proximity (*i.e.*, < 50 nm), there is likely to be cross talk. Therefore, it is important to appropriately space out the target macromolecules on the surface.

This oligonucleotide model system enables a full characterization of the recording capability of binding agent history. Figure 25 illustrates information transfer via ligation rather than primer extension. After initial optimization on gels, various binding and assay protocols are performed and assessed by sequencing. A unique molecular identifier (UMI) sequence is used for counting purposes, and enables identification of reads originating from a single macromolecule and provides a measure of overall total macromolecule complexity in the original sample. Exemplary historical binding protocols include: A-B-C-B-A, A-B-A-A-B-A, A-B-C-D-A-C, etc. The resultant final products should read: UMI-Sp-A-Sp-B-Sp-B-Sp-A-Sp + UMI-Sp-C-Sp; UMI-Sp-A-Sp-B-Sp-A-Sp-A-Sp-B-Sp-A; UMI-A-Sp-B-Sp-A + UMI-Sp-C-Sp-D-Sp-C-Sp, respectively. The results of this analysis allow further optimization.

EXAMPLE 9: OLIGONUCLEOTIDE-PEPTIDE MODEL SYSTEM - RECORDING OF BINDING AGENT HISTORY BY TRANSFER OF IDENTIFYING INFORMATION OF CODING TAG TO RECORDING TAG IN CYCLIC FASHION

After validating the oligonucleotide model system, a peptide model system is constructed from the oligonucleotide system by conjugating a peptide epitope tag to the 5' end of the exemplary target oligonucleotide sequence (Figures 26A and 26B). Exemplary peptide epitope tags include: FLAG (DYKDDDDK) (SEQ ID NO:171), V5 (GKPIPNPLLGLDST) (SEQ ID NO:172), c-Myc (EQKLISEEDL) (SEQ ID NO:173), HA (YPYDVPDYA) (SEQ ID NO:174), V5 (GKPIPNPLLGLDST) (SEQ ID NO:175), StrepTag II (NWSHPQFEK) (SEQ ID NO:176), etc. An optional Cys-Ser-Gly linker can be included for coupling of the peptide epitope tag to the oligonucleotide. The AB oligonucleotide template of Example 7 is replaced with an A_oligonucleotide-cMyc peptide construct, and the CD oligonucleotide template of Example 7 is replaced with an C_oligonucleotide-HA peptide construct (*see,* Figure

26). The A_oligonucleotide-cMyc peptide construct also contains a CSG linker and N-terminal phosphotyrosine. Likewise, the cognate peptide binding agents, cMyc antibody and HA antibody, are tagged with the B oligonucleotide coding tag, and D oligonucleotide coding tag, respectively. The phosphotyrosine specific antibody is tagged with a separate "E" coding tag. In this way, the peptide model system parallels the oligonucleotide system, and both oligo binding and antibody binding are tested in this model system.

Antibody staining of the immobilized DNA-peptide construct using anti-c-myc antibody (2G8D5, mouse monoclonal, GenScript), anti-HA antibody (5E11D8, mouse monoclonal, GenScript), strep-tag II antibody (5A9F9, mouse monoclonal, GenScript), or anti-FLAG antibody (5AE85, mouse monoclonal, GenScript) is performed using 0.1 – 1 μg/ml in 1X PBST (PBS + 0.1% Tween 20). Incubations are typically done at room temperature for 30 min. Standard pre-blocking using 1% PVP in 1X PBST, and post-stain washing are also performed. Antibody de-staining is effectively accomplished by washing with a high salt (1 M NaCl), and either low pH (glycine, pH 2.5) or high pH (triethylamine, pH 11.5).

The target oligonucleotide contains an internal alkyne label for attachment to azide beads, and the 5' terminus contains an amino group for an SMCC-mediated attachment to a C-terminal cysteine of the peptide as described by Williams et al. (2010, Curr Protoc Nucleic Acid Chem. Chapter 4:Unit 4.41). Alternatively, standard carbodiimide coupling is used for a conjugation reaction of the oligonucleotide and peptide (Lu et al., 2010, Bioconjug. Chem. 21:187-202). In this case, an excess of oligo is used to drive the carbodiimide reaction and minimized peptide-peptide coupling. After conjugation, the final product is purified by excision and elution from a PAGE gel.

EXAMPLE 10: CODING TAG TRANSFER VIA LIGATION OF DNA/PNA CODING TAG COMPLEMENT TO RECORDING TAG

A coding tag is transferred either directly or indirectly by ligation to the recording tag to generate an extended recording tag. In one implementation, an annealed complement of the coding tag is ligated to the recording tag (Figure 25). This coding tag complement can either be a nucleic acid (DNA or RNA), peptide nucleic acid (PNA), or some other coding molecule capable of being ligated to a growing recording tag. The ligation can be enzymatic in the case of DNA and RNA using standard ATP-dependent and NADH-dependent ligases, or ligation can be chemical-mediated for both DNA/RNA and especially the peptide nucleic acid, PNA.

For enzymatic ligation of DNA, the annealed coding tag requires a 5' phosphate to ligate to the 3' hydroxyl of the recording tag. Exemplary enzymatic ligation conditions are as follows (Gunderson, Huang et al. 1998): The standard T4 DNA ligation reaction includes: 50 mM Tris- HCl (pH 7.8), 10 mM MgCl2, 10 mM DTT, 1 mM ATP, 50 μg/ml BSA, 100 mM NaCl, 0.1% TX-100 and 2.0 U/μl T4 DNA ligase (New England Biolabs). *E. coli* DNA ligase reaction includes 40 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 5 mM DTT, 0.5 mM NADH, 50 μg/ml BSA, 0.1% TX-100, and 0.025 U/μl *E. coli* DNA ligase (Amersham). Taq DNA ligation reaction includes 20 mM Tris-HCl (pH 7.6), 25 mM potassium acetate, 10 mM magnesium acetate, 10 mM DTT, 1 mM NADH, 50 μg/ml BSA, 0.1% Triton X-100, 10% PEG, 100 mM NaCl, and 1.0 U/μl Taq DNA ligase (New England Biolabs). T4 and *E. coli* DNA ligase reactions are performed at room temperature for 1 hr., and Taq DNA ligase reactions are performed at 40°C for 1 hr.

Several methods of chemical ligation of templated of DNA/PNA can be employed for DNA/PNA coding tag transfer. These include standard chemical ligation and click chemistry approaches. Exemplary chemical ligation conditions for template DNA ligation is as follows (Gunderson, Huang et al. 1998): ligation of a template 3' phosphate reporter tag to a 5' phosphate coding tag takes place within 1 hr. at room temperature in a reaction consisting of 50 mM 2-[N- morpholino]ethanesulfonic acid (MES) (pH 6.0 with KOH), 10 mM $MgCl_2$, 0.001% SDS, freshly prepared 200 mM EDC, 50 mM imidazole (pH 6.0 with HCl) or 50 mM HOBt (pH 6.0 with HCl) and 3.0–4.0 M TMACl (Sigma).

Exemplary conditions for template-dependent ligation of PNA include ligation of NH_2-PNA-CHO polymers (e.g., coding tag complement and extended recorder tag) and are described by Brudno et al. (Brudno, Birnbaum et al. 2010). PNA has a 5' amine equivalent and a 3' aldehyde equivalent wherein chemical ligation couples the two moieties to create a Schiff base which is subsequently reduced with sodium cyanoborohydride. The typical reaction conditions for this coupling are: 100 mM TAPS (pH 8.5), 80 mM NaCl, and 80 mM sodium cyanoborohydride at room temperature for 60 min. Exemplary conditions for native chemical ligation using functionalized PNAs containing 5' amino terminal 1,2-aminothiol modifications and 3' C-terminal thioester modifications is described by Roloff et al. (2014, Methods Mol. Biol. 1050:131-141). Other N- and C- terminal PNA moieties can also be used for ligation. Another example involves the chemical ligation of PNAs using click chemistry. Using the approach of Peng et al. (2010, European J. Org. Chem. 2010: 4194-4197), PNAs can be derivitized with 5' azide and 3' alkyne and ligated using click chemistry. An exemplary reaction condition for the "click" chemical ligation is: 1-2 mg beads with templated PNA-PNA in 100 µl of reaction mix containing 10 mM potassium phosphate buffer, 100 mM KCl, 5 mM THPTA (tris-hydroxypropyl trizolyl amine), 0.5 mM $CuSO_4$, and 2.5 mM Na-ascorbate. The chemical ligation reaction is incubated at room temperature for 1 hr. Other exemplary methods of PNA ligation are described by Sakurai et al. (Sakurai, Snyder et al. 2005).

EXAMPLE 11: PNA TRANSLATION TO DNA

PNA is translated into DNA using click chemistry-mediated polymerization of DNA oligonucleotides annealed onto the PNA template. The DNA oligos contain a reactive 5' azide and 3' alkyne to create an inter-nucleotide triazole linkage capable of being replicated by DNA polymerases (El-Sagheer et al., 2011, Proc. Natl. Acad. Sci. USA 108:11338-11343). A complete set of DNA oligos (10 nM, in 1X hybridization buffer: 10 mM Na-borate (pH 8.5), 0.2 M NaCl) complementary to all possible coding tags in the PNA is incubated (23-50 °C) for 30 minutes with the solid-phase bound PNA molecules. After annealing, the solid-phase bound PNA-DNA constructs are washed 1X with sodium ascorbate buffer (10 mM sodium ascorbate, 200 mM NaCl). The 'click chemistry' reaction conditions are as follows: PNA-DNA on beads are incubated in fresh sodium ascorbate buffer and combined 1:1 with a mix of 10 mM THPTA + 2 mM $CuSO_4$ and incubated for 1 hr. at room temperature. The beads are then washed 1X with hybridization buffer and 2X with PCR buffer. After chemical ligation, the resultant ligated DNA product is amplified by PCR under conditions as described by El-Sagheer et al. (2011, Proc. Natl. Acad. Sci. USA 108:11338-11343).

EXAMPLE 12: MILD N-TERMINAL EDMAN DEGRADATION COMPATIBLE WITH NUCLEIC ACID RECORDING AND CODING TAGS

Compatibility between N-terminal Edman degradation and DNA encoding allows this approach to work for peptide sequencing. The standard conditions for N-terminal Edman degradation, employing anhydrous TFA, destroys DNA. However, this effect is mitigated by developing milder cleavage conditions and developing modified DNA with greater acid resistance. Milder conditions for N-terminal Edman degradation are developed using a combination of cleavage optimization of phenylthiocarbamoyl (PTC)-peptides and measured stability of DNA/PNA encoded libraries under the cleavage conditions. Moreover, native DNA can be stabilized against acid hydrolysis, by using base modifications, such as 7-deaza purines which reduce depurination at low pH, and 5' methyl modified cytosine which reduces depyrimidation (Schneider and Chait, 1995, Nucleic Acids Res. 23:1570-1575). T-rich coding tags may also be useful given that thymine is the most stable base to acid fragmentation. The conditions for mild N-terminal Edman degradation replace anhydrous TFA cleavage with a mild 10 min. base cleavage using triethylamine acetate in acetonitrile at 60°C as described by Barrett et al. (1985, Tetrahedron Lett. 26:4375-4378, incorporated by reference in its entirety). These mild conditions are compatible with most types of DNA reporting and coding tags. As an alternative, PNAs are used in coding tags since they are completely acid-stable (Ray and Norden, 2000, FASEB J. 14:1041-1060).

The compatibility of using DNA coding tags/recording tags to encode the identity of NTAA binders and perform mild N-terminal Edman degradation reaction is demonstrated using the following assay. Both anti-phosphotyrosine and anti-cMyc antibodies are used to read out the model peptide. C-Myc and N-terminal phosphotyrosine detection, coding tag writing, and removal of the N-terminal phosphotyrosine using a single Edman degradation step. After this step, the peptide is stained again with anti-phosphotyrosine and anti-cMyc antibodies. Stability of the recording tag to N-terminal degradation is assessed by qPCR. Effective removal of the phosphotyrosine is indicated by absence of the E-oligonucleotide coding tag information in the final recording tag sequence as analyzed by sequencing, qPCR, or gel electrophoresis.

EXAMPLE 13: PREPARATION OF COMPARTMENT TAGGED BEADS.

For preparation of compartment tagged beads, barcodes are incorporated into oligonucleotides immobilized on beads using a split-and-pool synthesis approach, using either phosphoramidite synthesis or through split-and-pool ligation. A compartment tag can further comprise a unique molecular identifier (UMI) to uniquely label each peptide or protein molecule to which the compartment tag is joined. An exemplary compartment tag sequence is as follows: 5'-NH_2-GCGCAATCAG-XXXXXXXXXXXX-NNNNN-TGCAAGGAT-3' (SEQ ID NO:177). The XXXXXXXXXXXX (SEQ ID NO:178) barcode sequence is a fixed population of nucleobase sequences per bead generated by split-pool on bead synthesis, wherein the fixed sequence differs from bead to bead . The NNNNN (SEQ ID NO:179) sequence is randomized within a bead to serve as a unique molecule identifier (UMI) for the peptide molecule that is subsequently joined thereto. The barcode sequence can be synthesized on beads using a split-and-pool approach as described by Macosko et al. (2015, Cell 161:1202-1214, incorporated by reference in its entirety). The UMI sequences can be created by synthesizing an oligonucleotide using a degenerate base mixture (mixture of all four phosphoramidite bases present at each coupling step). The 5'-NH_2 is activated with succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) and a cysteine containing butelase I peptide substrate with the sequence from N-terminus to C-terminus "CGGSSGSNHV"(SEQ ID NO:180) is coupled to the SMCC activated compartment tagged beads using a modified protocol described by Williams et al. (2010, Curr Protoc Nucleic Acid Chem. Chapter 4:Unit 4.41). Namely, 200 µl of magnetic beads (10 mg/ml) are placed in a 1.5 ml Eppendorf tube. 1 ml of coupling buffer (100 mM KH_2PO_4 buffer, pH 7.2 with 5 mM EDTA, 0.01% Tween 20, pH 7.4) is added to the tube and vortexed briefly. Freshly prepared 40 µl Sulfo-SMCC (50 mg/ml in DMSO, ThermoFisher) is added to the magnetic beads and mixed. The reaction is incubated for 1 hr. at room temperature on a rotary mixer. After incubation, the beads are separated from the supernatant on a magnet, and washed 3X with 500 µl coupling buffer. The beads are re-suspended in 400 µl coupling buffer. 1 mL of CGGSSGSNHV (SEQ ID NO:180) peptide is added (1 mg/mL in coupling buffer after TCEP-reduction (5 mM) and ice cold acetone precipitation) to the magnetic beads. The reaction is incubated at room temperature for 2 hours on a rotary mixer. The reaction is washed 1X with coupling buffer. 400 µl quenching buffer (100 mM KH_2PO_4 buffer, pH 7.2 with 10 mg/mL Mercaptosuccinic Acid, pH 7.4) is added to the reaction mixture and incubated for 2 hrs. on a rotary mixer. The reaction mixture is washed 3X with coupling buffer. The resultant beads are re-suspended in storage buffer (10 mM KH_2PO_4 buffer, pH 7.2 with 0.02% NaN_3, 0.01% Tween 20, pH 7.4) and stored at 4 °C.

EXAMPLE 14: GENERATION OF ENCAPSULATED BEADS AND PROTEINS

Compartment tagged beads and proteins are combined with a zinc metallo-endopeptidase, such as endoproteinase AspN (Endo AspN), an optional photo-caged Zn chelator (*e.g.*, ZincCleav I), and an engineered thermos-tolerant butelase I homolog (Bandara, Kennedy et al. 2009, Bandara, Walsh et al. 2011, Cao, Nguyen et al. 2015). Compartment tagged beads from Example 12 are mixed with proteins and emulsified through a T-junction microfluidic or flow focusing device (see Figure 21). In a two-aqueous flow configuration, the protein and Zn^{2+} in one flow can be combined with the metallo-endopeptidase from the other flow to initiate digestion immediately upon droplet formation. In the one flow configuration, all reagents are premixed and emulsified together. This requires use of the optional photo-caged Zn chelator (e.g., ZincCleav I) to initiate protein digestion post droplet formation via exposure to UV light. The concentrations and flow conditions are adjusted such that, on average, there is less than one bead per droplet. In an optimized experiment, 10^8 femto-droplets can be made with an occupancy of about 10% of the droplets containing beads (Shim et al., 2013, ACS Nano 7:5955-5964). In the one flow approach, after forming droplets, the protease is activated by exposing the emulsion to UV-365 nm light to release the photo-caged Zn^{2+}, activating the Endo AspN protease. The emulsion is incubated for 1 hr. at 37 °C to digest the proteins into peptides. After digestion, the Endo AspN is inactivated by heating the emulsion to 80°C for 15 min. In the two-flow formulation, the Zn^{2+} is introduced during the combining of the two flows into a droplet. In this case, the Endo AspN can be inactivated by using a photo-activated Zn^{2+} caging molecule in which the chelator is activated upon exposure to UV light, or by adding an amphipathic Zn^{2+} chelating agent to the oil phase, such as 2-alkylmalonic acid, or EDTA-MO. Examples of amphipathic EDTA molecules include: EDTA-MO, EDTA-BO, EDTA-BP, DPTA-MO, DPTA-BO, DPTA-BP, etc. (Ojha, Singh et al. 2010, Moghaddam, de Campo et al. 2012). Other modalities can also be used to control the reaction within the droplet interior including changing the pH of the droplet through addition of amphipathic acids or bases to the emulsion oil. For example, droplet pH can be lowered using water/oil soluble acetic acid. Addition of acetic acid to a fluoro-emulsion leads to reduction of pH within the droplet compartment due to the amphipathic nature of the acetic acid molecule (Mashaghi and van Oijen, 2015, Sci Rep 5:11837). Likewise, addition of the base, propyl amine, alkalinizes the droplet interior. Similar approaches can be used for other types of amphipathic molecules such as oil/water soluble redox reagents, reducing agents, chelating agents and catalysts.

After digestion of the compartmentalized proteins into peptides, the peptides are ligated to the compartment tags (oligonucleotide peptide barcode chimeras) on the bead using butelase I or a chemical ligation (*e.g.*, aldehyde-amino, etc.) (*see*, Figure 16 and Figure 22A). In an optional approach, an oligo-thiodepsipeptide “chemical substrate” is employed to make the butelase I ligation irreversible (Nguyen, Cao et al. 2015). After ligation, the emulsion is “cracked”, and the beads with immobilized compartment tagged peptide constructs collected in bulk, or the compartment tagged peptides are cleaved from the beads, and collected in bulk. If the bead immobilized compartment tagged peptides comprise a recording tag, these beads can be used directly in nucleic acid encoding based peptide analysis methods described herein. In contrast, if the compartment tagged peptides are cleaved from the bead substrate, the compartment tagged peptides are then associated with a recording tag by conjugation to the C-terminus of the compartment tagged peptide, and immobilized on a solid support for subsequent binding cycles with coding tagged binding agents and sequencing analysis as described herein. Association of a recording tag with a compartment tagged peptide can be accomplished using a trifunctional linker molecule. After immobilization of the compartment tagged peptide with an associated recording tag to a solid support for cyclic sequencing analysis, the compartment information is transferred to the associated recording tag using primer extension or ligation (*see,* Figure 22B). After transferring the compartment tag information to the recording tag, the compartment tag can be cleaved from the peptide using the same enzyme used in the original peptide digestion (*see,* Figure 22B). This restores the original N-terminal end of the peptide, thus enabling N-terminal degradation peptide sequencing methods as described herein.

EXAMPLE 15: DI-TAG GENERATION BY ASSOCIATING RECORDING TAGS OF PEPTIDES COVALENTLY MODIFIED WITH AMINO ACID-SPECIFIC CODING TAGS VIA THREE PRIMER FUSION EMULSION PCR

Peptides with recording tags comprised of a compartment tag and a molecular UMI are chemically modified with coding tag site-specific chemical labels. The coding tag also contains a UMI to enable counting of the number of amino acids of a given type within a modified peptide. Using a modified protocol from Tyson and Armor (Tyson and Armour 2012), emulsion PCRs are prepared in a total aqueous volume of 100 µl, containing 1x PHUSION™ GC reaction buffer (Thermo Fisher Scientific), 200 µM each dNTPs (New England Biolabs), 1 µM primer U1, 1 µM primer U2tr, 25 nM primer Sp, 14 units PHUSION™ high fidelity DNA polymerase (Thermo Fisher Scientific). 10 µl aqueous phase is added every 5 to 10 seconds to 200 µl oil phase (4.5% vol./vol.) Span 80, 0.4% vol./vol. Tween 80 and 0.05% Triton X-100 dissolved in light mineral oil (Sigma)) in a 2 ml cryo-vial while stirring at 1000 rpm for a total of 5 minutes as previously described by Turner and Hurles (2009, Nat. Protoc. 4:1771-1783). Average droplet size of the resultant emulsion was about 5 microns. Other methods of emulsion generation, such as the use of T-junctions and flow focusing, can also be employed (Brouzes, Medkova et al. 2009). After emulsion generation, 100 µl of aqueous/oil mixture is transferred to 0.5 ml PCR tubes and first-round amplification carried out at the following conditions: 98°C for 30 seconds; 40 cycles of 98°C for 10 seconds, 70°C for 30 seconds and 72°C for 30 seconds; followed by extension at 72°C for 5 minutes. A second-round amplification reaction is carried out at the following conditions: 98°C for 30 seconds; 40 cycles of 98°C for 10 seconds, 55°C for 30 seconds and 72°C for 30 seconds; followed by hold at 4°C. Emulsions are disrupted as soon as possible after the final cycle of the PCR by adding 200 µl hexane (Sigma) directly to the PCR tube, vortexing for 20 seconds, and centrifuging at 13,000 g for 3 minutes.

Example 16: Sequencing Extended Recording Tag, Extended Coding Tag, or Di-Tag Constructs

The spacer (Sp) or universal priming sites of a recording tag or coding tag can be designed using only three bases (*e.g.*, A, C, and T) in the body of the sequence, and a fourth base (*e.g.*, G) at the 5' end of the sequence. For sequencing by synthesis (SBS), this enables rapid dark base incorporation across the spacer sequence using a mix of standard dark (unlabeled and non-terminated) nucleotides (dATP, dGTP, and dTTP) and a single ffC dye-labeled reversible terminator (e.g., fully functional cytosine triphosphate). In this way, only the relevant encoder sequence, unique molecular identifier(s), compartment tags, binding cycle sequence of the extended reporter tag, extended coding tag, or di-tag are SBS sequenced, and the non-relevant spacer or universal priming sequences are "skipped over". The identities of the bases for the spacer and the fourth base at the 5' end of the sequence may be changed and the above identities are provided for purposes of illustration only.

EXAMPLE 17: PREPARATION OF PROTEIN LYSATES.

There are a wide variety of protocols known in the art for making protein lysates from various sample types. Most variations on the protocol depend on cell type and whether the extracted proteins in the lysate in are to be analyzed in a non-denatured or denatured state. For the NGPA assay, either native conformation or denatured proteins can be immobilized to a solid substrate (see Figure 32). Moreover, after immobilization of native proteins, the proteins immobilized on the substrate's surface can be denatured. The advantage of employing denatured proteins are two-fold. First of all, many antibody reagents bind linear epitopes (e.g., Western Blot Abs), and denatured proteins provide better access to linear epitopes. Secondly, the NGPA assay workflow is simplified when using denatured proteins since the annealed coding tag can be stripped from the extended recording tag using alkaline (e.g., 0.1 NaOH) stripping conditions since the immobilized protein is already denatured. This contrasts with the removal of annealed coding tags using assays comprising proteins in their native conformation, that require an enzymatic removal of the annealed coding tag following binding event and information transfer.

Examples of non-denaturing protein lysis buffers include: RPPA buffer consisting of 50 mm HEPES (pH 7.4), 150 mM NaCl, 1% Triton X-100, 1.5 mM MgCl2, 10% glycerol; and commercial buffers such as M-PER mammalian protein extraction reagent (Thermo-Fisher). A denaturing lysis buffer comprises 50 mm HEPES (pH 8.), 1% SDS. The addition of Urea (1M-3M) or Guanidine HCl (1-8M) can also be used in denaturing the protein sample. In addition to the above components of lysis buffers, protease and phosphatase inhibitors are also generally included. Examples of protease inhibitors and typical concentrations include aptrotinin (2 μg/ml), leupeptin (5-10 μg/ml), benzamidine (15 μg/ml), pepstatin A (1 μg/ml), PMSF (1 mM), EDTA (5 mM), and EGTA (1 mM). Examples of phosphatase inhibitors include Na pyrophosphate (10 mM), sodium fluoride (5-100 mM) and sodium orthovanadate (1mM). Additional additives can include DNAaseI to remove DNA from the protein sample, and reducing agents such as DTT to reduce disulfide bonds.

An example of a non-denaturing protein lysate protocol prepared from tissue culture cells is as follows: Adherent cells are trypsinized (0.05% trypsin-EDTA in PBS), collected by centrifugation (200g for 5 min.), and washed 2X in ice cold PBS. Ice-cold M-PER mammalian extraction reagent (~1 mL per 10^7 cells/100 mm dish or 150 cm^2 flask) supplemented with protease/phosphatase inhibitors and additives (e.g., EDTA free complete inhibitors (Roche) and PhosStop (Roche) is added. The resulting cell suspension is incubated on a rotating shaker at 4°C for 20 min. and then centrifuged at 4°C at ~12,000 rpm (depending on cell type) for 20 min to isolate the protein supernatant. The protein is quantitated using the BCA assay, and resuspended at 1 mg/ml in PBS. The protein lysates can be used immediately or snap frozen in liquid nitrogen and stored at -80 ℃.

An example of a denaturing protein lysate protocol, based on the SP3 protocol of Hughs et al., prepared from tissue culture cells is as follows: adherent cells are trypsinized (0.05% trypsin-EDTA in PBS), collected by centrifugation (200g for 5 min.), and washed 2X in ice cold PBS. Ice-cold denaturing lysis buffer (~1 mL per 10^7 cells/100 mm dish or 150 cm^2 flask) supplemented with protease/phosphatase inhibitors and additives (e.g. 1X cOmplete Protease Inhibitor Cocktail (Roche)) is added. The resulting cell suspension is incubated at 95°C for 5 min. and placed on ice for 5 min. Benzonase Nuclease (500 U/ml) is added to the lysate and incubated at 37°C for 30 min. to remove DNA and RNA.

The proteins are reduced by addition of 5 μL of 200 mM DTT per 100 uL of lysate and incubated for 45°C for 30 min. Alklylation of protein cysteine groups is accomplished by addition of 10 uL of 400 mM iodoacetamide per 100 uL of lysate and incubated in the dark at 24° for 30 min. Reactions are quenched by addition of 10 uL of 200 mM DTT per 100 uL of lysate. Proteins are optionally acylated by adding 2 ul an acid anhydride and 100 ul of 1 M Na2CO3 (pH 8.5) per 100 ul of lysate. Incubate for 30 min. at room temp. Valeric, benzoic, and proprionic anhydride are recommended rather than acetic anhydride to enable “in vivo” acetylated lysines to be distinguished from “in situ” blocking of lysine groups by acylation (Sidoli, Yuan et al. 2015). The reaction is quenched by addition of 5 mg of Tris(2-aminoethyl)amine, polymer (Sigma) and incubation at room temperature for 30 min. Polymer resin is removed by centrifuging lysate at 2000g for 1 min. through a 0.45 um cellulose acetate Spin-X tube (Corning). The protein is quantitated using the BCA assay, and resuspended at 1 mg/ml in PBS.

In additional examples, labeled peptides are generated using a filter-aided sample preparation (FASP) protocol, as described by Erde et al. in which a MWCO filtration device is used for protein entrapment, alkylation, and peptidase digestion (Erde, Loo et al. 2014, Feist and Hummon 2015).

EXAMPLE 18: GENERATION OF PARTITION-TAGGED PEPTIDES.

A DNA tag (with an optional sample barcode, and an orthogonal attachment moiety) is used to label the ε-amino groups on lysines of denatured polypeptides using standard bioconjugation methods (Hermanson 2013), or alternatively, are attached to the polypeptide using photoaffinity labeling (PAL) methods such as benzophenone (Li, Liu et al. 2013). After labeling of the polypeptide with DNA tags at lysine groups or randomly on CH groups (via PAL) and blocking unlabeled groups via acylation with an acyl anhydride, the DNA-tag labeled, acylated polypeptides are annealed to compartment beads with attached DNA oligonucleotides comprising a universal priming sequence, a compartment barcode, an optional UMI, and a primer sequence complementary to a portion of the DNA tag attached to the polypeptides. Because of the cooperativity of multiple DNA hybridization tags, single polypeptide molecule interacts primarily with a single bead enabling writing of the same compartment barcode to all DNA tags of the polypeptide molecule. After annealing, the polypeptide-bound DNA tag primes a polymerase extension reaction on the annealed bead-bound DNA sequence. In this manner, the compartment barcodes and other functional elements are written onto the DNA tags attached to the bound polypeptide. Upon completion of this step, the polypeptide has a plurality of recording tags attached, wherein the recording tag has a common spacer sequence, barcode sequences (e.g. sample, fraction, compartment, spatial, etc.), optional UMIs and other functional elements. This labeled polypeptide can be digested into peptide fragments using standard endoproteases such as trypsin, GluC, proteinase K, etc. Note: if trypsin is used for digestion of lysine-labeled polypeptides, the polypeptide is only cleaved at Arg residues not Lys residues (since Lys residues are labeled). The protease digestion can be done on directly on the beads or after removal of the labeled polypeptide from the barcoded beads.

EXAMPLE 19: PREPARING DNA RECORDING TAG-PEPTIDE CONJUGATES FOR MODEL SYSTEM.

The recording tag oligonucleotides are synthesized with a 5' NH_2 group, and an internal mTetrazine group for later coupling to beads (alkyne-dT is converted to mTetrazine-dT via an mTet-PEG-N_3 heterobifunctional crosslinking agent). The 5' NH_2 of the oligonucleotide is coupled to a reactive cysteine on a peptide using an NHS/maleimide heterobifunctional cross-linker, such as LC-SMCC (ThermoFisher Scientific), as described by Williams et al. (Williams and Chaput 2010). In particular, 20 nmols of 5' NH_2-labeled oligonucleotides are ethanol precipitated and resuspended in 180 ul of phosphate coupling buffer (0.1 M potassium phosphate buffer, pH 7.2) in a siliconized tube. 5 mg of LC-SMCC is resuspended in 1 mL of DMF (5 mg/ml) (store in aliquots at -20). An aliquot of 20 ul LC-SMCC (5 mg/ml) is added to 180 ul of the resuspended oligonucleotides, mixed and incubated at room temperature for 1 hr. The mixture is 2X ethanol precipitated. The resultant malemide-derivitized oligonucleotide is resuspended in 200 ul phosphate coupling buffer. A peptide containing a cysteine residue (>95% purity, desalted) is resuspended at 1 mg/ml (~0.5 mM) in DMSO. Approximately 50 nmol of peptide (100 ul) are added to the reaction mix, and incubated at room temperature overnight. The resultant DNA recording tag-peptide conjugate is purified using native-PAGE as described by William et al. (Williams and Chaput 2010). Conjugates are resuspended in phosphate coupling buffer at 100 uM concentration in siliconized tubes.

EXAMPLE 20: DEVELOPMENT OF SUBSTRATE FOR DNA-PEPTIDE IMMOBILIZATION.

Magnetic beads suitable for click-chemistry immobilization are created by converting M-270 amine magnetic Dynabeads to either azide or TCO-derivatized beads capable of coupling to alkyne or methyl Tetrazine-labeled oligo-peptide conjugates, respectively (see, e.g., Figures 29D-E; Figures 30D-E). Namely, 10 mg of M-270 beads are washed and resuspended in 500 ul borate buffer (100 mM sodium borate, pH 8.5). A mixture of TCO-PEG (12-120)-NHS (Nanocs) and methyl-PEG (12-120)-NHS is resuspended at 1 mM in DMSO and incubated with M-270 amine beads at room temperature overnight. The ratio of the Methyl to TCO PEG is titrated to adjust the final TCO surface density on the beads such that there is < 100 TCO moieties/um^2 (see, e.g., Figure 31E; Figure 34). Unreacted amine groups are capped with a mixture of 0.1M acetic anhydride and 0.1M DIEA in DMF (500 ul for 10 mg of beads) at room temperature for 2 hrs. After capping and washing 3X in DMF, the beads are resuspended in phosphate coupling buffer at 10 mg/ml.

EXAMPLE 21: IMMOBILIZATION OF RECORDING TAG LABELED PEPTIDES TO SUBSTRATE.

Recording tag labeled peptides are immobilized on a substrate via an IEDDA click chemistry reaction using an mTet group on the recording tag and a TCO group on the surface of activated beads or substrate. This reaction is fast and efficient, even at low input concentrations of reactants. Moreover, the use of methyl tetrazine confers greater stability to the bond (Selvaraj and Fox 2013, Knall, Hollauf et al. 2014, Wu and Devaraj 2016). 200 ng of M-270 TCO beads are resuspended in 100 ul phosphate coupling buffer. 5 pmol of DNA recording tag labeled peptides comprising an mTet moiety on the recording tag is added to the beads for a final concentration of ~ 50 nM. The reaction is incubated for 1 hr. at room temperature. After immobilization, unreacted TCO groups on the substrate are quenched with 1 mM methyl tetrazine acid in phosphate coupling buffer for 1 hr. at room temperature.

EXAMPLE 22: N-TERMINAL AMINO ACID (NTAA) MODIFICATION

<u>Chemical NTAA Acetylation</u>:

The NTAA of a peptide is acetylated using either acetic anhydride or NHS-acetate in organic or aqueous solutions (sulfo-NHS-acetate). For acetic anhydride derivatization, 10 mM of acetic anhydride in DMF is incubated with the peptide for 30 min. at RT

(Halpin, Lee et al. 2004). Alternatively, the the peptide is acetylated in aqueous solution using 50 mM acetic anhydride in 100 mM 2-(N-morpholino)ethanesulfonate (MES) buffer (pH 6.0) and 1M NaCl at RT for 30 min (Tse, Snyder et al. 2008). For NHS-acetate derivatization, a stock solution of sulfo-NHS-acetate (100 mM in DMSO) is prepared and added at a final concentration of 5-10 mM in 100 mM sodium phosphate buffer (pH 8.0) or 100 mM borate buffer (pH 9.4) and incubated for 10-30 min. at RT (Goodnow 2014).

<u>Enzymatic NTAA Acetylation:</u>

NTAA of a peptide is enzymatically acetylated by exposure to N-Acetyl Transferase (SsArd1 from *Sulfolobus solfataricus*) using the following conditions: peptides are incubated with 2 μM SsArd1 in NAT buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 1 mM acetyl-CoA) at 65°C for 10 min (Chang and Hsu 2015).

<u>Chemical NTAA Amidination (Guanidination):</u>

Peptides are incubated with 10 mM N,N-bis(tert-butoxycarbonyl) thiourea, 20 mM trimethylamine, and 12 mM Mukayama's reagent (2-chloro-1-methylpyridinium iodide) in DMF at RT for 30 min. Alternatively, the peptides are incubated with 10 mM 1H-Pyrazole-1-carboxamidine Hydrochloride, 10 mM DIEA in DMF at RT for 30 min. Standard deblocking methods are used to remove protecting groups. Alternatively, the peptides are incubated with 10 mM S-methylisothiourea in PBS buffer (pH 8.0) or 100 mM borate buffer (pH 8.0) for 30 min. at 10°C (Tse, Snyder et al. 2008).

<u>PITC Labeling:</u>

Peptide is incubated with 5% (vol./vol.) PITC in ionic liquid [Bmim][BF4] at room temperature for 5 min. The reaction time is optimized for quantitative PITC labelling of NTAA while minimizing ectopic labeling of the exocyclic amines on nucleotide bases present in the extended DNA recording tag.

DNFB Labeling:

2,4- Dinitrofluorobenzene (DNFB) is prepared as a 5 mg/ml stock in methanol. The solution is protected from light and prepared fresh daily. Peptides are labeled by incubation in 0.5 -5.0 ug/ml DNFB in 10 mM borate buffer (pH 8.0) at 37°C for 5-30 min.

SNFB Labeling:

4-sulfonyl-2-nitro-fluorobenzene (SNFB) is prepared as a 5 mg/ml stock in methanol. The solution should be protected from light and prepared fresh daily. Peptides are labeled by incubation in 0.5-5.0 ug/ml DNFB in 10 mM borate buffer (pH 8.0) at 37°C for 5-30 min.

Cleavage of Acetylated NTAA Peptides:

The acetylated NTAA is cleaved from the peptide by incubation with 10 uM acylpeptide hydrolase (APH) enzyme (from Sulfolobus solfataricus, SSO2693) in 25 mM Tris-HCl (pH 7.5) at 90°C for 10 min (Gogliettino, Balestrieri et al. 2012).

Cleavage of Amidinated NTAA Peptides:

The amidinated (guanidinated) NTAA is cleaved from the peptide by incubation in 0.1N NaOH for 10 min. at 37°C (Hamada 2016).

EXAMPLE 23: DEMONSTRATION OF INTRAMOLECULAR TRANSFER OF CODING TAG INFORMATION TO RECORDING TAGS WITH MODEL SYSTEM

DNA model system was used to test the “intra-molecular” transfer of coding tag information to recording tags that are immobilized to beads (see, Figure 36A). Two different types of recording tag oligonucleotides were used. saRT_Abc_v2 (SEQ ID NO:141) contained an “A” DNA capture sequence (SEQ ID NO:153) (mimic epitope for “A’” binding agent) and a corresponding “A” barcode (rtA_BC); saRT_Bbc_V2 (SEQ ID NO:142) contained a “B” DNA capture sequence (SEQ ID NO:154) (mimic epitope for “B’” binding agent) and a corresponding “B” barcode (rtB_BC). These barcodes were combinations of the elementary 65 set of 15-mer barcodes (SEQ ID NOS:1-65) and their reverse complementary sequences (SEQ ID NOS:66-130). rtA_BC is a collinear combination of two barcodes, BC_1 and BC_2, and rtB_BC is just the one barcode, BC_3. Likewise the barcodes (encoder sequences) on the coding tags were also comprised of barcodes from the elementary set of 65 15-mer barcodes (SEQ ID NOS:1-65). CT_A'-bc_1PEG (SEQ ID NO:144) and CT_B'-bc (SEQ ID NO:147) coding tags were comprised of complementary capture sequences, A' and B', respectively, and were assigned the 15-mer barcodes, BC_5, and BC_5 & BC_6, respectively. This design set-up for the recording tags and coding tags enables easy gel analysis. The desired "intra-molecular" primer extension generates oligonucleotide products of similar size, whereas the undesired "inter-molecular" extension generates one oligo product 15 bases larger and another oligo product 15 bases shorter than the "intra-molecular" product (Figure 36B).

The effect of recording tag density on "intra-molecular" vs. "inter-molecular" information transfer was evaluated. For correct information transfer, "intra-molecular" information transfer ("A'" coding tag to A recording tag; B' coding tag to B recording tag), should be observed rather than "inter-molecular" information transfer (A'coding tag binding to A recording tag but transferring information to B recording tag, and vice versa). To test the effect of recording tags spacing on the bead surface, biotinylated recording tag oligonucleotides, saRT_Abc_v2 (SEQ ID NO:141) and saRT_Bbc_v2 (SEQ ID NO:142), were mixed in a 1:1 ratio, and then titrated against the saDummy-T10 oligonucleotide (SEQ ID NO:143) in ratios of 1:0, 1:10, 1:10^2, 1:10^3, and 1:10^4. A total of 20 pmols of recording tag oligonucleotides was incubated with 5 ul of M270 streptavidin beads (Thermo) in 50 ul Immobilization buffer (5 mM Tris-Cl (pH 7.5), 0.5 mM EDTA, 1 M NaCl) for 15 min. at 37 °C. The beads were washed 3X with 100 ul Immobilization buffer at room temperature. Most subsequent wash steps used a volume of 100 ul. Coding tags (duplex annealing with DupCT sequences required for later cycles) were annealed to the recording tags immobilized on the beads by resuspending the beads in 25 ul of 5X Annealing buffer (50 mM Tris-Cl (pH 7.5), 10 mM MgCl2) and adding the coding tag mix. The coding tags annealed to the recording tags by heating to 65 ºC for 1 min, and then allowed to slow cool to room temperature (0.2 ºC/sec). Alternatively, coding tags can be annealed in PBST buffer at 37 ºC. Beads were washed PBST (PBS + 0.1% Tween-20) at room temp, and washed 2X with PBST at 37 ºC for 5 min. and washed 1X with PBST at room temp. and a final wash in 1X Annealing buffer. The beads were resuspended in 19.5 ul Extension buffer (50 mM Tris-Cl (pH 7.5), 2 mM MgSO4, 125 uM dNTPs, 50 mM NaCl, 1 mM dithiothreitol, 0.1% Tween-20, and 0.1 mg/ml BSA) and incubated at 37 ºC for 15 min. Klenow exo-DNA polymerase (NEB, 5 U/ul) was added to the beads for a final concentration of 0.125 U/ul, and incubated at 37 ºC for 5 min. After primer extension, beads were washed 2X with PBST, and 1X with 50 ul 0.1 NaOH at room temp for 5 min., and 3X with PBST and 1X with PBS. To add the downstream PCR adapter sequence, R1', the EndCap2T oligo (comprised of R1 (SEQ ID NO:152) was hybridized and extended on the beads as done for the coding tag oligonucleotides. After adding the adapter sequence, the final extended recording tag oligonucleotides were eluted from the streptavidin beads by incubation in 95% formamide/10 mM EDTA at 65 ºC for 5 min. Approximately 1/100th of the eluted product was PCR amplified in 20 ul for 18 cycles, and 1 ul of PCR product analyzed on a 10% denaturing PAGE gel. The resulting gels demonstrates proof of principle of writing coding tag information to the recording tag by polymerase extension (Figure 36C), and the ability to generate a primarily "intra-molecular" extension events relative to "inter-molecular" extension events upon dilution of recording tag density on the surface of the bead.

In this model system, the size of PCR products from recording tags RT_ABC and RT_BBC that contain the corresponding encoder sequence and universal reverse primer site is 100 base pairs (Figure 36C), while the products by incorrect pairings of saRT_ABC (SEQ ID NO:141)/CT_B'BC (SEQ ID NO:147) and saRT_BBC (SEQ ID NO:142)/CT_A'BC (SEQ ID NO:144) are 115 and 85 base pairs, respectively. As shown in Figure 36D, three bands were observed in the presence of saRT_ABC (SEQ ID NO:141) and saRT_BBC (SEQ ID NO:142) on beads at high density. It was expected that the recoding tag extended on proximal coding tag binding to itself (intra-molecular event) or neighbor recoding tag (inter-molecular event) at the high density.

However the bands of products by incorrect pairings decreased by diluting the recoding tags in dummy oligonucleotide, and disappeared at a ratio of 1:10000. This result demonstrated that the recording tags were spaced out on beads surface at the low density, resulting in decreased intermolecular events.

Table 1. Model System Sequences

Name	Sequence (5'-3')	SEQ ID NO:
saRT_Abc_v2	/5Biosg/TTTTTGCAAATGGCATTCTGACATCCCGTAGTCCGCGACACTAGATGTCTAGCATGCCGCCGTGTCATGTGGAAACTGAGTG	141
saRT_Bbc_v2	/5Biosg/TTTTTTTTTTGACTGGTTCCAATTGACAAGCCGTAGTCCGCGACACTAGTAAGCCGGTATATCAACTGAGTG	142
saDummy-pT10	/5Biosg/TTTTTTTTTT/3SpC3/	143
CT_A'-bc	GGATGTCAGAATGCCATTTGCTTTTTTTTTT/iSP18/CACTCAGTCCTAACGCGTATACGCACTCAGT/3SpC3/	144
CT_A'-bc_1PEG	GGATGTCAGAATGCCATTTGCTTTTTTTTTT/iSP18/CACTCAGTCCTAACGCGTATACGTCACTCAGT/3SpC3/	145
CT_A'bc_5PEG	GGATGTCAGAATGCCATTTGCTTTTTTTTTT/iSP18//iSP18//iSP18//iSP18//iSP18/CACTCAGTCCTAACGCGTATACGTCACTCAGT/3SpC3/	146
CT_B'bc	GCTTGTCAATTGGAACCAGTCTTTT/iSp18/CACTCAGTCCTAACGCGTATACGGGAATCTCGGCAGTTCACTCAGT/3SpC3/	147
EndCap2T	CGATTTGCAAGGATCACTCGTCACTCAGTCCTAACGCGTATACG/3SpC3/	148
Sp	ACTGAGTG	149
Sp'	CACTCAGT	150
P1_f2	CGTAGTCCGCGACACTAG	151
R1	CGATTTGCAAGGATCACTCG	152
dupCT_A'BC	CGTATACGCGTTAGGACTGAGTG/3SpC3/	153
dupCT_B'BC	AACTGCCGAGATTCCCGTATACGCGTTAGGACTGAGTG/3SpC3/	154

/3SpC3/ = 3' C3 (three carbon) spacer

/5Biosg/ = 5' Biotin

/iSP18/ = 18-atom hexa-ethyleneglycol spacer

EXAMPLE 24: SEQUENCING EXTENDED RECORDING TAG, EXTENDED CODING TAG, OR DI-TAG CONSTRUCTS ON NANOPORE SEQUENCERS

DNA barcodes can be designed to be tolerant to highly-error prone NGS sequencers, such as nanopore-based sequencers where the current base call error rate is on the order of 10% or more. A number of error correcting code systems have been described in the literature. These include Hamming codes, Reed-Solomon codes, Levenshtein codes, Lee codes, etc. Error-tolerant barcodes were based on Hamming and Levenshtein codes using R Bioconductor package, “DNAbarcodes” capable of correcting insertion, deletion, and substitution errors, depending on the design parameters chosen (Buschmann and Bystrykh 2013). A set of 65 different 15-mer Hamming barcodes are shown in Figure 27A (as set forth in SEQ ID NOS:1-65 and their reverse complementary sequences in SEQ ID NOS:66-130, respectively). These barcodes have a minimum Hamming distance of 10 and are self-correcting out to four substitution errors and two indel errors, more than sufficient to be accurately readout on a nanopore sequencer with a 10% error rate. Moreover, these barcodes have been filtered from a set of 77 original barcodes using the predicted nanopore current signatures (see Figure 27B). They were filtered to have large current level differences across the barcode, and to be maximally uncorrelated with other barcodes in the set. In this way, actual raw nanopore current level plots from assays using these barcodes can be mapped directly to the predicted barcode signature without using base calling algorithms (Laszlo, Derrington et al. 2014).

To mimic the analysis of extended recording tags, extended coding tags, or di-tag constructs using nanopore sequencing, PCR products comprised of a small subset of 15-mer barcodes using four forward primers (DTF1 (SEQ ID NO:157), DTF2 (SEQ ID NO:158), DTF3 (SEQ ID NO:159), DTF4 (SEQ ID NO:160)) and four reverse primers (DTR9 (SEQ ID NO:161), DTR10 (SEQ ID NO:162), DTR11 (SEQ ID NO:163), DTR12 (SEQ ID NO:164)) were generated (Figure 27C). This set of 8 primers was included in a PCR reaction along with a flanking forward primer F1 (SEQ ID NO:165), and reverse primer R1 (SEQ ID NO:166). The DTF and DTR primers annealed via an complementary 15-mer spacer sequence (Sp15) (SEQ ID NO:167). The combination of 4 DTF forward and 4 DTR reverse primers leads to a set of 16 possible PCR products.

PCR Conditions:

Reagent	Final Conc.
F1 (5' phosphorylated) (SEQ ID NO:165)	1 uM
R1 (5' phosphorylated) (SEQ ID NO:166)	1 uM
DTF1-4 (SEQ ID NOS:157-160); DTR9-12 (SEQ ID NOS:161-164)	0.3 nM ea
VeraSeq Buffer 2	1X
dNTPs	200 uM
water	
VeraSeq 2.0 Ultra Pol	2 U/100 ul

PCR Cycling:

98 °C	30 sec
50 °C	2 min
98 °C	10 sec
55 °C	15 sec
72 °C	15 sec
Repeat last 3 steps for 19 cycles	
72 °C	5 min

After PCR, the amplicons were concatenated by blunt end ligation (Figure 27C) as follows: 20 ul PCR product was mixed directly with 20 ul Quick Ligase Mix (NEB) and incubated overnight at room temp. The resultant ligated product, ~ 0.5 - 2 kb in length, was purified using a Zymo purification column and eluted into 20 ul water. About 7 ul of this purified ligation product was used directly in the MinIon Library Rapid Sequencing Prep kit (SQK-RAD002) and analyzed on a MinION Mk 1B (R9.4) device. An example of a 734 bp nanopore read of quality score 7.2 (~80% accuracy) is shown in Figure 27D. Despite the poor sequencing accuracy, a large number of barcodes are easily readable in the sequence as indicated by *lalign-based* alignment of the barcodes to the MinIon sequence read (Figure 27D).

EXAMPLE 25: ENCAPSULATED SINGLE CELLS IN GEL BEADS

Single cells are encapsulated into droplets (~50 μm) using standard techniques (Tamminen and Virta 2015, Spencer, Tamminen et al. 2016) (see Figure 38).

A Polyacrylamide (Acrylamide:bisacrylamide (29:1) (30% w/vol.)), benzophenone methacrylamide (BM) , and APS is included in the discontinuous phase along with the cells to create droplets capable of polymerizing upon addition of TEMED in the continuous oil phase (diffuses into droplets). Benzophenone is cross-linked into the matrix of the polyacrylamide gel droplet. This allows subsequent photoaffinity crosslinking of the proteins to the polyacrylamide matrix (Hughes, Spelke et al. 2014, Kang, Yamauchi et al. 2016). The proteins immobilized within the resulting single cell gel bead, can be single cell barcoded using a variety of methods. In one embodiment, DNA tags are chemically or photo-chemically attached to the immobilized proteins in the single cell gel beads using amine-reactive agents or a photo-active benzophenone DNA tag as previously described. The single cell gel beads can be encapsulated in droplets containing barcodes via co-encapsulation of barcoded beads as previously described and the DNA barcode tag transferred to the proteins, or alternatively proteins within single cell gel beads can be combinatorically indexed through a series of pool-and-split steps as described by Amini, Cusanovich, and Gunderson et al. (Amini, Pushkarev et al. 2014, Cusanovich, Daza et al. 2015)(Gunderson, Steemers et al. 2016). In the simplest implementation, the proteins within single cell gel beads are first labeled with “click-chemistry” moieties (see Figure 40), and then combinatorial DNA barcodes are clicked onto the protein samples using the pool-and-split approach.

References:

Harlow, Ed, and David Lane. Using Antibodies. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1999.

Hennessy BT, Lu Y, Gonzalez-Angulo AM, et al. A Technical Assessment of the Utility of Reverse Phase Protein Arrays for the Study of the Functional Proteome in Non-microdissected Human Breast Cancers. *Clinical proteomics*. 2010;6(4):129-151.

Davidson, G. R., S. D. Armstrong and R. J. Beynon (2011). "Positional proteomics at the N-terminus as a means of proteome simplification." Methods Mol Biol **753**: 229-242.

Zhang, L., Luo, S., and Zhang, B. (2016). The use of lectin microarray for assessing glycosylation of therapeutic proteins. mAbs *8*, 524–535.

Akbani, R., K. F. Becker, N. Carragher, T. Goldstein, L. de Koning, U. Korf, L. Liotta, G. B. Mills, S. S. Nishizuka, M. Pawlak, E. F. Petricoin, 3rd, H. B. Pollard, B. Serrels and J. Zhu (2014). "Realizing the promise of reverse phase protein arrays for clinical, translational, and basic research: a workshop report: the RPPA (Reverse Phase Protein Array) society." Mol Cell Proteomics **13**(7): 1625-1643.

Amini, S., D. Pushkarev, L. Christiansen, E. Kostem, T. Royce, C. Turk, N. Pignatelli, A. Adey, J. O. Kitzman, K. Vijayan, M. Ronaghi, J. Shendure, K. L. Gunderson and F. J. Steemers (2014). "Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing." Nat Genet **46**(12): 1343-1349.
Assadi, M., J. Lamerz, T. Jarutat, A. Farfsing, H. Paul, B. Gierke, E. Breitinger, M. F. Templin, L. Essioux, S. Arbogast, M. Venturi, M. Pawlak, H. Langen and T. Schindler (2013). "Multiple protein analysis of formalin-fixed and paraffin-embedded tissue samples with reverse phase protein arrays." Mol Cell Proteomics **12**(9): 2615-2622.
Bailey, J. M. and J. E. Shively (1990). "Carboxy-terminal sequencing: formation and hydrolysis of C-terminal peptidylthiohydantoins." Biochemistry **29**(12): 3145-3156.
Bandara, H. M., D. P. Kennedy, E. Akin, C. D. Incarvito and S. C. Burdette (2009). "Photoinduced release of Zn2+ with ZinCleav-1: a nitrobenzyl-based caged complex." Inorg Chem **48**(17): 8445-8455.
Bandara, H. M., T. P. Walsh and S. C. Burdette (2011). "A Second-generation photocage for Zn2+ inspired by TPEN: characterization and insight into the uncaging quantum yields of ZinCleav chelators." Chemistry **17**(14): 3932-3941.
Basle, E., N. Joubert and M. Pucheault (2010). "Protein chemical modification on endogenous amino acids." Chem Biol **17**(3): 213-227.
Bilgicer, B., S. W. Thomas, 3rd, B. F. Shaw, G. K. Kaufman, V. M. Krishnamurthy, L. A. Estroff, J. Yang and G. M. Whitesides (2009). "A non-chromatographic method for the purification of a bivalently active monoclonal IgG antibody from biological fluids." J Am Chem Soc **131**(26): 9361-9367.
Bochman, M. L., K. Paeschke and V. A. Zakian (2012). "DNA secondary structures: stability and function of G-quadruplex structures." Nat Rev Genet **13**(11): 770-780.
Borgo, B. and J. J. Havranek (2014). "Motif-directed redesign of enzyme specificity." Protein Sci **23**(3): 312-320.
Brouzes, E., M. Medkova, N. Savenelli, D. Marran, M. Twardowski, J. B. Hutchison, J. M. Rothberg, D. R. Link, N. Perrimon and M. L. Samuels (2009). "Droplet microfluidic technology for single-cell high-throughput screening." Proc Natl Acad Sci U S A **106**(34): 14195-14200.
Brudno, Y., M. E. Birnbaum, R. E. Kleiner and D. R. Liu (2010). "An in vitro translation, selection and amplification system for peptide nucleic acids." Nat Chem Biol **6**(2): 148-155.
Calcagno, S. and C. D. Klein (2016). "N-Terminal methionine processing by the zinc-activated Plasmodium falciparum methionine aminopeptidase 1b." Appl Microbiol Biotechnol.
Cao, Y., G. K. Nguyen, J. P. Tam and C. F. Liu (2015). "Butelase-mediated synthesis of protein thioesters and its application for tandem chemoenzymatic ligation." Chem Commun (Camb) **51**(97): 17289-17292.
Carty, R. P. and C. H. Hirs (1968). "Modification of bovine pancreatic ribonuclease A with 4-sulfonyloxy-2-nitrofluorobenzene. Isolation and identification of modified proteins." J Biol Chem **243**(20): 5244-5253.
Chang, L., D. M. Rissin, D. R. Fournier, T. Piech, P. P. Patel, D. H. Wilson and D. C. Duffy (2012). "Single molecule enzyme-linked immunosorbent assays: theoretical considerations." J Immunol Methods **378**(1-2): 102-115.
Chang, Y. Y. and C. H. Hsu (2015). "Structural basis for substrate-specific acetylation of Nalpha-acetyltransferase Ard1 from Sulfolobus solfataricus." Sci Rep **5**: 8673.

Christoforou, A., C. M. Mulvey, L. M. Breckels, A. Geladaki, T. Hurrell, P. C. Hayward, T. Naake, L. Gatto, R. Viner, A. Martinez Arias and K. S. Lilley (2016). "A draft map of the mouse pluripotent stem cell spatial proteome." Nat Commun **7**: 8992.
Creighton, C. J. and S. Huang (2015). "Reverse phase protein arrays in signaling pathways: a data integration perspective." Drug Des Devel Ther **9**: 3519-3527.
Crosetto, N., M. Bienko and A. van Oudenaarden (2015). "Spatially resolved transcriptomics and beyond." Nat Rev Genet **16**(1): 57-66.
Cusanovich, D. A., R. Daza, A. Adey, H. A. Pliner, L. Christiansen, K. L. Gunderson, F. J. Steemers, C. Trapnell and J. Shendure (2015). "Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing." Science **348**(6237): 910-914.
Derrington, I. M., T. Z. Butler, M. D. Collins, E. Manrao, M. Pavlenok, M. Niederweis and J. H. Gundlach (2010). "Nanopore DNA sequencing with MspA." Proc Natl Acad Sci U S A **107**(37): 16060-16065.
El-Sagheer, A. H., V. V. Cheong and T. Brown (2011). "Rapid chemical ligation of oligonucleotides by the Diels-Alder reaction." Org Biomol Chem **9**(1): 232-235.
El-Sagheer, A. H., A. P. Sanzone, R. Gao, A. Tavassoli and T. Brown (2011). "Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in Escherichia coli." Proc Natl Acad Sci U S A **108**(28): 11338-11343.
Emili, A., M. McLaughlin, K. Zagorovsky, J. B. Olsen, W. C. W. Chan and S. S. Sidhu (2017). Protein Sequencing Method and Reagents. USPTO. USA, The Governing Council of University of Toronto. **9,566,335 B1**.
Erde, J., R. R. Loo and J. A. Loo (2014). "Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments." J Proteome Res **13**(4): 1885-1895.
Farries, T. C., A. Harris, A. D. Auffret and A. Aitken (1991). "Removal of N-acetyl groups from blocked peptides with acylpeptide hydrolase. Stabilization of the enzyme and its application to protein sequencing." Eur J Biochem **196**(3): 679-685.
Feist, P. and A. B. Hummon (2015). "Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples." Int J Mol Sci **16**(2): 3537-3563.
Friedmann, D. R. and R. Marmorstein (2013). "Structure and mechanism of non-histone protein acetyltransferase enzymes." FEBS J **280**(22): 5570-5581.
Frokjaer, S. and D. E. Otzen (2005). "Protein drug stability: a formulation challenge." Nat Rev Drug Discov **4**(4): 298-306.
Fujii, Y., M. Kaneko, M. Neyazaki, T. Nogi, Y. Kato and J. Takagi (2014). "PA tag: a versatile protein tagging system using a super high affinity antibody against a dodecapeptide derived from human podoplanin." Protein Expr Purif **95**: 240-247.
Gebauer, M. and A. Skerra (2012). "Anticalins small engineered binding proteins based on the lipocalin scaffold." Methods Enzymol **503**: 157-188.
Gerry, N. P., N. E. Witowski, J. Day, R. P. Hammer, G. Barany and F. Barany (1999). "Universal DNA microarray method for multiplex detection of low abundance point mutations." J Mol Biol **292**(2): 251-262.
Gogliettino, M., M. Balestrieri, E. Cocca, S. Mucerino, M. Rossi, M. Petrillo, E. Mazzella and G. Palmieri (2012). "Identification and characterisation of a novel acylpeptide hydrolase from Sulfolobus solfataricus: structural and functional insights." PLoS One 7(5): e37921.

Gogliettino, M., A. Riccio, M. Balestrieri, E. Cocca, A. Facchiano, T. M. D'Arco, C. Tesoro, M. Rossi and G. Palmieri (2014). "A novel class of bifunctional acylpeptide hydrolases--potential role in the antioxidant defense systems of the Antarctic fish Trematomus bernacchii." FEBS J **281**(1): 401-415.

Granvogl, B., M. Ploscher and L. A. Eichacker (2007). "Sample preparation by in-gel digestion for mass spectrometry-based proteomics." Anal Bioanal Chem **389**(4): 991-1002.

Gunderson, K. L., X. C. Huang, M. S. Morris, R. J. Lipshutz, D. J. Lockhart and M. S. Chee (1998). "Mutation detection by ligation to complete n-mer DNA arrays." Genome Res **8**(11): 1142-1153.

Gunderson, K. L., F. J. Steemers, J. S. Fisher and R. Rigatti (2016). Methods and Compositions for Analyzing Cellular Components. WIPO, Illumina, Inc.

Gunderson, K. L., F. J. Steemers, J. S. Fisher and R. Rigatti (2016). Methods and compositions for analyzing cellular components, Illumina, Inc.

Guo, H., W. Liu, Z. Ju, P. Tamboli, E. Jonasch, G. B. Mills, Y. Lu, B. T. Hennessy and D. Tsavachidou (2012). "An efficient procedure for protein extraction from formalin-fixed, paraffin-embedded tissues for reverse phase protein arrays." Proteome Sci **10**(1): 56.

Hamada, Y. (2016). "A novel N-terminal degradation reaction of peptides via N-amidination." Bioorg Med Chem Lett **26**(7): 1690-1695.

Hermanson, G. (2013). Bioconjugation Techniques, Academic Press.

Hernandez-Moreno, A. V., F. Villasenor, E. Medina-Rivero, N. O. Perez, L. F. Flores-Ortiz, G. Saab-Rincon and G. Luna-Barcenas (2014). "Kinetics and conformational stability studies of recombinant leucine aminopeptidase." Int J Biol Macromol **64**: 306-312.

Hori, M., H. Fukano and Y. Suzuki (2007). "Uniform amplification of multiple DNAs by emulsion PCR." Biochem Biophys Res Commun **352**(2): 323-328.

Horisawa, K. (2014). "Specific and quantitative labeling of biomolecules using click chemistry." Front Physiol **5**: 457.

Hoshika, S., F. Chen, N. A. Leal and S. A. Benner (2010). "Artificial genetic systems: self-avoiding DNA in PCR and multiplexed PCR." Angew Chem Int Ed Engl **49**(32): 5554-5557.

Hughes, A. J., D. P. Spelke, Z. Xu, C. C. Kang, D. V. Schaffer and A. E. Herr (2014). "Single-cell western blotting." Nat Methods **11**(7): 749-755.

Hughes, C. S., S. Foehr, D. A. Garfield, E. E. Furlong, L. M. Steinmetz and J. Krijgsveld (2014). "Ultrasensitive proteome analysis using paramagnetic bead technology." Mol Syst Biol **10**: 757.

Kang, C. C., K. A. Yamauchi, J. Vlassakis, E. Sinkala, T. A. Duncombe and A. E. Herr (2016). "Single cell-resolution western blotting." Nat Protoc **11**(8): 1508-1530.

Kang, T. S., L. Wang, C. N. Sarkissian, A. Gamez, C. R. Scriver and R. C. Stevens (2010). "Converting an injectable protein therapeutic into an oral form: phenylalanine ammonia lyase for phenylketonuria." Mol Genet Metab **99**(1): 4-9.

Katritzky, A. R. and B. V. Rogovoy (2005). "Recent developments in guanylating agents." ARKIVOC **iv**(Issue in Honor of Prof. Nikolai Zefirov): 49-87.

Klein, A. M., L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz and M. W. Kirschner (2015). "Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells." Cell **161**(5): 1187-1201.

Knall, A. C., M. Hollauf and C. Slugovc (2014). "Kinetic studies of inverse electron demand Diels-Alder reactions (iEDDA) of norbornenes and 3,6-dipyridin-2-yl-1,2,4,5-tetrazine." Tetrahedron Lett **55**(34): 4763-4766.

Le, Z. G., Z. C. Chen, Y. Hu and Q. G. Zheng (2005). "Organic Reactions in Ionic Liquids: Ionic Liquid-promoted Efficient Synthesis of Disubstituted and Trisubstituted Thioureas Derivatives." Chinese Chemical Letters **16**(2): 201-204.

Lesch, V., A. Heuer, V. A. Tatsis, C. Holm and J. Smiatek (2015). "Peptides in the presence of aqueous ionic liquids: tunable co-solutes as denaturants or protectants?" Phys Chem Chem Phys **17**(39): 26049-26053.

Li, G., Y. Liu, Y. Liu, L. Chen, S. Wu, Y. Liu and X. Li (2013). "Photoaffinity labeling of small-molecule-binding proteins by DNA-templated chemistry." Angew Chem Int Ed Engl **52**(36): 9544-9549.

Litovchick, A., M. A. Clark and A. D. Keefe (2014). "Universal strategies for the DNA-encoding of libraries of small molecules using the chemical ligation of oligonucleotide tags." Artif DNA PNA XNA **5**(1): e27896.

Liu, Y. and S. Liang (2001). "Chemical carboxyl-terminal sequence analysis of peptides and proteins using tribenzylsilyl isothiocyanate." J Protein Chem **20**(7): 535-541.

Lundblad, R. L. (2014). Chemical reagents for protein modification. Boca Raton, CRC Press, Taylor & Francis Group.

Mashaghi, S. and A. M. van Oijen (2015). "External control of reactions in microdroplets." Sci Rep **5**: 11837.

McCormick, R. M. (1989). "A solid-phase extraction procedure for DNA purification." Anal Biochem **181**(1): 66-74.

Mendoza, V. L. and R. W. Vachet (2009). "Probing protein structure by amino acid-specific covalent labeling and mass spectrometry." Mass Spectrom Rev **28**(5): 785-815.

Mikami, T., T. Takao, K. Yanagi and H. Nakazawa (2012). "N (alpha) Selective Acetylation of Peptides." Mass Spectrom (Tokyo) **1**(2): A0010.

Moghaddam, M. J., L. de Campo, N. Kirby and C. J. Drummond (2012). "Chelating DTPA amphiphiles: ion-tunable self-assembly structures and gadolinium complexes." Phys Chem Chem Phys **14**(37): 12854-12862.

Mukherjee, S., M. Ura, R. J. Hoey and A. A. Kossiakoff (2015). "A New Versatile Immobilization Tag Based on the Ultra High Affinity and Reversibility of the Calmodulin-Calmodulin Binding Peptide Interaction." J Mol Biol **427**(16): 2707-2725.

Namimatsu, S., M. Ghazizadeh and Y. Sugisaki (2005). "Reversing the effects of formalin fixation with citraconic anhydride and heat: a universal antigen retrieval method." J Histochem Cytochem **53**(1): 3-11.

Nguyen, G. K., Y. Cao, W. Wang, C. F. Liu and J. P. Tam (2015). "Site-Specific N-Terminal Labeling of Peptides and Proteins using Butelase 1 and Thiodepsipeptide." Angew Chem Int Ed Engl **54**(52): 15694-15698.

Nguyen, G. K., S. Wang, Y. Qiu, X. Hemu, Y. Lian and J. P. Tam (2014). "Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis." Nat Chem Biol **10**(9): 732-738.

Nishizuka, S. S. and G. B. Mills (2016). "New era of integrated cancer biomarker discovery using reverse-phase protein arrays." Drug Metab Pharmacokinet **31**(1): 35-45.

Ohkubo, A., R. Kasuya, K. Sakamoto, K. Miyata, H. Taguchi, H. Nagasawa, T. Tsukahara, T. Watanobe, Y. Maki, K. Seio and M. Sekine (2008). "'Protected DNA

Probes' capable of strong hybridization without removal of base protecting groups." Nucleic Acids Res **36**(6): 1952-1964.
Ojha, B., A. K. Singh, M. D. Adhikari, A. Ramesh and G. Das (2010). "2-Alkylmalonic acid: amphiphilic chelator and a potent inhibitor of metalloenzyme." J Phys Chem B **114**(33): 10835-10842.
Peng, X., H. Li and M. Seidman (2010). "A Template-Mediated Click-Click Reaction: PNA-DNA, PNA-PNA (or Peptide) Ligation, and Single Nucleotide Discrimination." European J Org Chem **2010**(22): 4194-4197.
Perbandt, M., O. Bruns, M. Vallazza, T. Lamla, C. Betzel and V. A. Erdmann (2007). "High resolution structure of streptavidin in complex with a novel high affinity peptide tag mimicking the biotin binding motif." Proteins **67**(4): 1147-1153.
Rauth, S., D. Hinz, M. Borger, M. Uhrig, M. Mayhaus, M. Riemenschneider and A. Skerra (2016). "High-affinity Anticalins with aggregation-blocking activity directed against the Alzheimer beta-amyloid peptide." Biochem J **473**(11): 1563-1578.
Ray, A. and B. Norden (2000). "Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future." FASEB J **14**(9): 1041-1060.
Riley, N. M., A. S. Hebert and J. J. Coon (2016). "Proteomics Moves into the Fast Lane." Cell Syst **2**(3): 142-143.
Roloff, A., S. Ficht, C. Dose and O. Seitz (2014). "DNA-templated native chemical ligation of functionalized peptide nucleic acids: a versatile tool for single base-specific detection of nucleic acids." Methods Mol Biol **1050**: 131-141.
Roloff, A. and O. Seitz (2013). "The role of reactivity in DNA templated native chemical PNA ligation during PCR." Bioorg Med Chem **21**(12): 3458-3464.
Sakurai, K., T. M. Snyder and D. R. Liu (2005). "DNA-templated functional group transformations enable sequence-programmed synthesis using small-molecule reagents." J Am Chem Soc **127**(6): 1660-1661.
Schneider, K. and B. T. Chait (1995). "Increased stability of nucleic acids containing 7-deaza-guanosine and 7-deaza-adenosine may enable rapid DNA sequencing by matrix-assisted laser desorption mass spectrometry." Nucleic Acids Res **23**(9): 1570-1575.
Selvaraj, R. and J. M. Fox (2013). "trans-Cyclooctene--a stable, voracious dienophile for bioorthogonal labeling." Curr Opin Chem Biol **17**(5): 753-760.
Sharma, A. K., A. D. Kent and J. M. Heemstra (2012). "Enzyme-linked small-molecule detection using split aptamer ligation." Anal Chem **84**(14): 6104-6109.
Shembekar, N., C. Chaipan, R. Utharala and C. A. Merten (2016). "Droplet-based microfluidics in drug discovery, transcriptomics and high-throughput molecular genetics." Lab Chip **16**(8): 1314-1331.
Shenoy, N. R., J. E. Shively and J. M. Bailey (1993). "Studies in C-terminal sequencing: new reagents for the synthesis of peptidylthiohydantoins." J Protein Chem **12**(2): 195-205.
Shim, J. U., R. T. Ranasinghe, C. A. Smith, S. M. Ibrahim, F. Hollfelder, W. T. Huck, D. Klenerman and C. Abell (2013). "Ultrarapid generation of femtoliter microfluidic droplets for single-molecule-counting immunoassays." ACS Nano **7**(7): 5955-5964.
Shim, J. W., Q. Tan and L. Q. Gu (2009). "Single-molecule detection of folding and unfolding of the G-quadruplex aptamer in a nanopore nanocavity." Nucleic Acids Res **37**(3): 972-982.
Sidoli, S., Z. F. Yuan, S. Lin, K. Karch, X. Wang, N. Bhanu, A. M. Arnaudo, L. M. Britton, X. J. Cao, M. Gonzales-Cope, Y. Han, S. Liu, R. C. Molden, S. Wein, L.

Afjehi-Sadat and B. A. Garcia (2015). "Drawbacks in the use of unconventional hydrophobic anhydrides for histone derivatization in bottom-up proteomics PTM analysis." Proteomics **15**(9): 1459-1469.
Sletten, E. M. and C. R. Bertozzi (2009). "Bioorthogonal chemistry: fishing for selectivity in a sea of functionality." Angew Chem Int Ed Engl **48**(38): 6974-6998.
Spencer, S. J., M. V. Tamminen, S. P. Preheim, M. T. Guo, A. W. Briggs, I. L. Brito, A. W. D, L. K. Pitkanen, F. Vigneault, M. P. Juhani Virta and E. J. Alm (2016). "Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers." ISME J **10**(2): 427-436.
Spicer, C. D. and B. G. Davis (2014). "Selective chemical protein modification." Nat Commun **5**: 4740.
Spiropulos, N. G. and J. M. Heemstra (2012). "Templating effect in DNA proximity ligation enables use of non-bioorthogonal chemistry in biological fluids." Artif DNA PNA XNA **3**(3): 123-128.
Switzar, L., M. Giera and W. M. Niessen (2013). "Protein digestion: an overview of the available techniques and recent developments." J Proteome Res **12**(3): 1067-1077.
Tamminen, M. V. and M. P. Virta (2015). "Single gene-based distinction of individual microbial genomes from a mixed population of microbial cells." Front Microbiol **6**: 195.
Tessler, L. (2011). Digital Protein Analysis: Technologies for Protein Diagnostics and Proteomics through Single-Molecule Detection. Ph.D., WASHINGTON UNIVERSITY IN ST. LOUIS.
Tyson, J. and J. A. Armour (2012). "Determination of haplotypes at structurally complex regions using emulsion haplotype fusion PCR." BMC Genomics **13**: 693.
Vauquelin, G. and S. J. Charlton (2013). "Exploring avidity: understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands." Br J Pharmacol **168**(8): 1771-1785.
Veggiani, G., T. Nakamura, M. D. Brenner, R. V. Gayet, J. Yan, C. V. Robinson and M. Howarth (2016). "Programmable polyproteams built using twin peptide superglues." Proc Natl Acad Sci U S A **113**(5): 1202-1207.
Wang, D., S. Fang and R. M. Wohlhueter (2009). "N-terminal derivatization of peptides with isothiocyanate analogues promoting Edman-type cleavage and enhancing sensitivity in electrospray ionization tandem mass spectrometry analysis." Anal Chem **81**(5): 1893-1900.
Williams, B. A. and J. C. Chaput (2010). "Synthesis of peptide-oligonucleotide conjugates using a heterobifunctional crosslinker." Curr Protoc Nucleic Acid Chem **Chapter 4**: Unit4 41.
Wu, H. and N. K. Devaraj (2016). "Inverse Electron-Demand Diels-Alder Bioorthogonal Reactions." Top Curr Chem (J) **374**(1): 3.
Xiong, A. S., R. H. Peng, J. Zhuang, F. Gao, Y. Li, Z. M. Cheng and Q. H. Yao (2008). "Chemical gene synthesis: strategies, softwares, error corrections, and applications." FEMS Microbiol Rev **32**(3): 522-540.
Yao, Y., M. Docter, J. van Ginkel, D. de Ridder and C. Joo (2015). "Single-molecule protein sequencing through fingerprinting: computational assessment." Phys Biol **12**(5): 055003.

Zakeri, B., J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy and M. Howarth (2012). "Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin." Proc Natl Acad Sci U S A **109**(12): E690-697.
Zhang, L., K. Zhang, S. Rauf, D. Dong, Y. Liu and J. Li (2016). "Single-Molecule Analysis of Human Telomere Sequence Interactions with G-quadruplex Ligand." Anal Chem **88**(8): 4533-4540.
Zhou, H., Z. Ning, A. E. Starr, M. Abu-Farha and D. Figeys (2012). "Advancements in top-down proteomics." Anal Chem **84**(2): 720-734.
Zilionis, R., J. Nainys, A. Veres, V. Savova, D. Zemmour, A. M. Klein and L. Mazutis (2017). "Single-cell barcoding and sequencing using droplet microfluidics." Nat Protoc **12**(1): 44-73.

These and other changes can be made to the embodiments in light of the above-detailed description. In general, in the following claims, the terms used should not be construed to limit the claims to the specific embodiments disclosed in the specification and the claims, but should be construed to include all possible embodiments along with the full scope of equivalents to which such claims are entitled. Accordingly, the claims are not limited by the disclosure.

The various embodiments described above can be combined to provide further embodiments. All U.S. patents, U.S. patent application publications, U.S. patent applications, foreign patents, foreign patent applications, and non-patent publications referred to in this specification and/or listed in the Application Data Sheet, including U.S. Provisional Patent Application No. 62/330,841, U.S. Provisional Patent Application No. 62/339,071, and U.S. Provisional Patent Application No. 62/376,886, are incorporated herein by reference, in their entirety. are incorporated herein by reference in their entirety. Aspects of the embodiments can be modified, if necessary to employ concepts of the various patents, applications and publications to provide yet further embodiments.

2017259794 16 Mar 2023

CLAIMS

1. A method for analyzing a macromolecule, comprising the steps of:

(a) providing the macromolecule and a nucleic acid recording tag associated or colocalized directly or indirectly with the macromolecule, the macromolecule and/or the nucleic acid recording tag being joined to a solid support;

(b) contacting the macromolecule with a first binding agent capable of binding to the macromolecule, wherein the first binding agent comprises a first nucleic acid coding tag that comprises identifying information regarding the first binding agent;

(c) following binding of the first binding agent to the macromolecule, transferring the identifying information regarding the first binding agent from the first nucleic acid coding tag to the nucleic acid recording tag to generate a first order extended nucleic acid recording tag joined to the solid support;

(d) contacting the macromolecule with a second binding agent capable of binding to the macromolecule, wherein the second binding agent comprises a second nucleic acid coding tag that comprises identifying information regarding the second binding agent;

(e) following binding of the second binding agent to the macromolecule, transferring the identifying information regarding the second binding agent from the second nucleic acid coding tag to the first order extended nucleic acid recording tag to generate a second order extended nucleic acid recording tag joined to the solid support; and

(f) analyzing the second order extended nucleic acid recording tag, wherein analyzing comprises a sequencing method, and obtaining the identifying information regarding the first binding agent and the identifying information regarding the second binding agent to provide information regarding the macromolecule, thereby analyzing the macromolecule.

2. The method of claim 1, further comprising, between steps (e) and (f), the following steps:

repeating steps (d) and (e) one or more times by replacing the second binding agent with a third or higher order binding agent capable of binding to the macromolecule, wherein the third or higher order binding agent comprises a third or higher order nucleic acid coding tag that comprises identifying information regarding the third or higher order binding agent; and by transferring the identifying information regarding the third or higher order binding agent from the third or higher order nucleic acid coding tag to the second or higher order extended nucleic acid recording tag to generate a third or higher order extended nucleic acid recording tag joined to the solid support;

and wherein the third or higher order extended nucleic acid recording tag is analyzed in step (f).

3. The method of claim 1 or claim 2, wherein:

step (a) comprises providing a plurality of macromolecules and associated nucleic acid recording tags joined to the solid support; and

each of steps (b) and (d) comprises contacting the plurality of macromolecules with a first or second plurality of binding agents capable of binding to the macromolecules, wherein the first or second plurality of binding agents comprise a first or second nucleic acid coding tags with identifying information regarding the first or second plurality of binding agents.

4. The method of any one of claims 1-3, wherein the first binding agent and/or the second binding agent form a non-covalent association with the macromolecule.

5. The method of any one of claims 1-4, wherein the macromolecule is a peptide, and wherein at least one binding agent binds to the N-terminal amino acid (NTAA) residue of the peptide.

6. The method of claim 5, further comprising modifying the NTAA of the peptide with a chemical moiety to produce a modified NTAA.

7. The method of claim 6, wherein at least one binding agent binds to the modified NTAA of the peptide.

8. The method of any one of claims 5-7, further comprising removing the NTAA to expose a new NTAA of the peptide.

9. The method of any one of claims 5-8, wherein the first binding agent and/or the second binding agent bind(s) to a di-peptide or a tri-peptide of the peptide.

10. The method of any one of claims 1-9, wherein the nucleic acid recording tag comprises:

a unique molecule identifier (UMI); and/or

a barcode.

11. The method of claim 3, wherein different macromolecules of the plurality of macromolecules are spaced apart from each other on a surface or within a volume of the solid support at an average distance > 50 nm.

12. The method of any one of claims 1-11, wherein transferring the information of the nucleic acid coding tag to the nucleic acid recording tag is mediated by a DNA ligase, by a DNA polymerase or by chemical ligation.

13. The method of claim 1, wherein contacting steps (b) and (d) are performed at the same time.

14. The method of any one of claims 1-13, wherein the first binding agent and/or the second binding agent is/are capable of selectively binding to the macromolecule.

15. The method of any one of claims 1-14, wherein the first binding agent and the second binding agent each comprises an aptamer or an antibody.

16. The method of any one of claims 1-15, wherein the first binding agent or the second binding agent binds to a post-translationally modified amino acid.

17. The method of claim 1-16, wherein an order of nucleic acid coding tag information contained on the extended nucleic acid recording tag provides information regarding an order of binding by the binding agents to the macromolecule.

18. The method of claim 1-17, wherein the solid support is a polystyrene bead, a polymer bead, an agarose bead, an acrylamide bead, a porous bead or a controlled pore bead.

19. A method for analyzing one or more peptides from a sample comprising a plurality of protein complexes, proteins, or polypeptides, the method comprising:

(a) partitioning the plurality of protein complexes, proteins, or polypeptides within the sample into a plurality of compartments, wherein each compartment comprises a plurality of compartment tags optionally joined to a solid support, wherein the plurality of compartment tags are the same within an individual compartment and are different from the compartment tags of other compartments;

(b) fragmenting the plurality of protein complexes, proteins, and/or polypeptides into a plurality of peptides;

(c) contacting the plurality of peptides to the plurality of compartment tags under conditions sufficient to permit annealing or joining of the plurality of peptides with the plurality of compartment tags within the plurality of compartments, thereby generating a plurality of compartment tagged peptides;

(d) collecting the compartment tagged peptides from the plurality of compartments; and

(e) analyzing one or more compartment tagged peptide according to the method of claim 1.

20. A method for analyzing one or more peptides from a sample comprising a plurality of protein complexes, proteins, or polypeptides, the method comprising:

(a) labeling the plurality of protein complexes, proteins, or polypeptides with a plurality of universal DNA tags;

(b) partitioning the plurality of labeled protein complexes, proteins, or polypeptides within the sample into a plurality of compartments, wherein each compartment comprises a plurality of compartment tags, wherein the plurality of compartment tags are the same within an individual compartment and are different from the compartment tags of other compartments;

(c) contacting the plurality of protein complexes, proteins, or polypeptides to the plurality of compartment tags under conditions sufficient to permit annealing or joining of the plurality of protein complexes, proteins, or polypeptides with the plurality of compartment tags within the plurality of compartments, thereby generating a plurality of compartment tagged protein complexes, proteins or polypeptides;

(d) collecting the compartment tagged protein complexes, proteins, or polypeptides from the plurality of compartments;

(e) optionally fragmenting the compartment tagged protein complexes, proteins, or polypeptides into a compartment tagged peptides; and

(f) analyzing one or more compartment tagged peptide according to the method of claim 1.

21. A method for analyzing a peptide, comprising:

(a) contacting the peptide with a first binding agent,

wherein the peptide is associated with a nucleic acid recording tag joined to a support,

wherein the first binding agent is associated with a first nucleic acid coding tag that comprises identifying information regarding the first binding agent, and

wherein the first binding agent binds to a terminal amino acid (TAA) or a modified TAA of the peptide;

(b) transferring the identifying information regarding the first binding agent from the first nucleic acid coding tag to the nucleic acid recording tag to generate a first order extended nucleic acid recording tag on the support, and/or transferring sequence information of the nucleic acid recording tag to the first nucleic acid coding tag to generate an extended first nucleic acid coding tag;

(c) cleaving the peptide to generate a cleaved peptide, thereby removing the TAA or the modified TAA to expose a new TAA, and, optionally, modifying the new TAA to yield a newly modified TAA;

(d) contacting the cleaved peptide with a second binding agent,

wherein the second binding agent is associated with a second nucleic acid coding tag that comprises identifying information regarding the second binding agent, and

wherein the second binding agent binds to the new TAA or the newly modified TAA;

(e) transferring the identifying information regarding the second binding agent from the second nucleic acid coding tag to the first order extended nucleic acid recording tag to generate a second order extended nucleic acid recording tag on the support, and/or transferring sequence information of the nucleic acid recording tag to the second nucleic acid coding tag to generate an extended second nucleic acid coding tag; and

(f) analyzing the second order extended nucleic acid recording tag and/or analyzing the extended first and extended second nucleic acid coding tags, wherein the analyzing comprises a sequencing method, to obtain the identifying information regarding the first binding agent and the identifying information regarding the second binding agent, thereby providing information regarding an amino acid sequence of the peptide.

22. The method of claim 21, further comprising, between steps (e) and (f):

repeating steps (c), (d) and (e) in one or more cycles, wherein each cycle comprises:

cleaving the cleaved peptide of a previous cycle to generate a cleaved peptide of the cycle, thereby removing the TAA or modified TAA of the previous cycle;

contacting the cleaved peptide of the cycle with a third or higher order binding agent that binds to a new TAA or a newly modified TAA in the cleaved peptide of the cycle, wherein the third or higher order binding agent is associated with a third or higher order nucleic acid coding tag that comprises identifying information regarding the third or higher order binding agent; and

transferring the identifying information regarding the third or higher order binding agent from the third or higher order nucleic acid coding tag to the second or higher order extended nucleic acid recording tag to generate a third or higher order extended nucleic acid recording tag on the support, and/or transferring sequence information of the nucleic acid recording tag to the third or higher order nucleic acid coding tag to generate an extended third or higher order nucleic acid coding tag,

wherein the third or higher order extended nucleic acid recording tag, and/or the extended first, extended second, and extended third or higher order nucleic acid coding tags, is/are analyzed in step (f).

23. The method of claim 21 or 22, wherein in step (a), 100 or more different peptides are each associated with a nucleic acid recording tag joined to the support, and wherein the 100 or more different peptides are analyzed simultaneously.

24. The method of any one of claims 21-23, wherein the first binding agent binds to the TAA and the second binding agent binds to the new TAA.

25. The method of any one of claims 21-24, wherein the first binding agent binds to the modified TAA and the second binding agent binds to the newly modified TAA.

26. The method of any one of claims 21-25, wherein the first binding agent binds to a penultimate terminal amino acid of the peptide in addition to the TAA or the modified TAA.

27. The method of any one of claims 21-26, wherein the peptide is covalently joined to the support before step (a).

REPLACEMENT SHEET
1/60

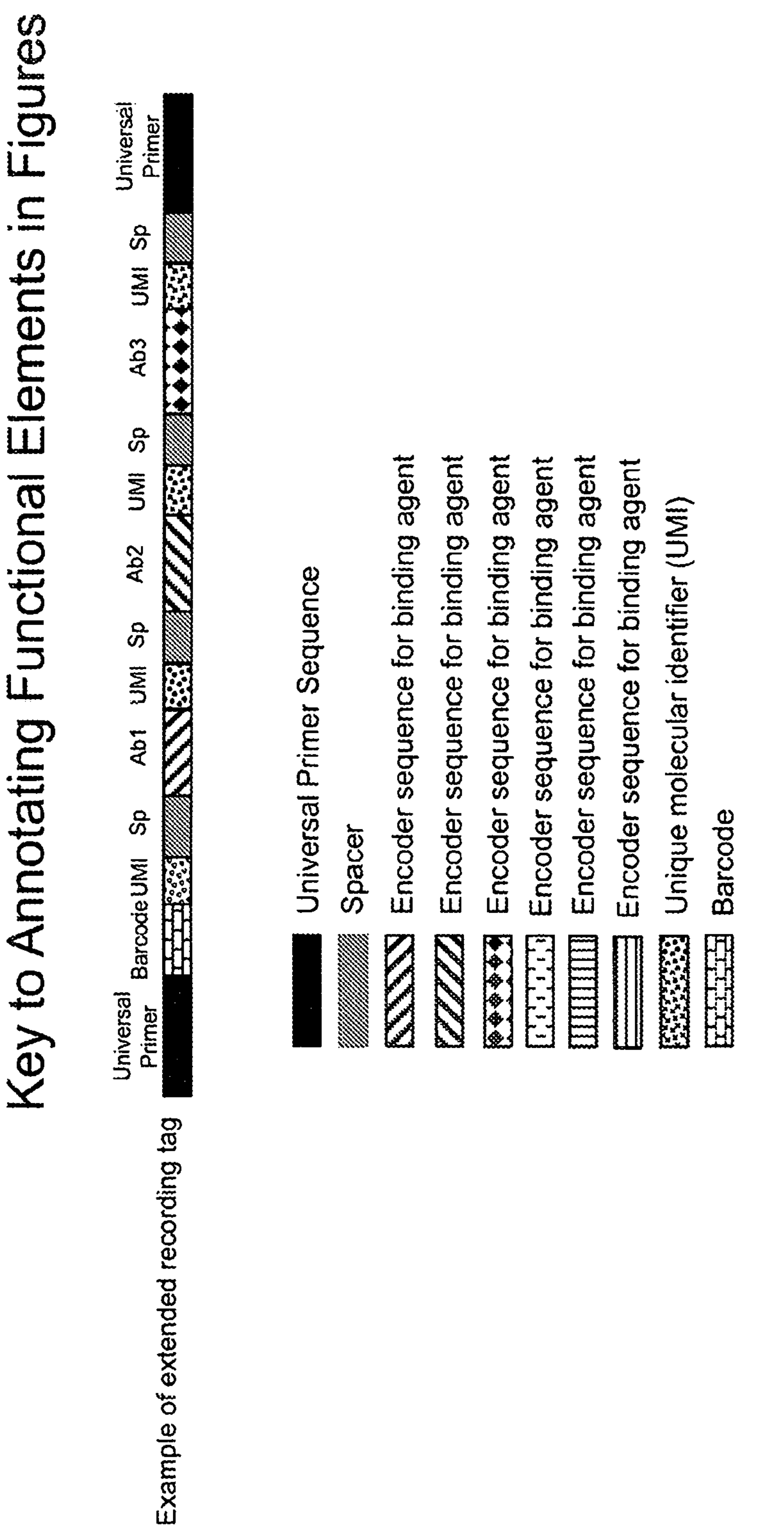


Fig. 1A

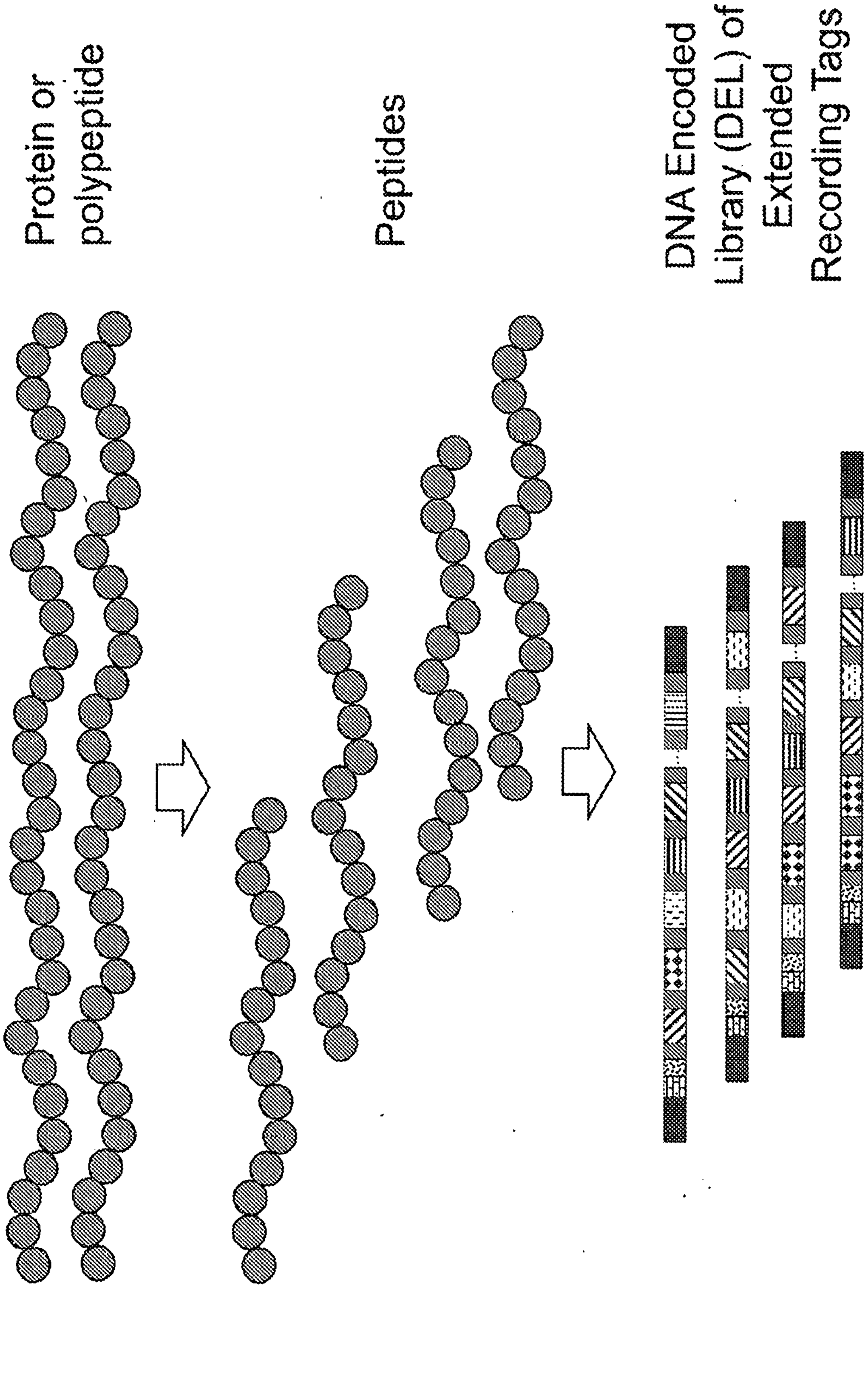

Protein or polypeptide
Peptides
DNA Encoded Library (DEL) of Extended Recording Tags

Fig. 1B

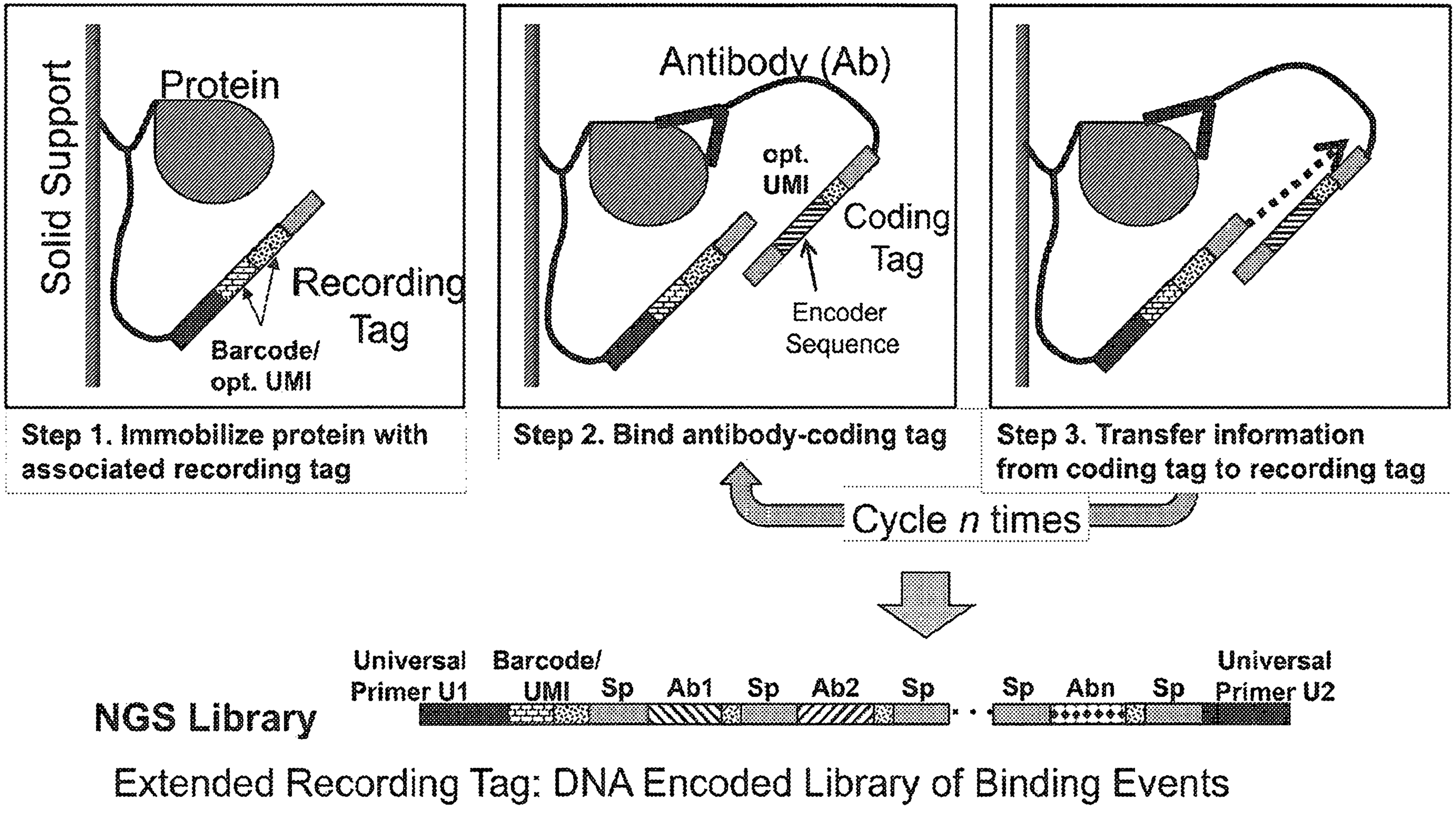

Solid Support
Protein
Recording Tag
Barcode/ opt. UMI
Step 1. Immobilize protein with associated recording tag
Antibody (Ab)
opt. UMI
Coding Tag
Encoder Sequence
Step 2. Bind antibody-coding tag
Step 3. Transfer information from coding tag to recording tag
Cycle n times
NGS Library
Universal Primer U1
Barcode/ UMI
Sp
Ab1
Sp
Ab2
Sp
Sp
Abn
Sp
Universal Primer U2
Extended Recording Tag: DNA Encoded Library of Binding Events

Fig. 2A

REPLACEMENT SHEET
4/60

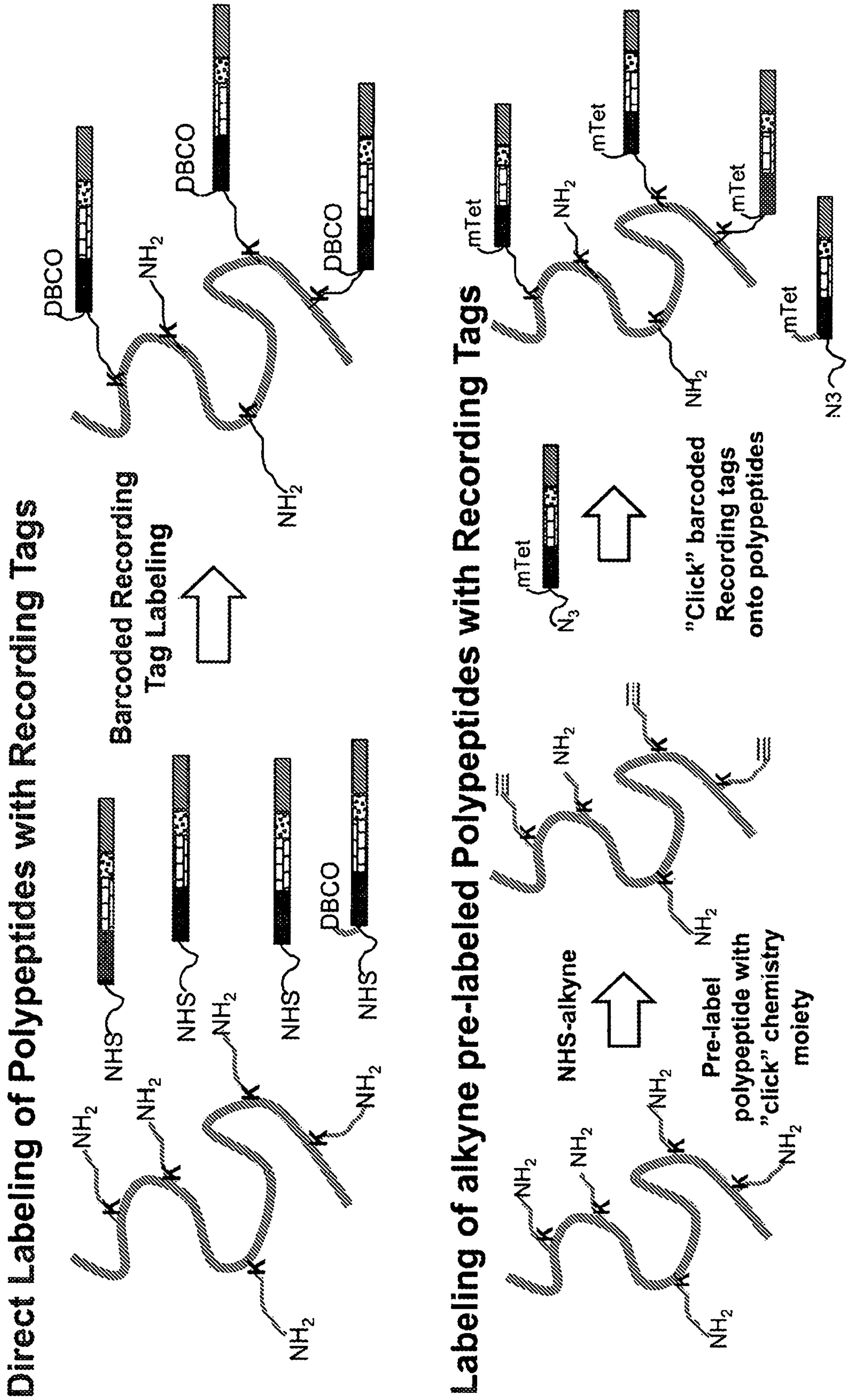


Fig. 2B

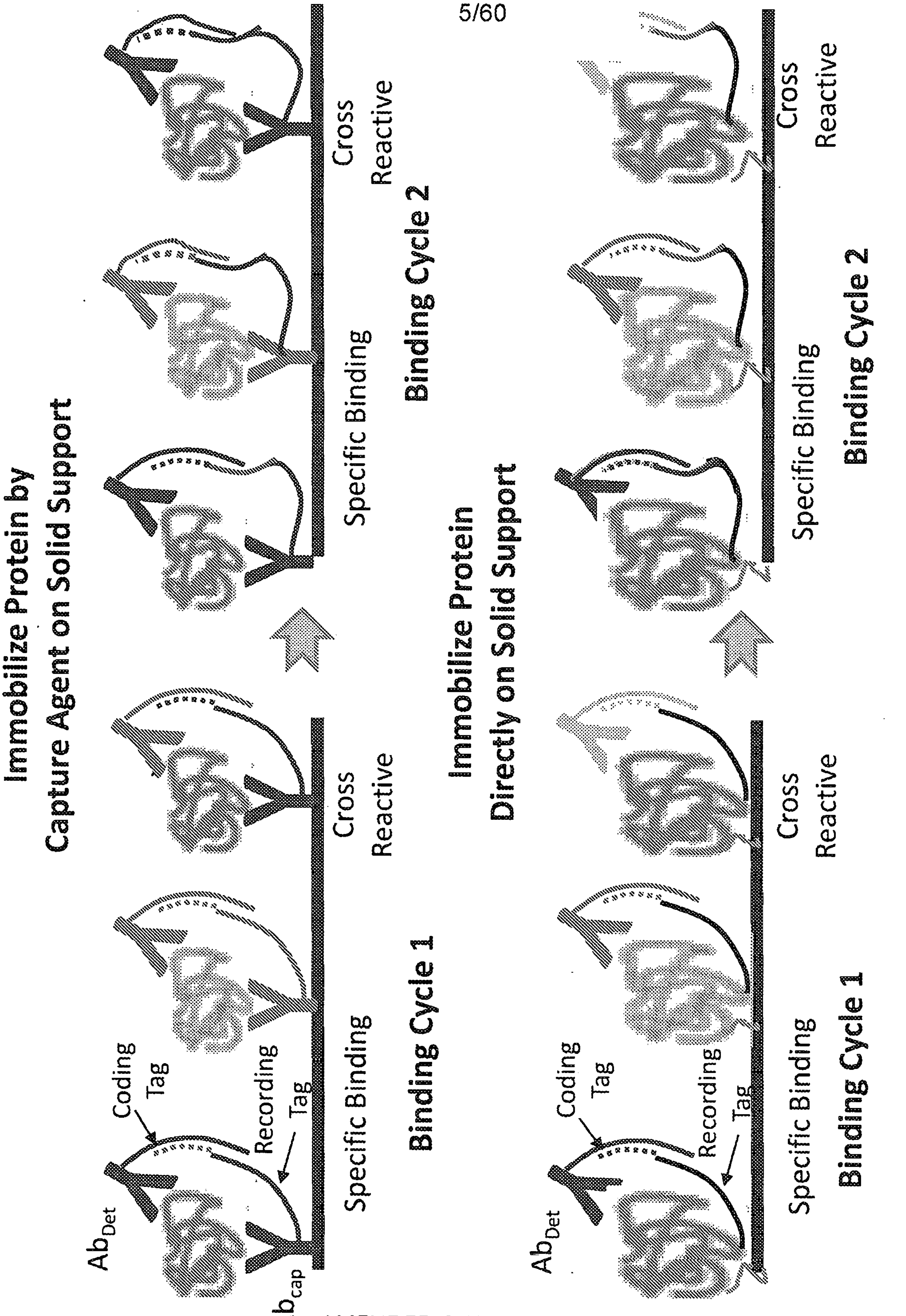

Immobilize Protein by
Capture Agent on Solid Support
Ab_{Det}
Coding
Tag
Recording
Tag
Ab_{cap}
Specific Binding
Cross
Reactive
Binding Cycle 1
Specific Binding
Cross
Reactive
Binding Cycle 2
Immobilize Protein
Directly on Solid Support
Ab_{Det}
Coding
Tag
Recording
Tag
Specific Binding
Cross
Reactive
Binding Cycle 1
Specific Binding
Cross
Reactive
Binding Cycle 2

Fig. 2C

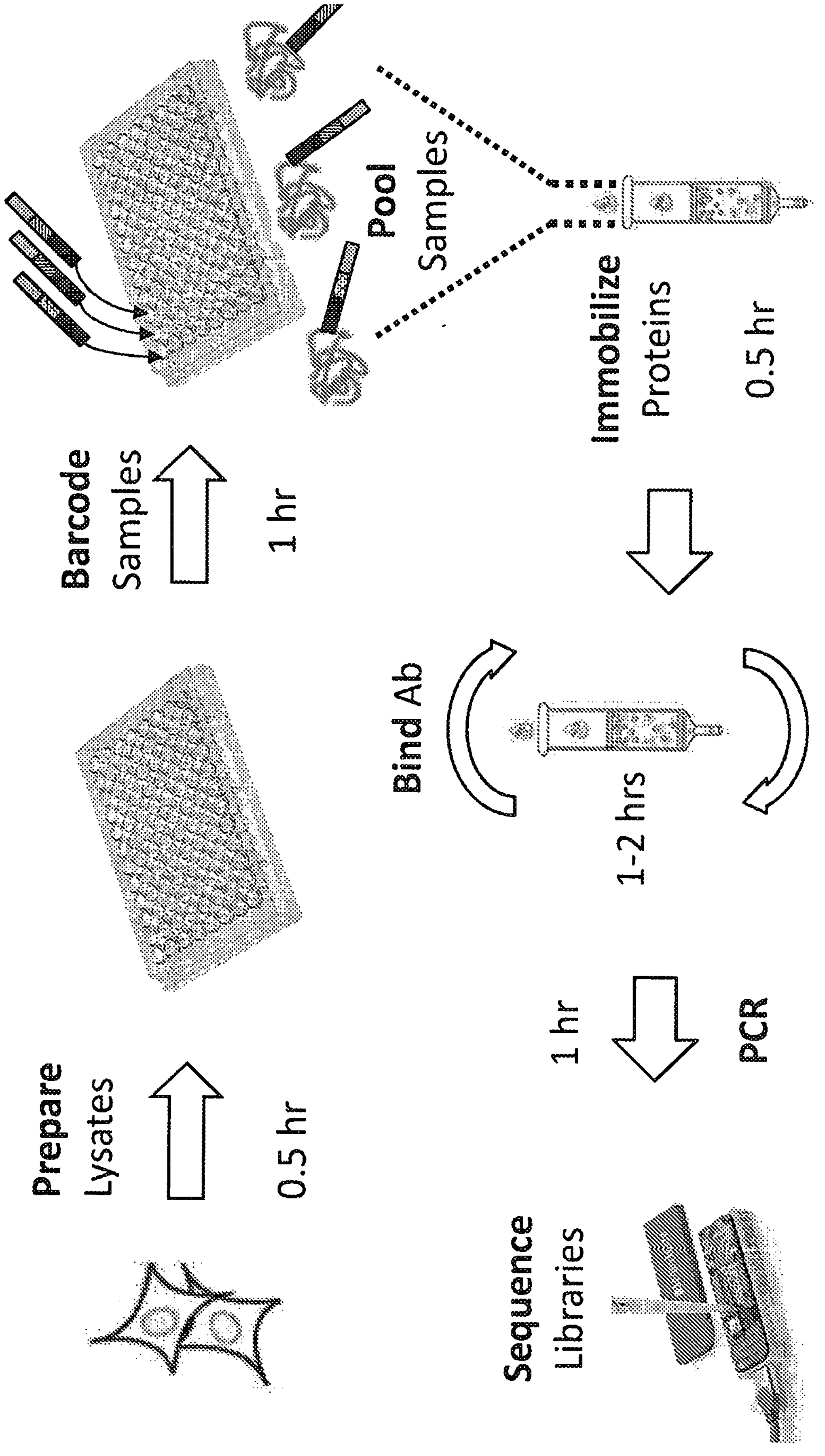

Prepare
Lysates
0.5 hr
Barcode
Samples
1 hr
Pool
Samples
Immobilize
Proteins
0.5 hr
Bind Ab
1-2 hrs
Transfer
binding event
information
1 hr
PCR
Sequence
Libraries

Fig. 2D

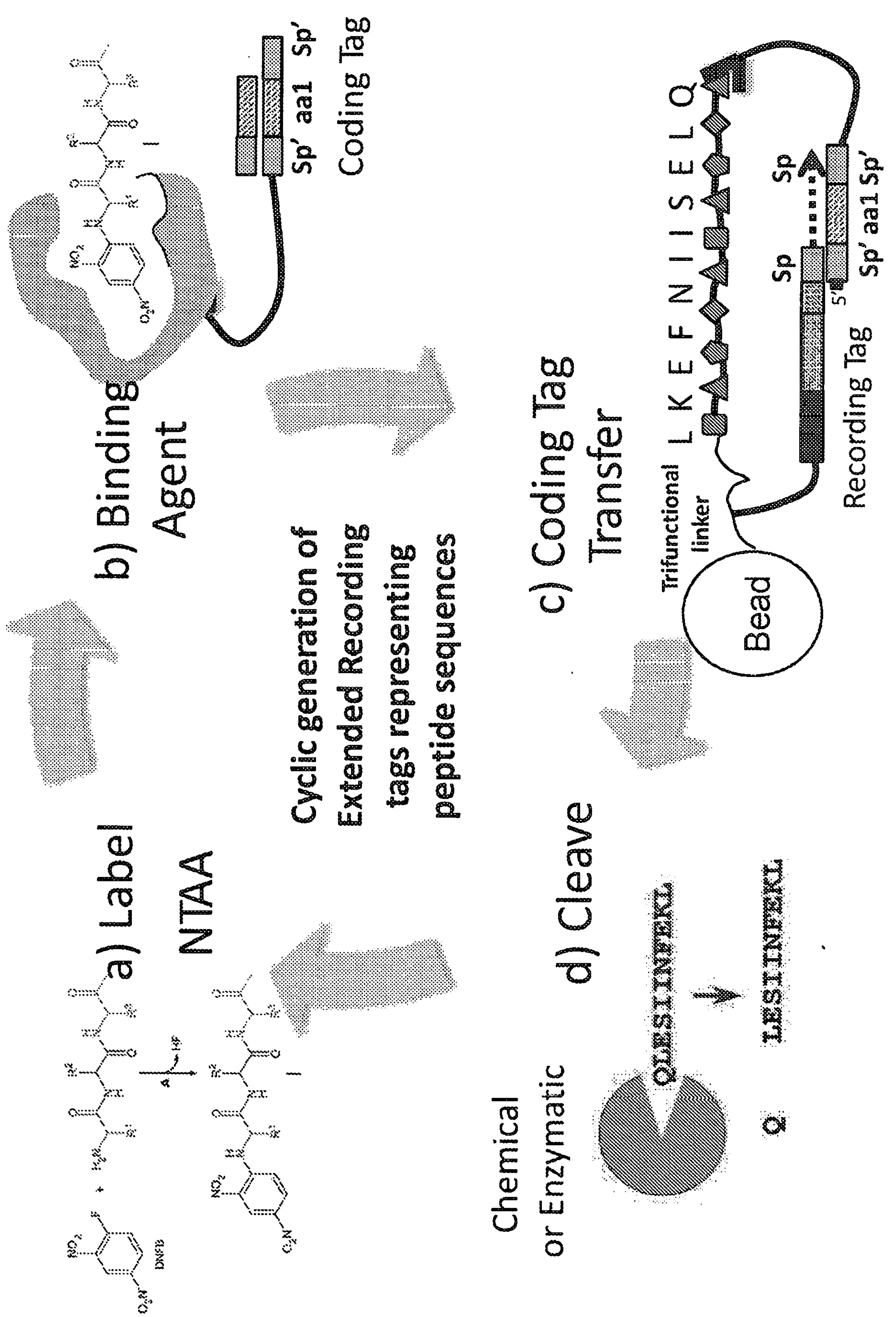

a) Label NTAA
DNFB
b) Binding Agent
Sp' aa1 Sp'
Coding Tag
Cyclic generation of Extended Recording tags representing peptide sequences
c) Coding Tag Transfer
L K E F N I I S E L Q
Trifunctional linker
Bead
Sp
Sp
Sp' aa1 Sp'
5'
Recording Tag
d) Cleave
Chemical or Enzymatic
QLESIINFEKL
LESIINFEKL
Q

Fig. 3

REPLACEMENT SHEET
8/60

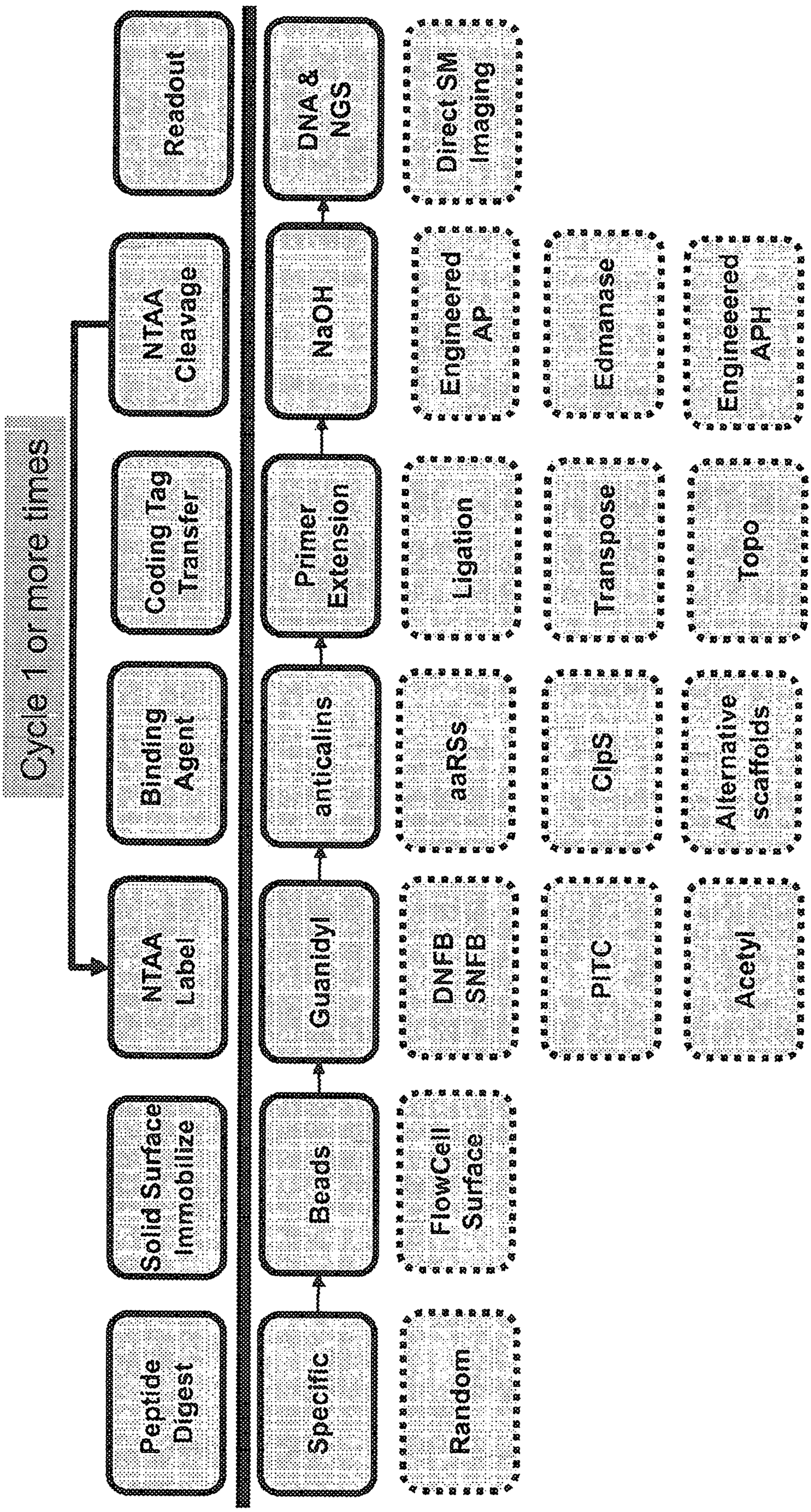


Fig. 4A

REPLACEMENT SHEET
9/60

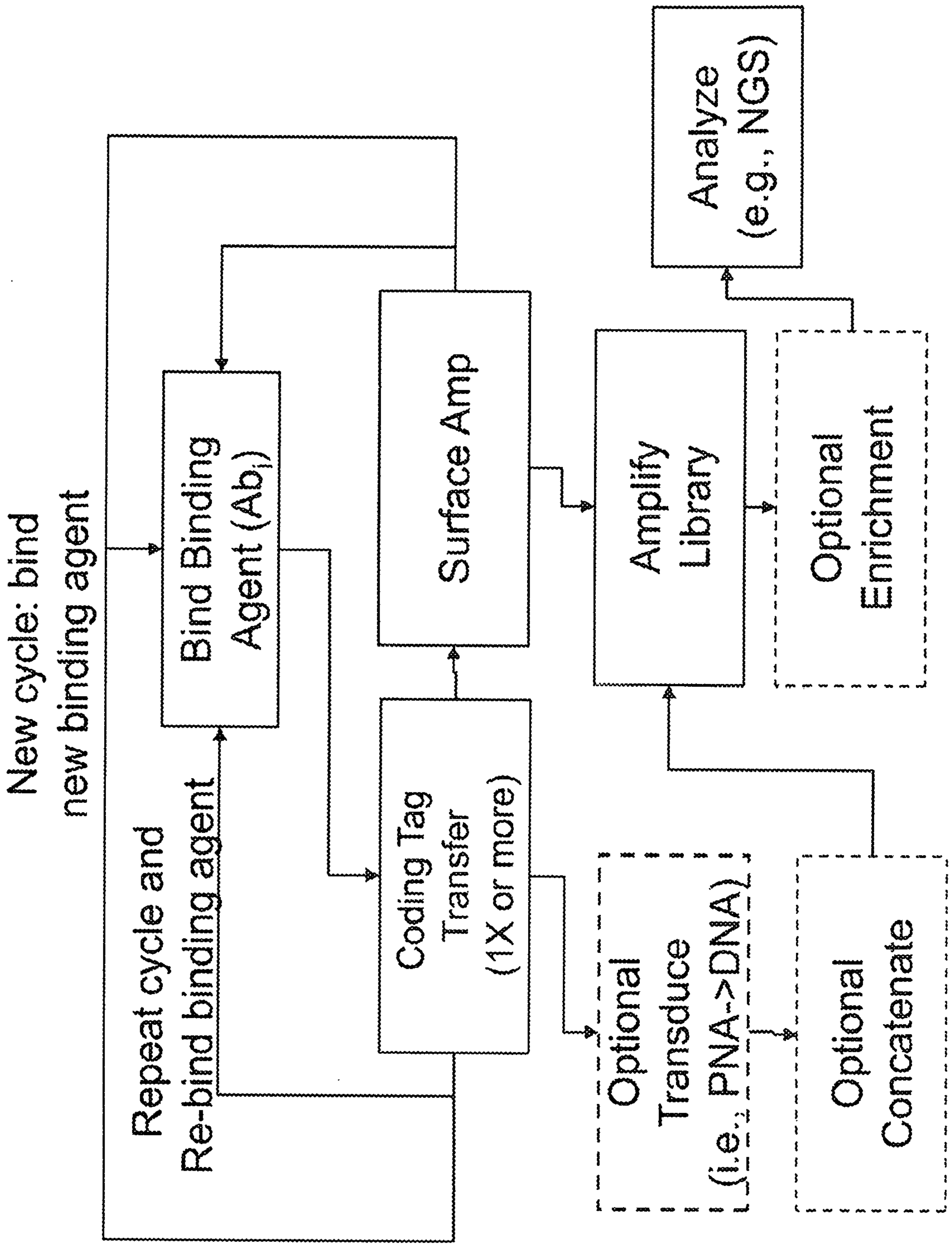


Fig. 4B

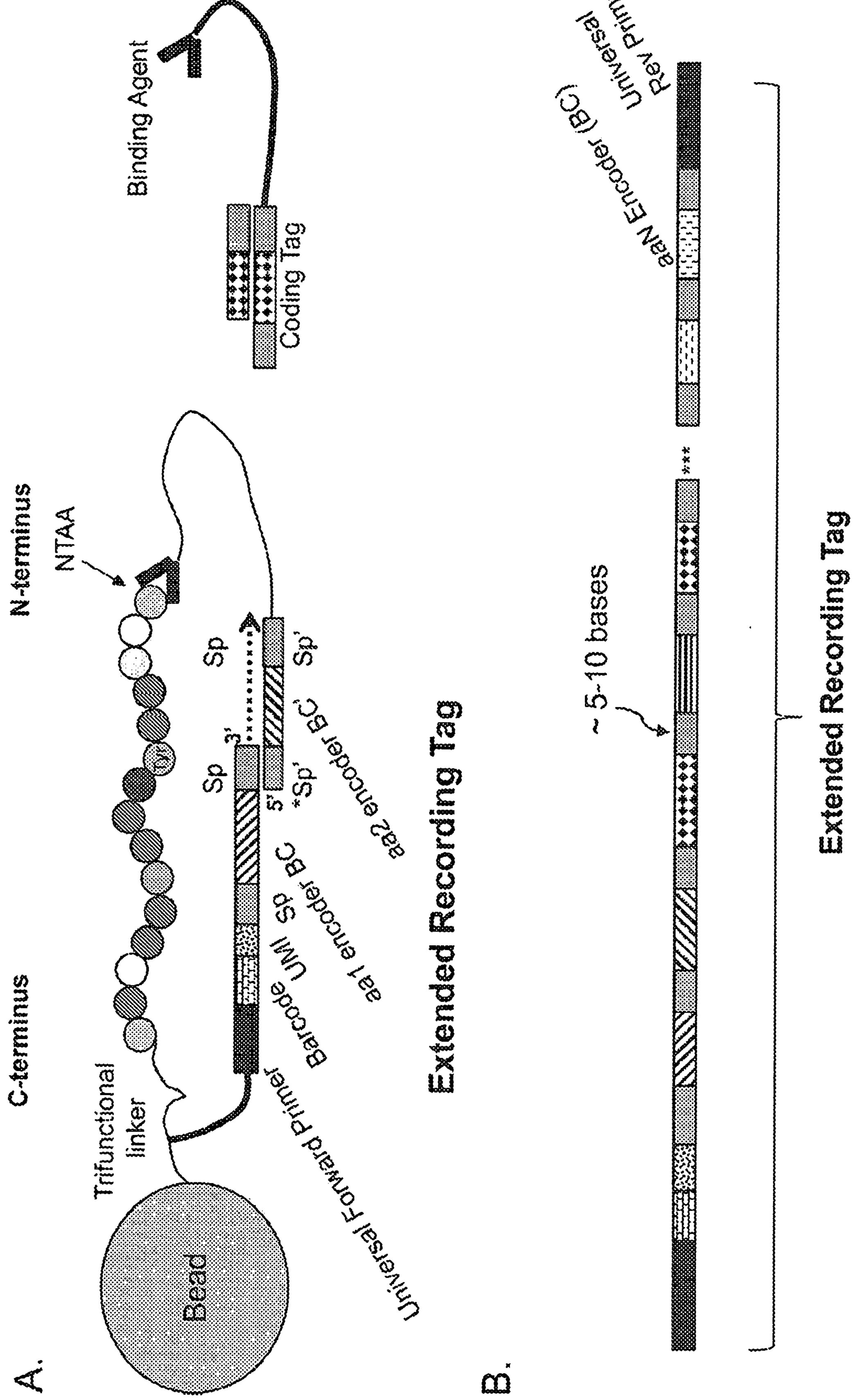

A.
B.
Binding Agent
Coding Tag
N-terminus
NTAA
C-terminus
Tyr
Trifunctional linker
Bead
Sp
Sp
3'
5'
*Sp'
Sp'
Barcode
UMI
Sp
BC
BC'
aa1 encoder
aa2 encoder
Universal Forward Primer
Extended Recording Tag
Universal Rev Primer
aaN Encoder (BC)
~5-10 bases
***
Extended Recording Tag

Fig. 5

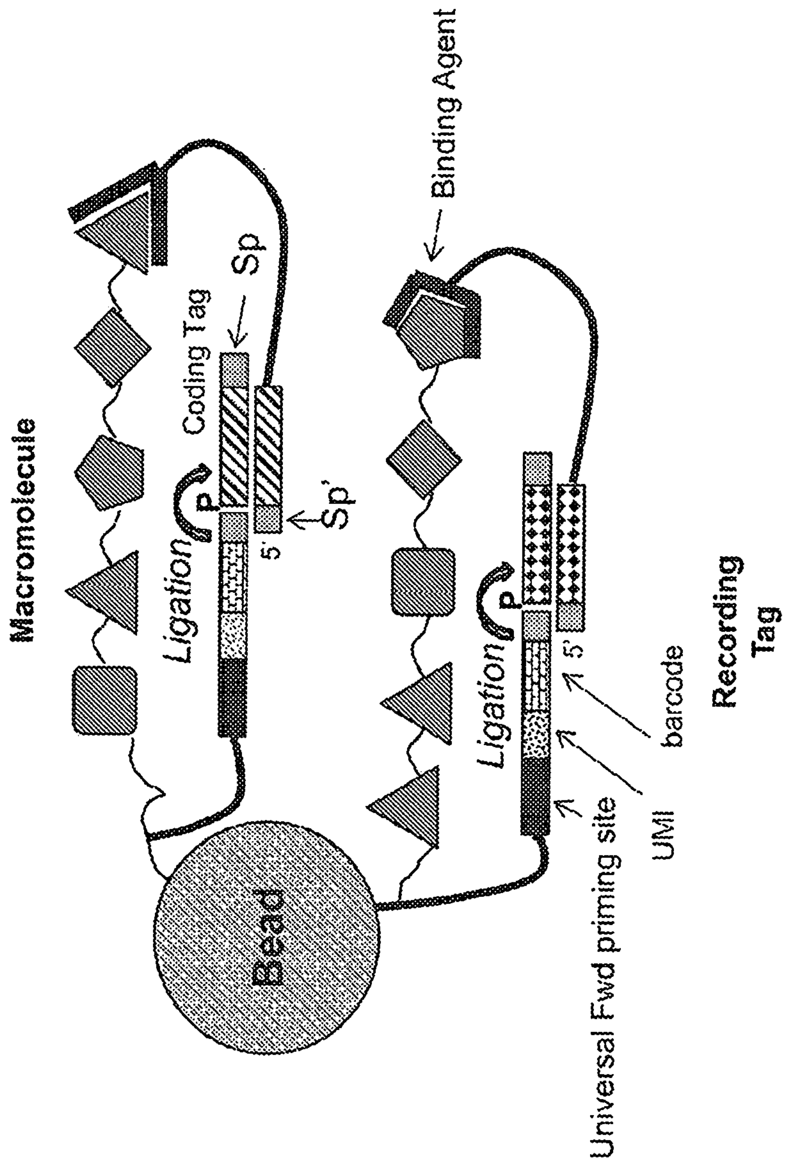

Macromolecule
Binding Agent
Sp
Coding Tag
Ligation
P
Sp'
5'
Bead
Ligation
P
5'
barcode
UMI
Universal Fwd priming site
Recording Tag

Fig. 6

REPLACEMENT SHEET
12/60

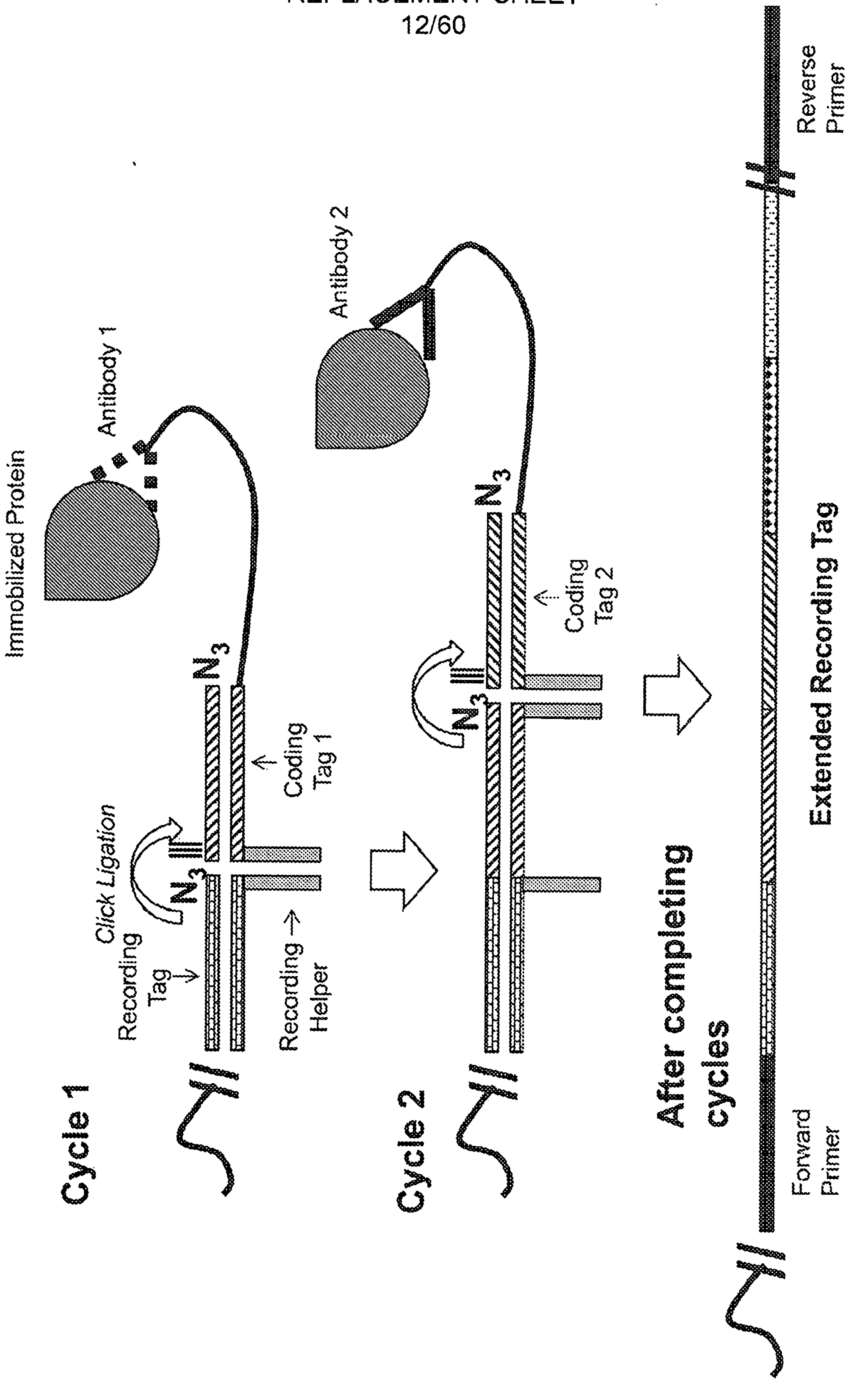


Fig. 7

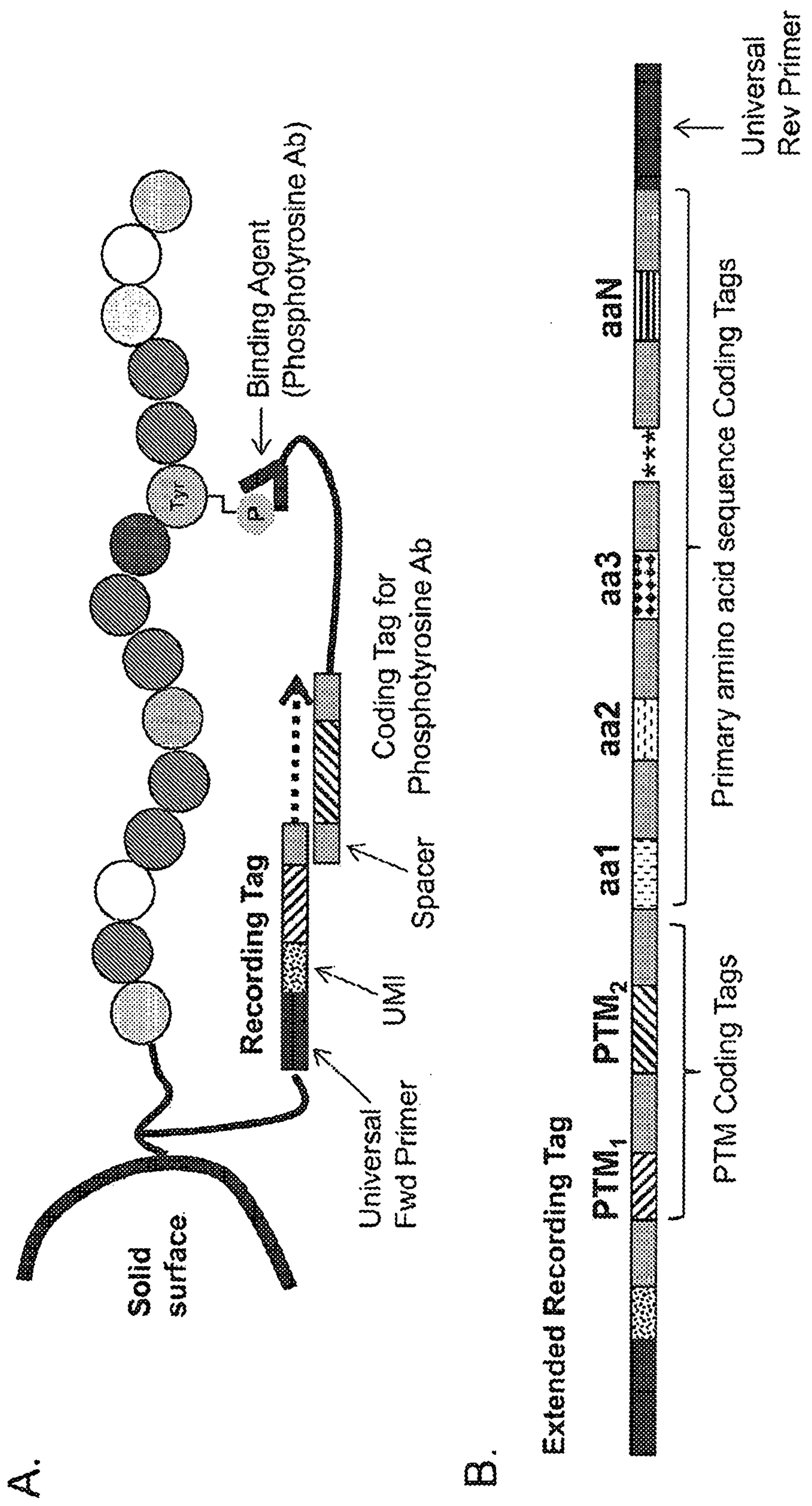

A.
Solid surface
Universal Fwd Primer
UMI
Recording Tag
Spacer
Coding Tag for Phosphotyrosine Ab
Binding Agent (Phosphotyrosine Ab)
Tyr
P
B.
Extended Recording Tag
PTM1
PTM2
aa1
aa2
aa3
aaN
PTM Coding Tags
Primary amino acid sequence Coding Tags
Universal Rev Primer

Fig. 8

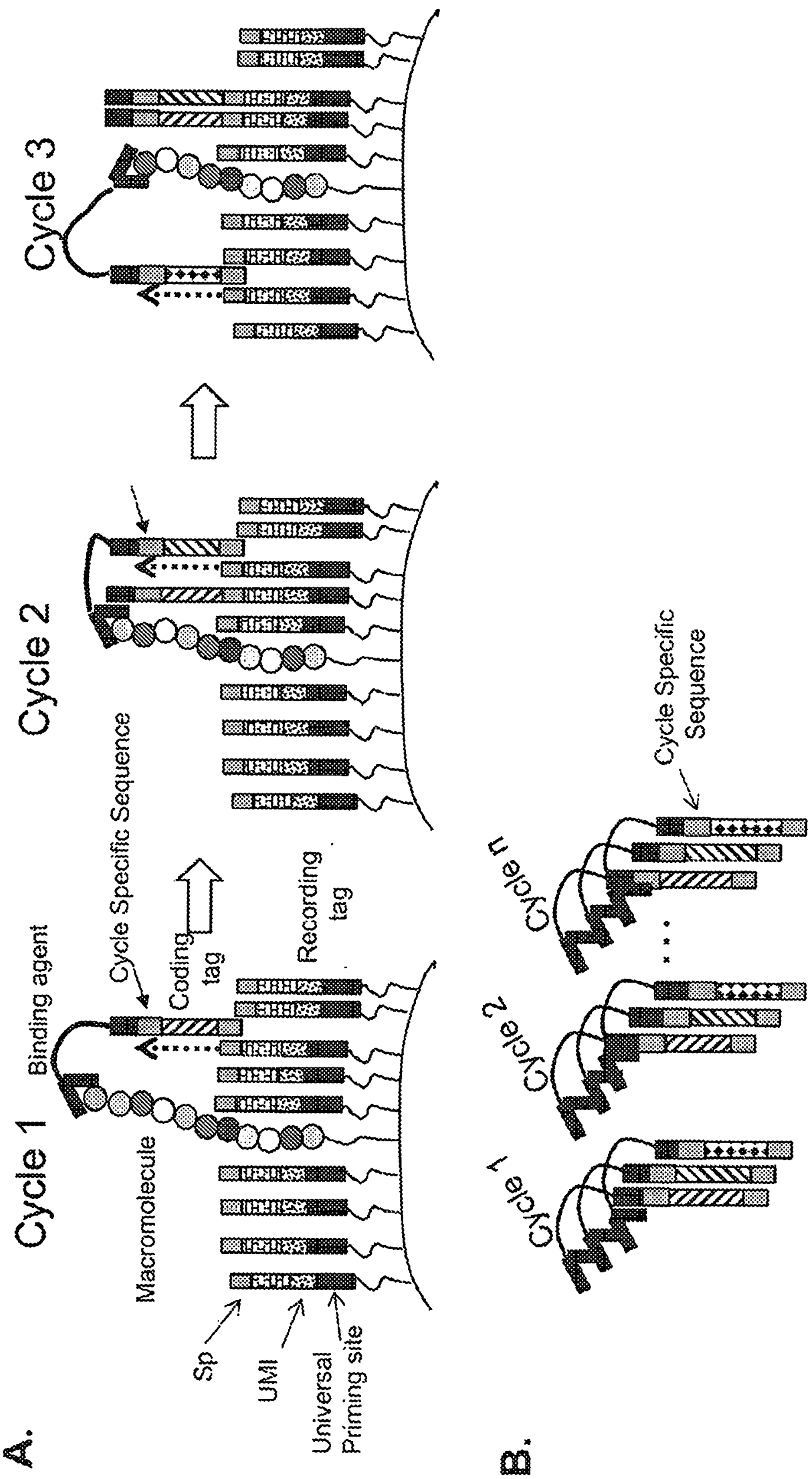

A.
Cycle 1
Cycle 2
Cycle 3
Binding agent
Cycle Specific Sequence
Coding tag
Recording tag
Macromolecule
Sp
UMI
Universal Priming site
B.
Cycle 1
Cycle 2
Cycle n
Cycle Specific Sequence

Fig. 9

REPLACEMENT SHEET
15/60

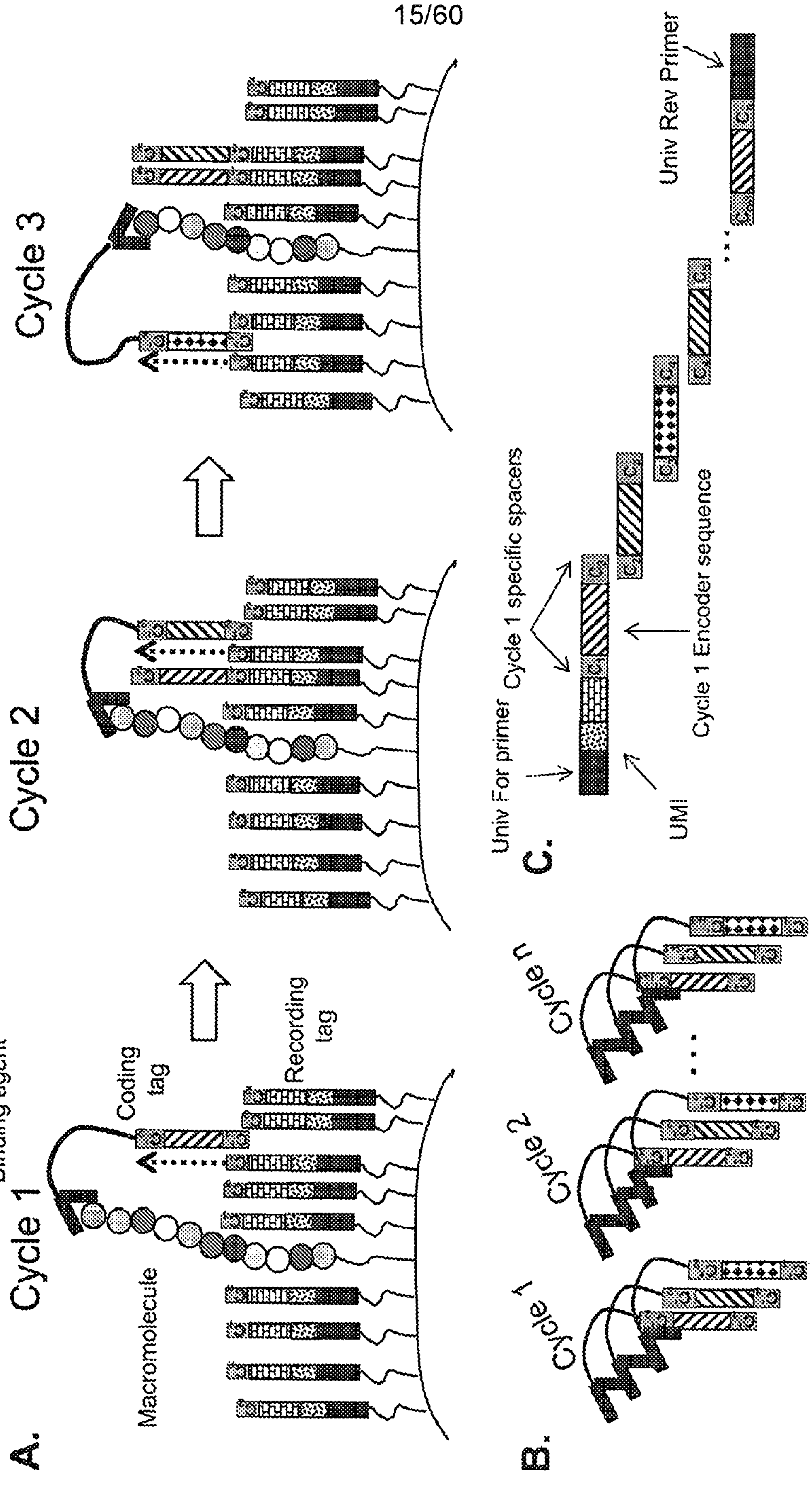


Fig. 10

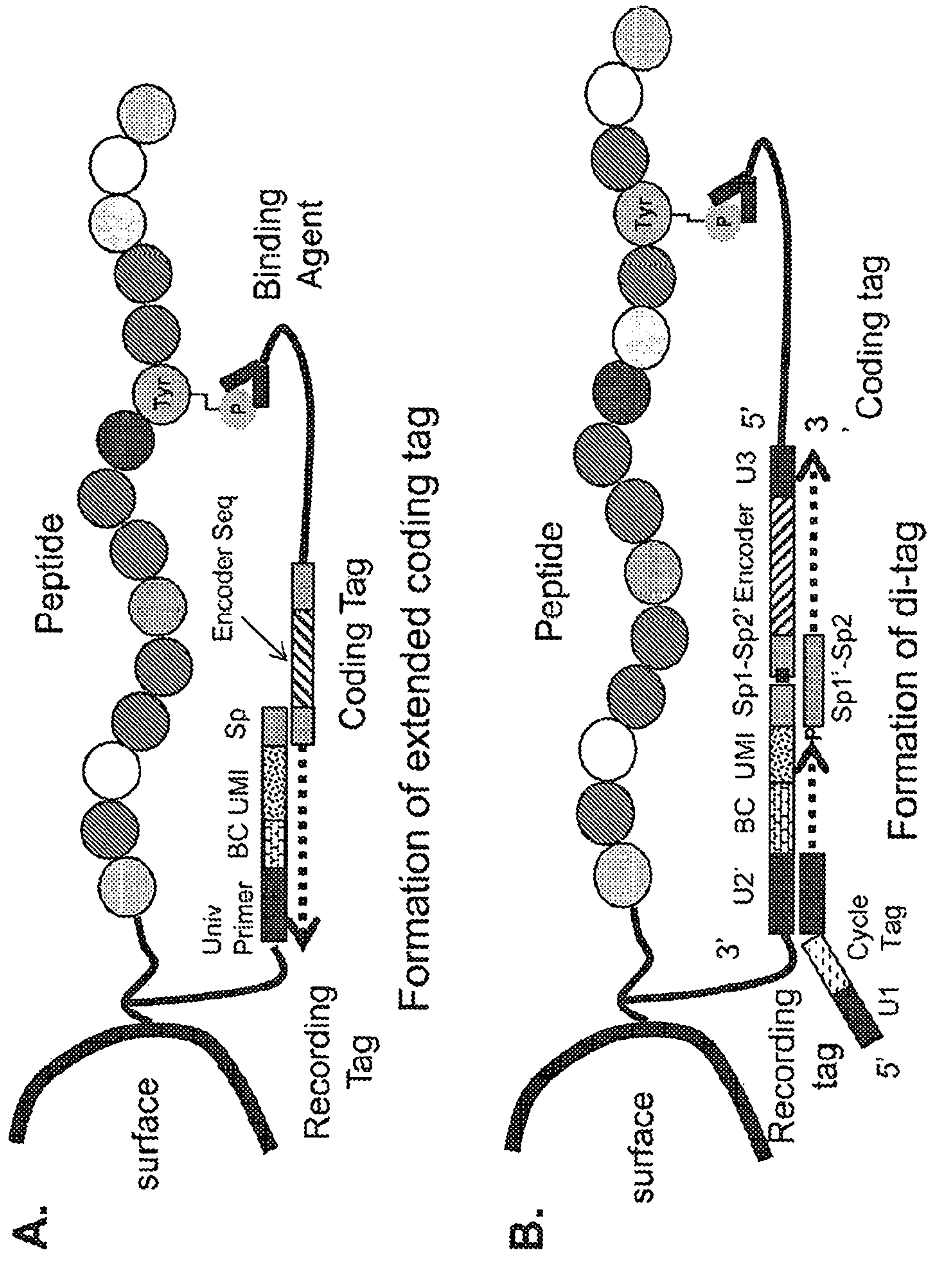

A.
surface
Recording Tag
Univ Primer
BC UMI
Sp
Encoder Seq
Coding Tag
Peptide
Tyr
P
Binding Agent
Formation of extended coding tag
B.
surface
Recording tag
5'
U1
Cycle Tag
3'
U2'
BC
UMI
Sp1-Sp2'
Encoder
U3
5'
3'
Sp1'-Sp2
Peptide
Tyr
P
Coding tag
Formation of di-tag

Fig. 11

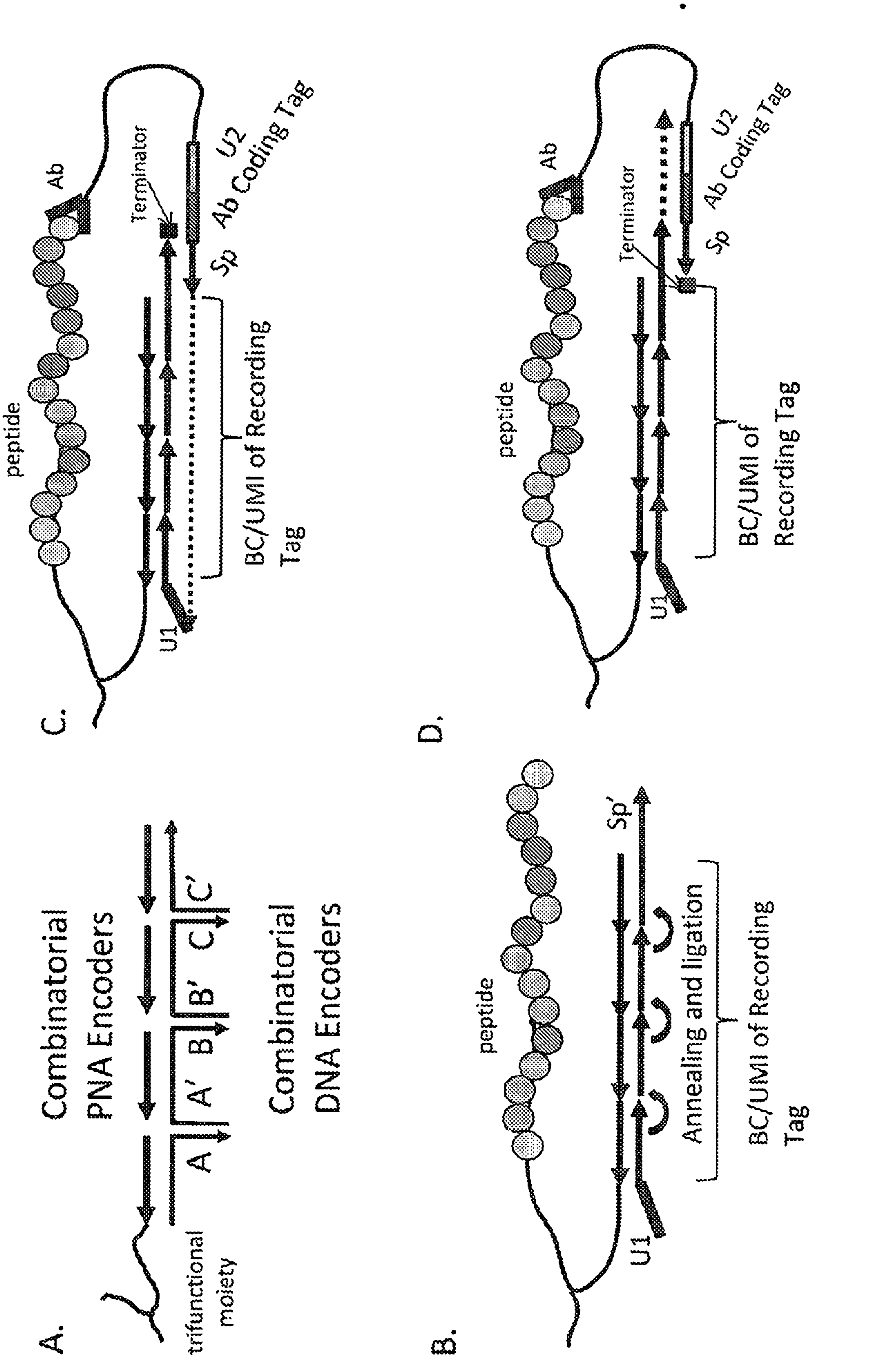

A.
Combinatorial
PNA Encoders
trifunctional
moiety
A
A'
B
B'
C
C'
Combinatorial
DNA Encoders
B.
peptide
U1
Sp'
Annealing and ligation
BC/UMI of Recording
Tag
C.
peptide
Ab
Terminator
U2
Ab Coding Tag
Sp
U1
BC/UMI of Recording
Tag
D.
peptide
Ab
Terminator
U2
Ab Coding Tag
Sp
U1
BC/UMI of
Recording Tag

Fig. 12

REPLACEMENT SHEET
18/60

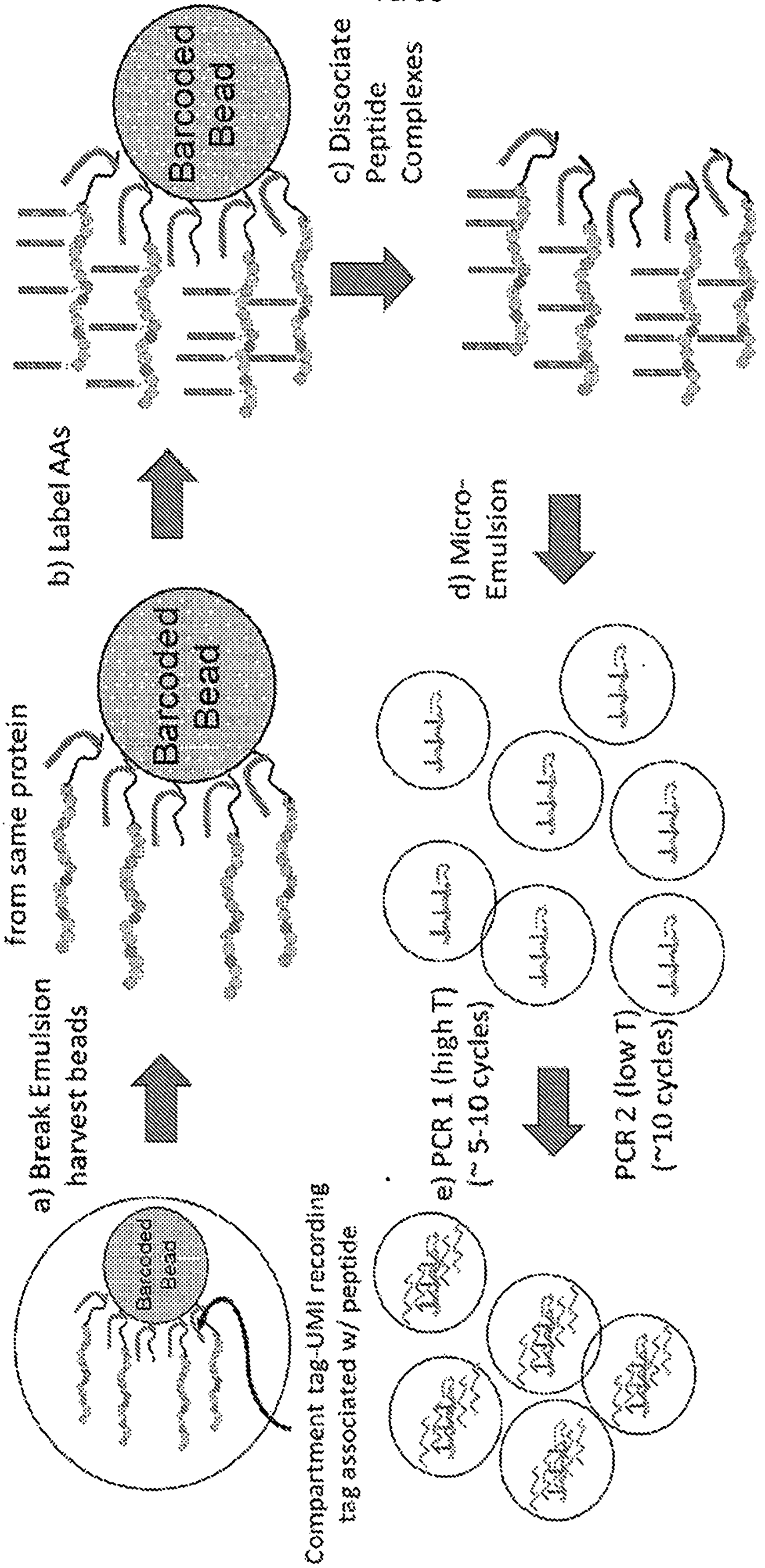


Fig. 13

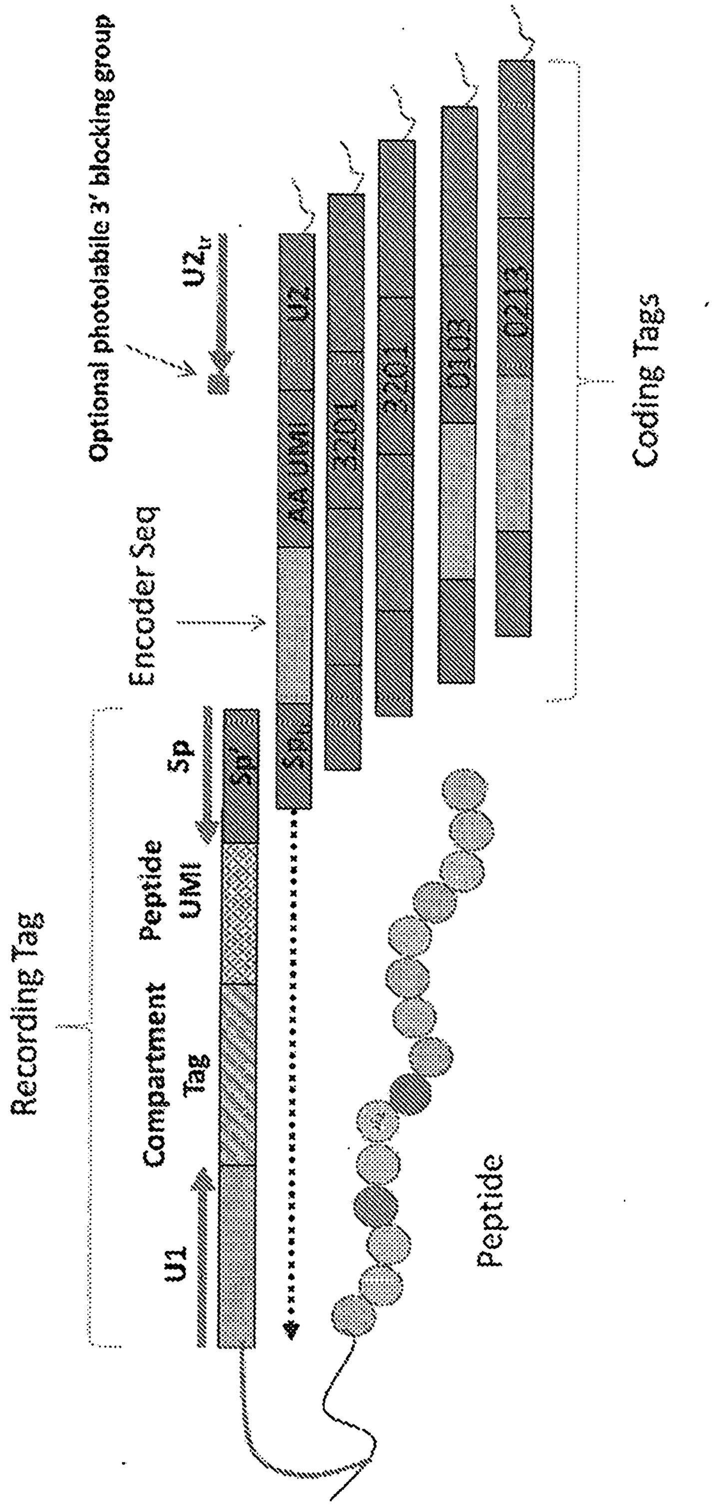

Optional photolabile 3' blocking group
Encoder Seq
Coding Tags
Recording Tag
Compartment Tag
Peptide UMI
Peptide
U1
Sp
AA UMI
U2
3201
0103
0213

Fig. 14

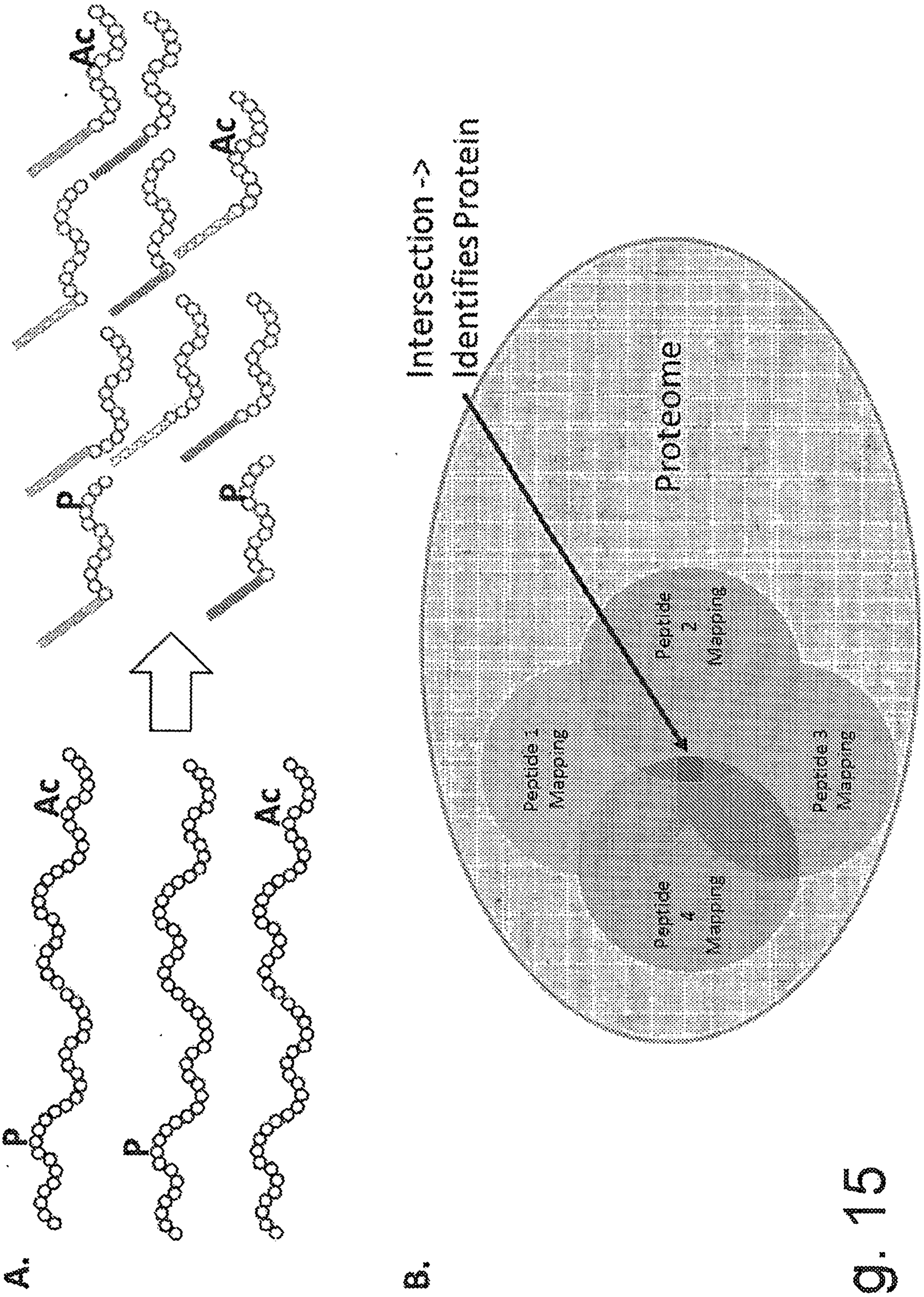

A.
P
Ac
B.
Intersection -> Identifies Protein
Proteome
Peptide 1 Mapping
Peptide 2 Mapping
Peptide 3 Mapping
Peptide 4 Mapping

Fig. 15

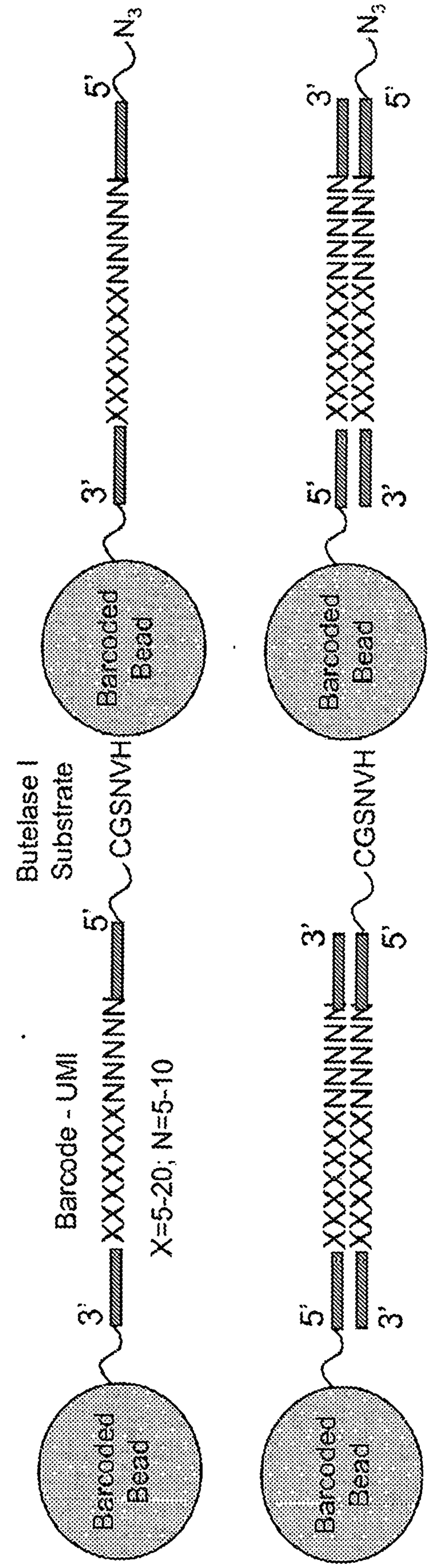

Barcoded Bead
3'
XXXXXXXXNNNNN
5'
Barcode - UMI
X=5-20, N=5-10
CGSNVH
Butelase I Substrate
N3

Fig. 16

REPLACEMENT SHEET
22/60

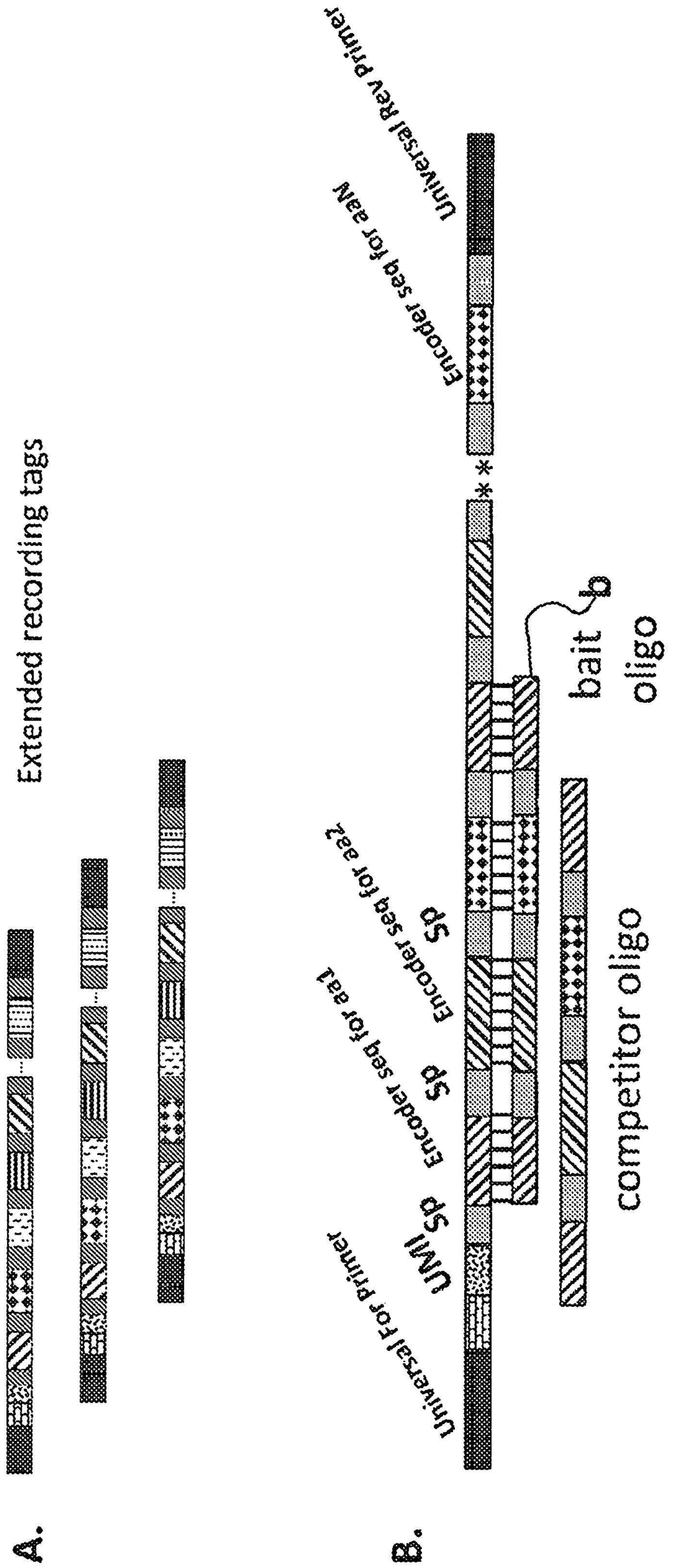


Fig. 17

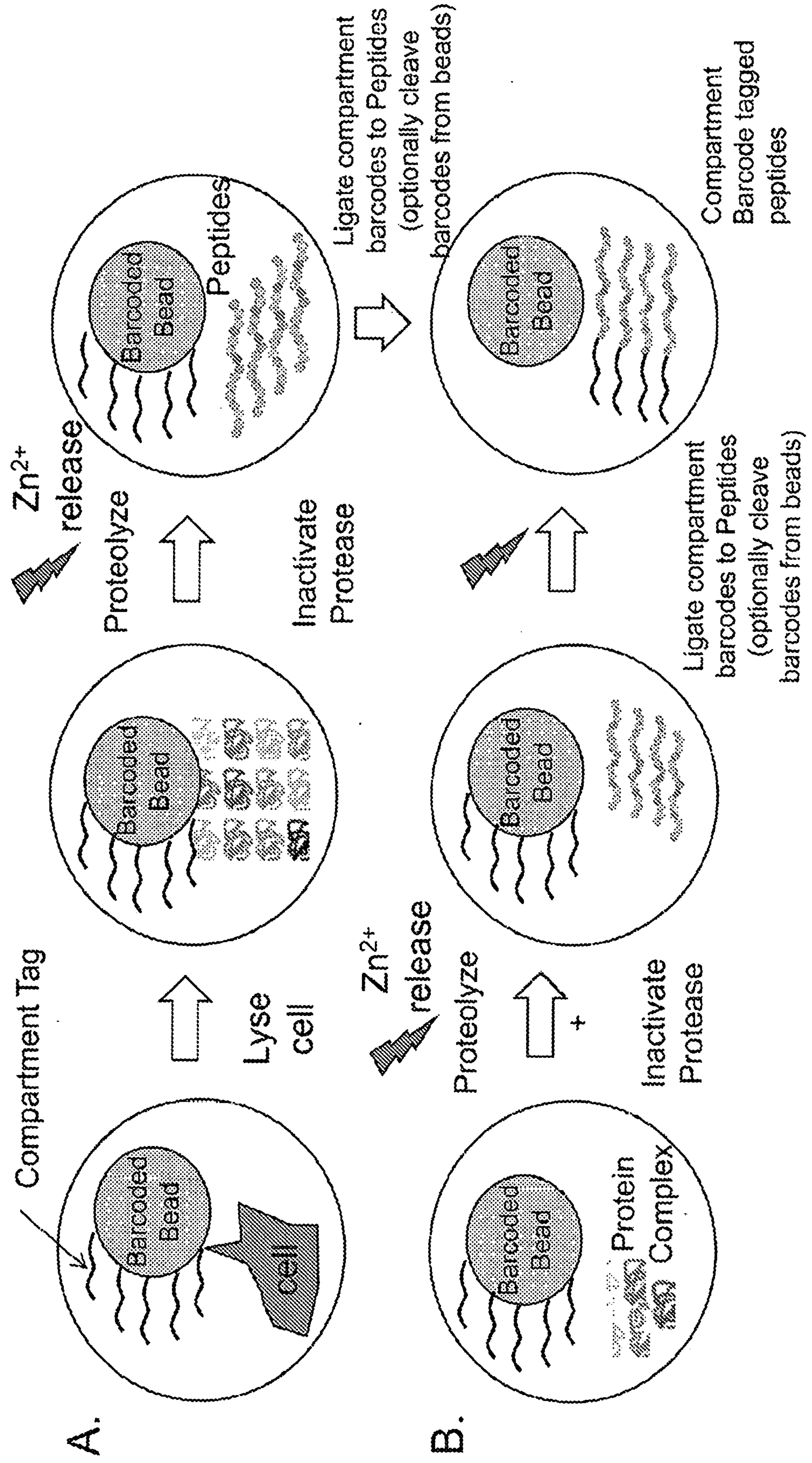

A.
Compartment Tag
Barcoded Bead
Cell
Lyse cell
Zn^{2+} release
Proteolyze
Inactivate Protease
Peptides
Ligate compartment barcodes to Peptides (optionally cleave barcodes from beads)
Compartment Barcode tagged peptides
B.
Protein Complex
+

Fig. 18

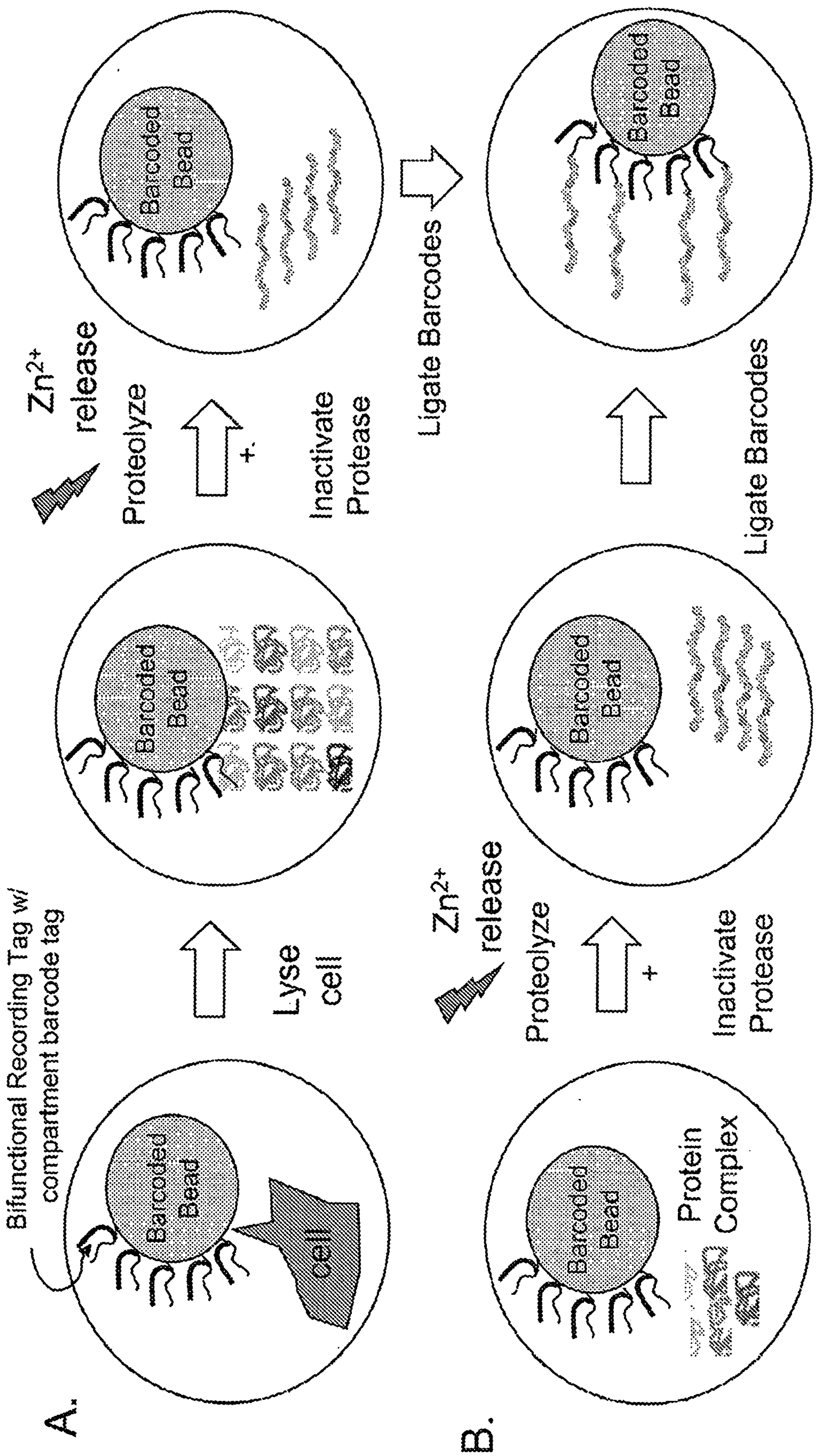

A.
B.
Bifunctional Recording Tag w/ compartment barcode tag
Barcoded Bead
cell
Lyse cell
Zn^{2+} release
Proteolyze
+
Inactivate Protease
Ligate Barcodes
Protein Complex

Fig. 19

REPLACEMENT SHEET
25/60

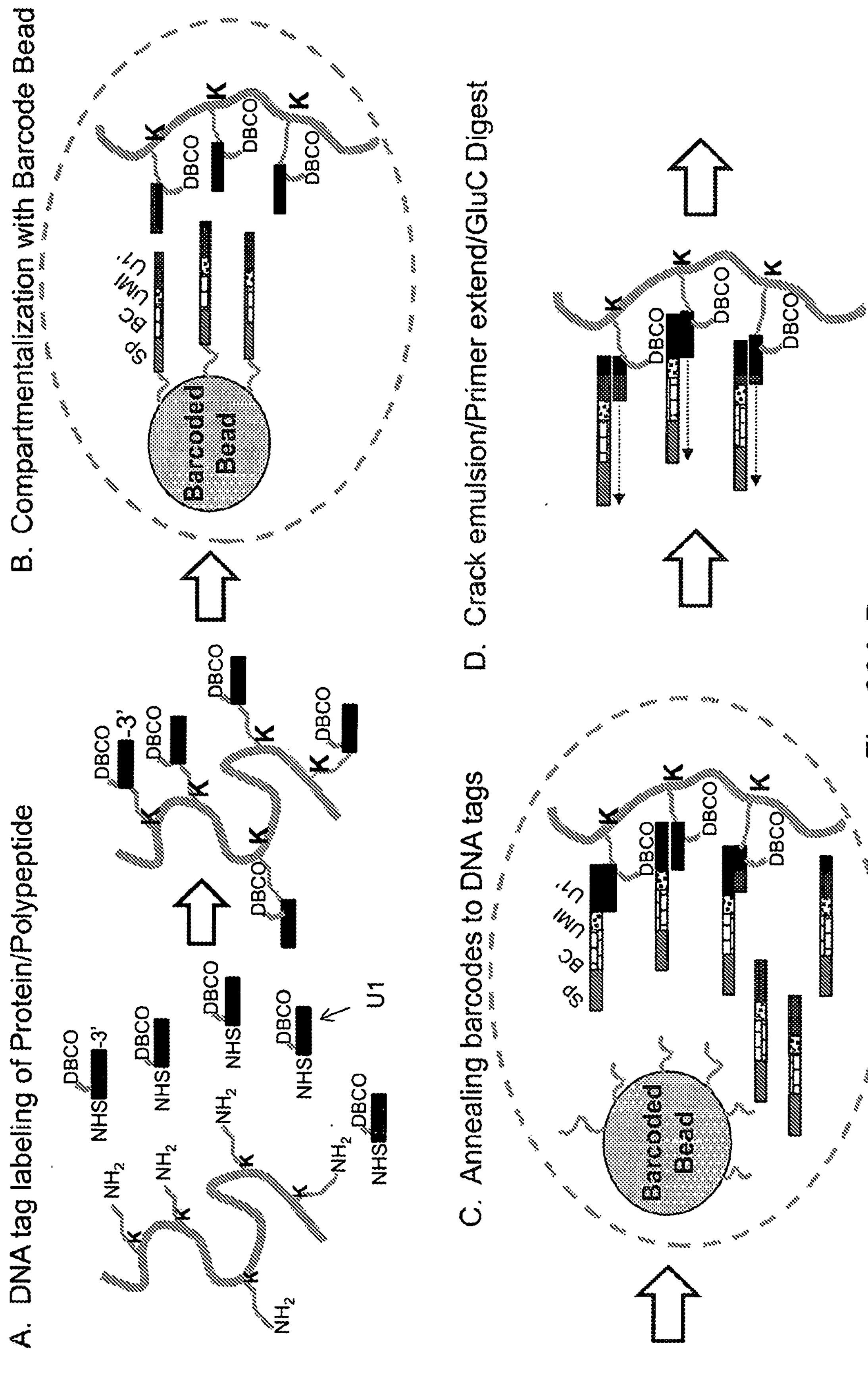


Fig. 20A-D

REPLACEMENT SHEET
26/60

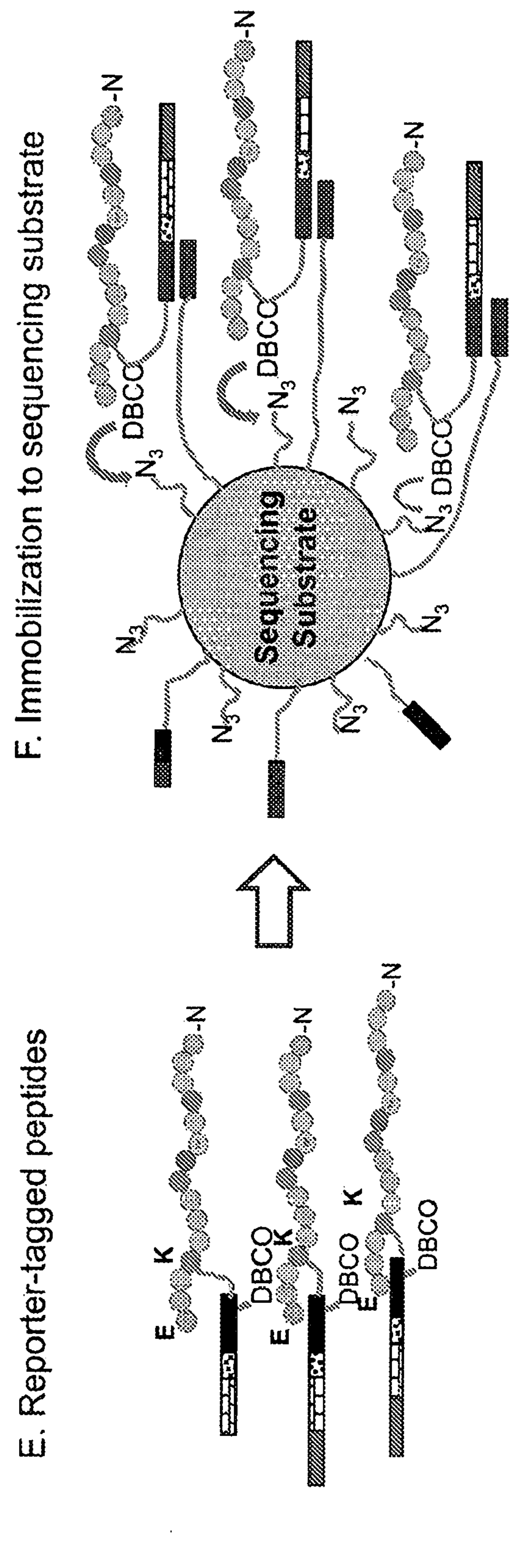


Fig. 20E-F

REPLACEMENT SHEET
27/60

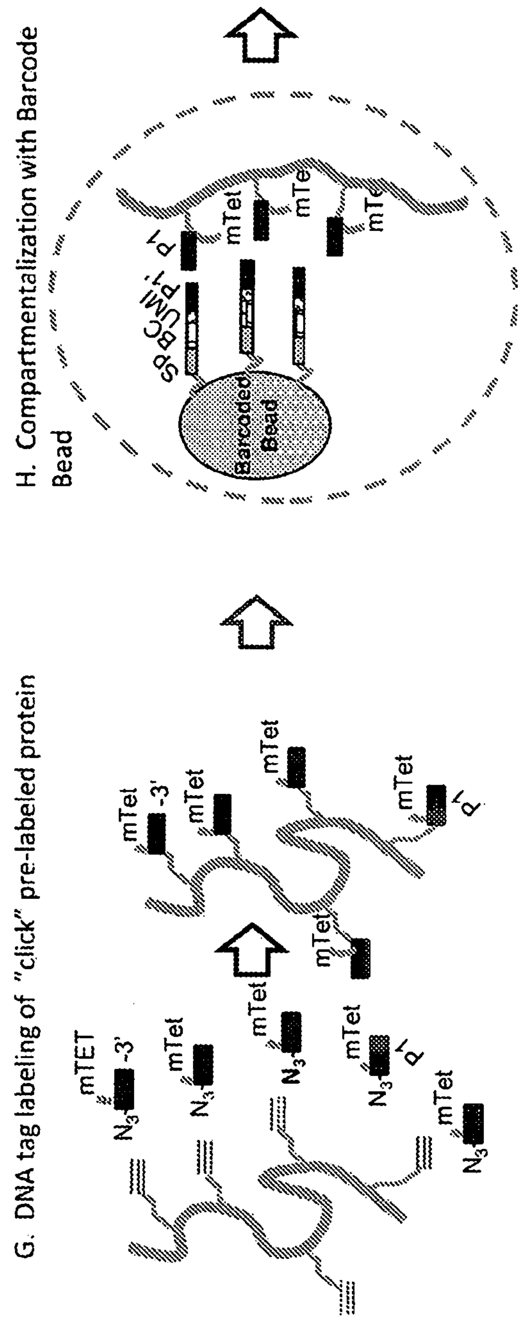


Fig. 20G-H

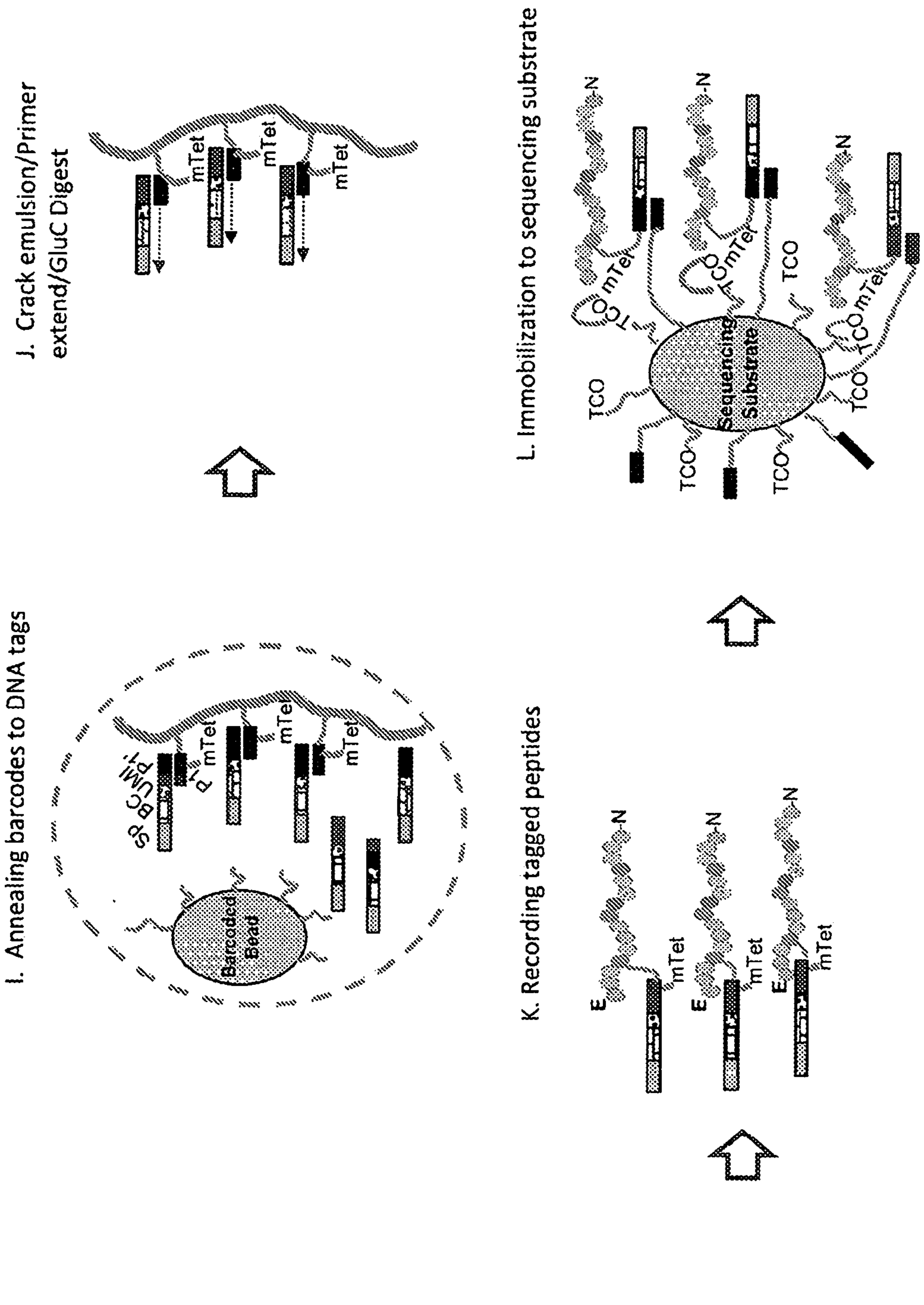

I. Annealing barcodes to DNA tags
Barcoded Bead
SP
BC
UMI
mTet
J. Crack emulsion/Primer extend/GluC Digest
mTet
K. Recording tagged peptides
E
N
mTet
L. Immobilization to sequencing substrate
Sequencing Substrate
TCO
mTet
N

Fig. 20I-L

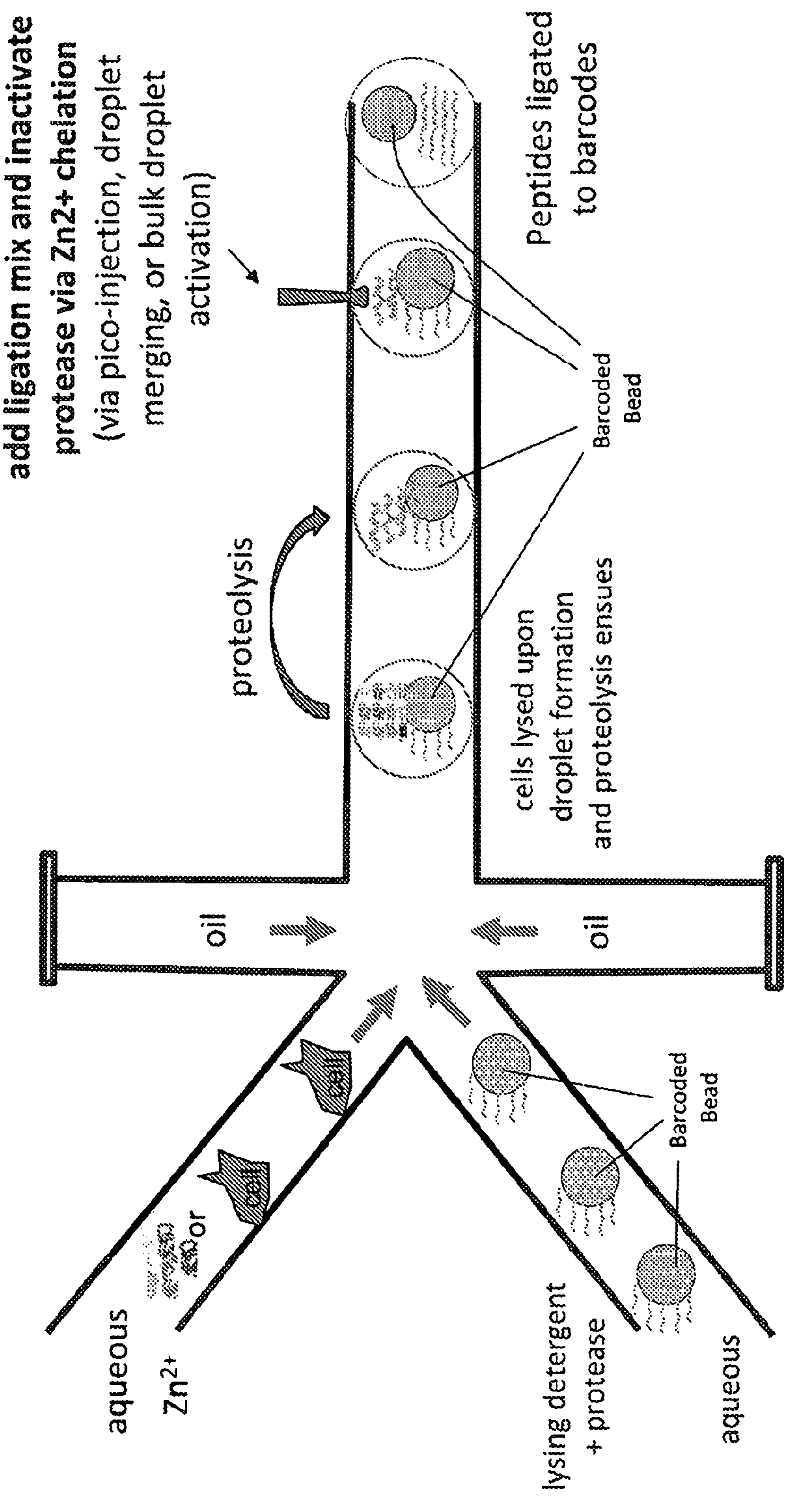

add ligation mix and inactivate protease via Zn2+ chelation
(via pico-injection, droplet merging, or bulk droplet activation)
proteolysis
Peptides ligated to barcodes
Barcoded Bead
cells lysed upon droplet formation and proteolysis ensues
oil
oil
cell
cell
or
aqueous
Zn2+
Barcoded Bead
lysing detergent + protease
aqueous

Fig. 21

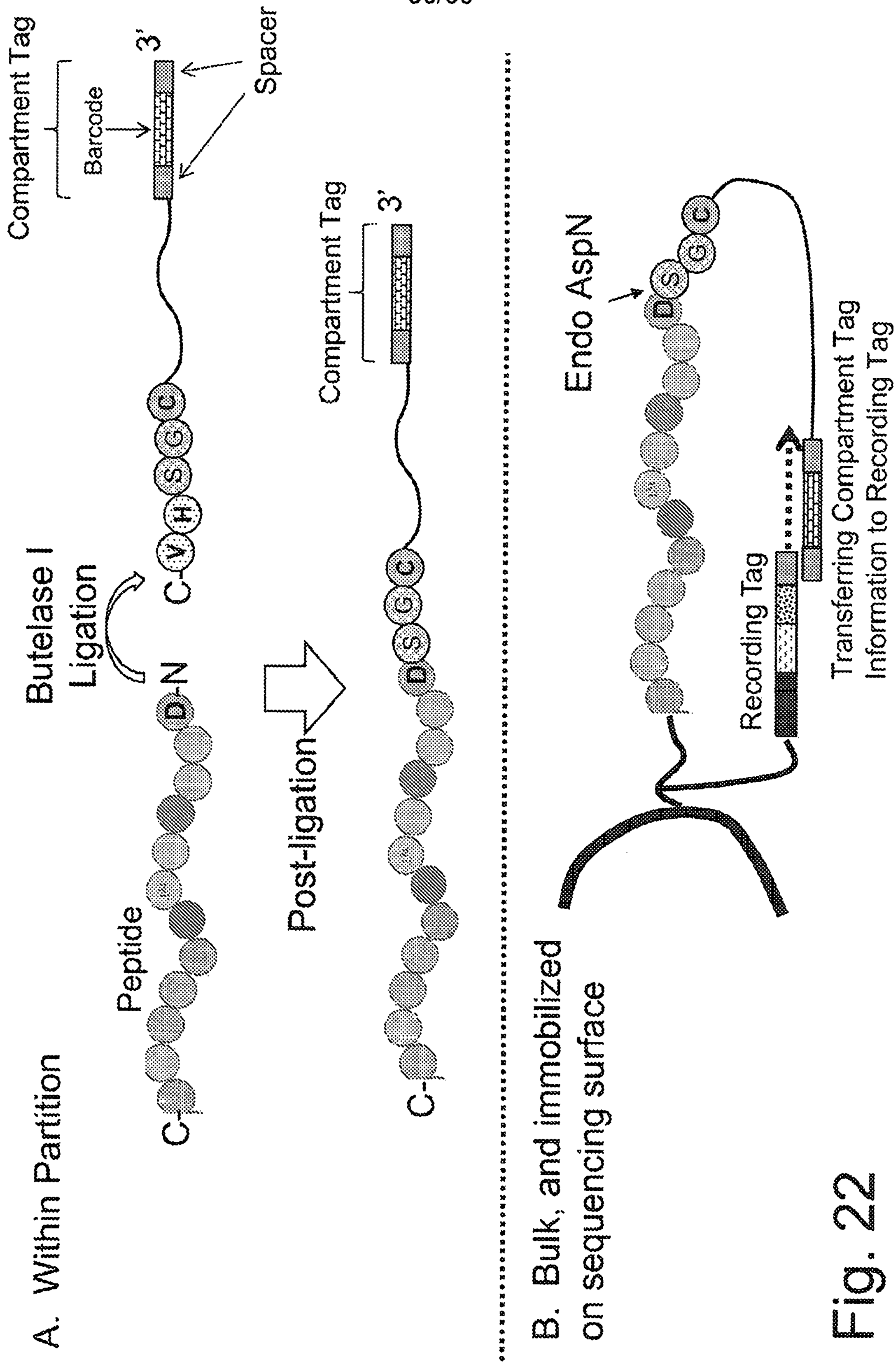

A. Within Partition
Peptide
C-
D-N
Butelase I Ligation
C-
V
H
S
G
C
Compartment Tag
Barcode
Spacer
3'
Post-ligation
C-
D
S
G
C
Compartment Tag
3'
B. Bulk, and immobilized on sequencing surface
Endo AspN
D
S
G
C
Recording Tag
Transferring Compartment Tag Information to Recording Tag

Fig. 22

REPLACEMENT SHEET
31/60

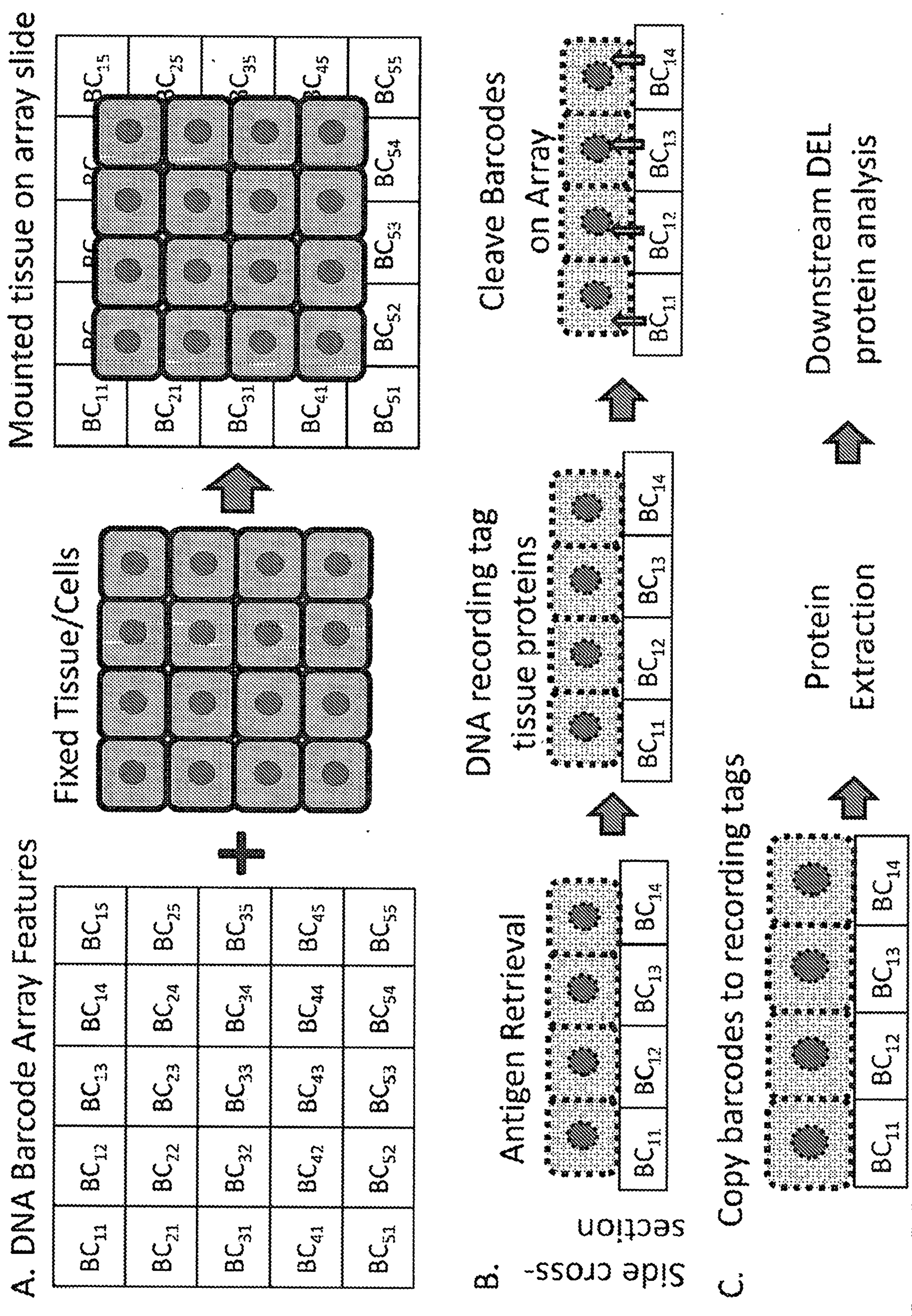


Fig. 23

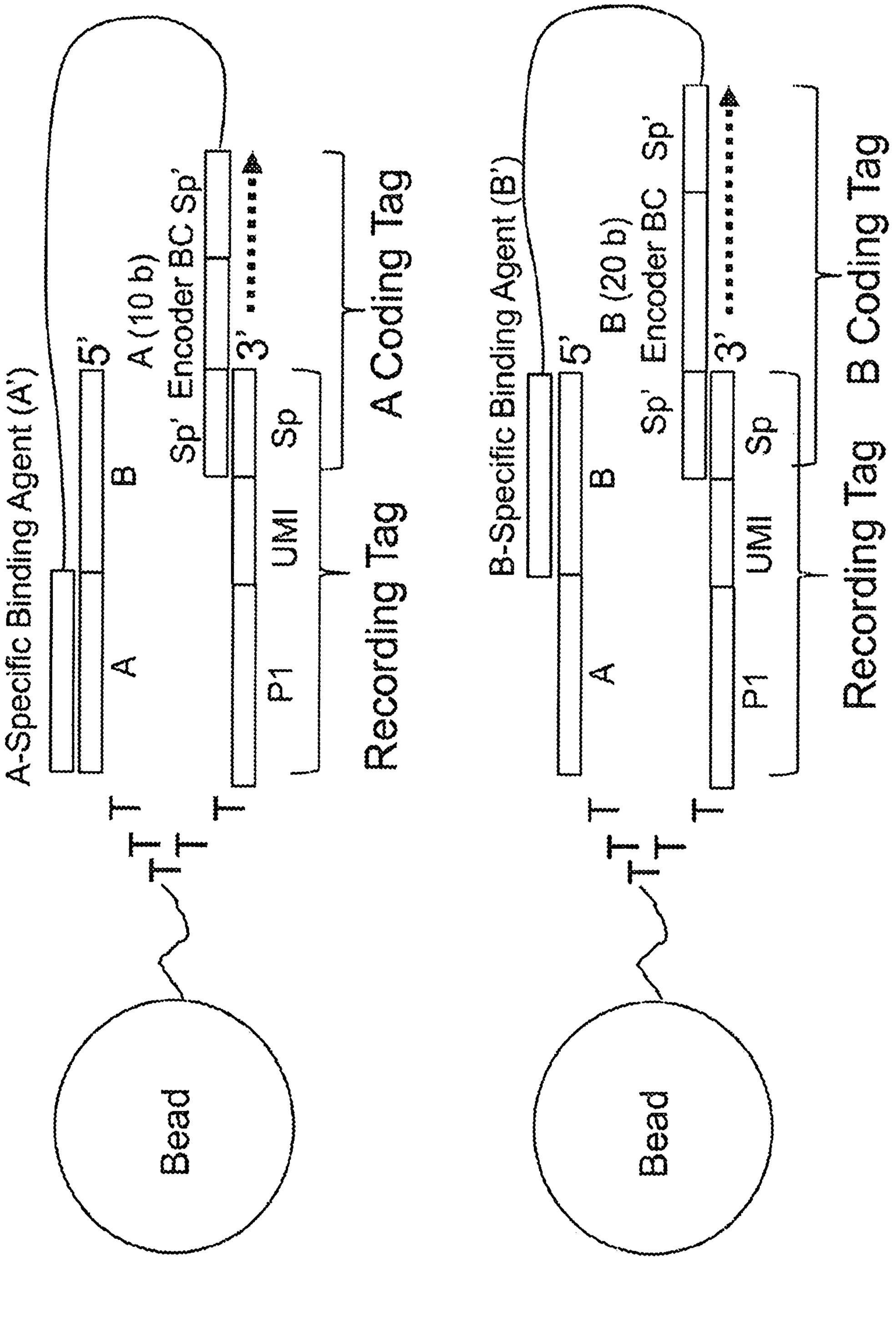

A-Specific Binding Agent (A')
A
B
5'
A (10 b)
Sp' Encoder BC Sp'
3'
P1
UMI
Sp
Recording Tag
A Coding Tag
Bead
B-Specific Binding Agent (B')
A
B
5'
B (20 b)
Sp' Encoder BC Sp'
3'
P1
UMI
Sp
Recording Tag
B Coding Tag
Bead

Fig. 24A

REPLACEMENT SHEET
33/60

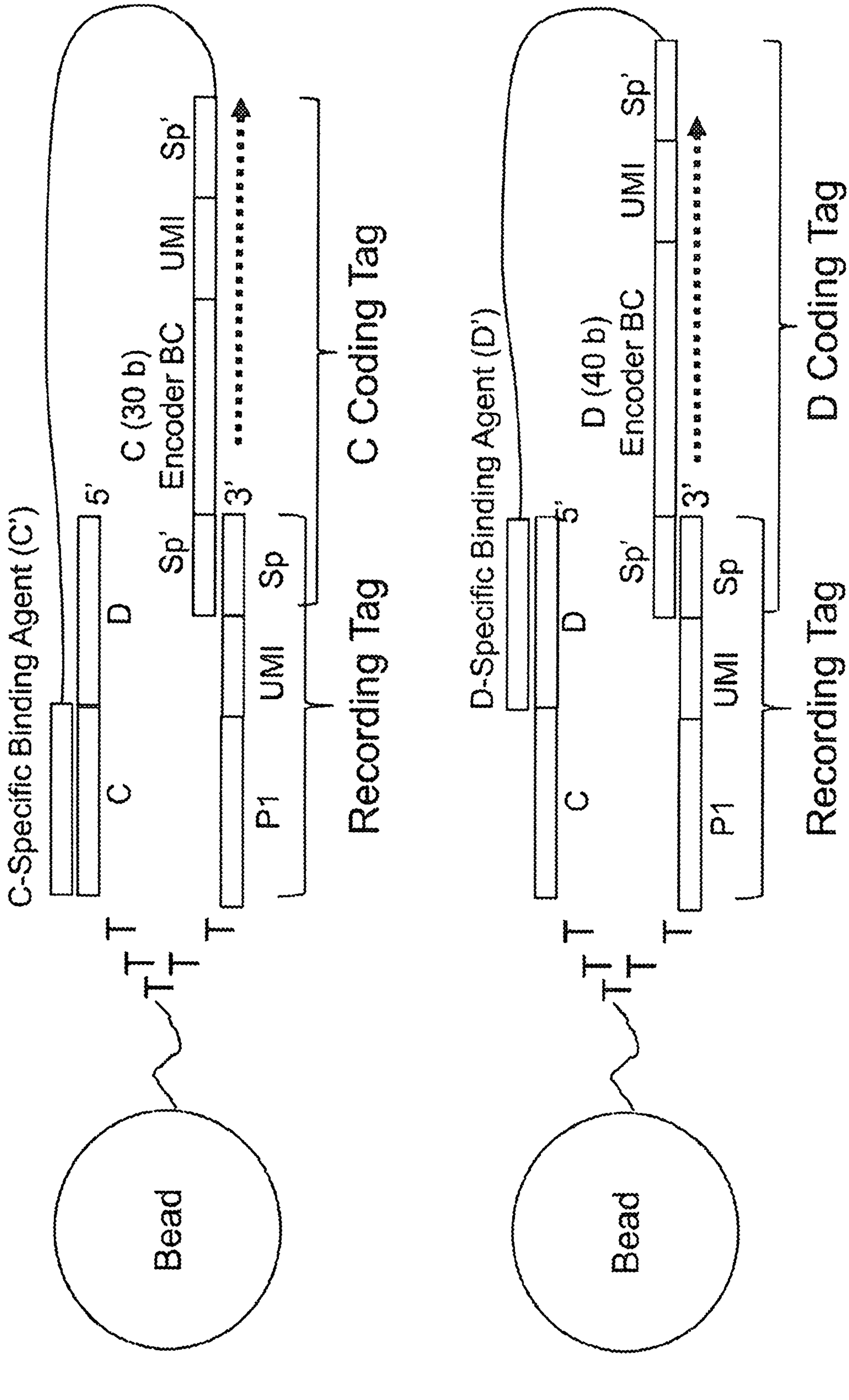


Fig. 24B

PCT/US17/30702 01 March 2018 (01.03.2018)

PCT/US2017/030702 16.05.2018

REPLACEMENT SHEET

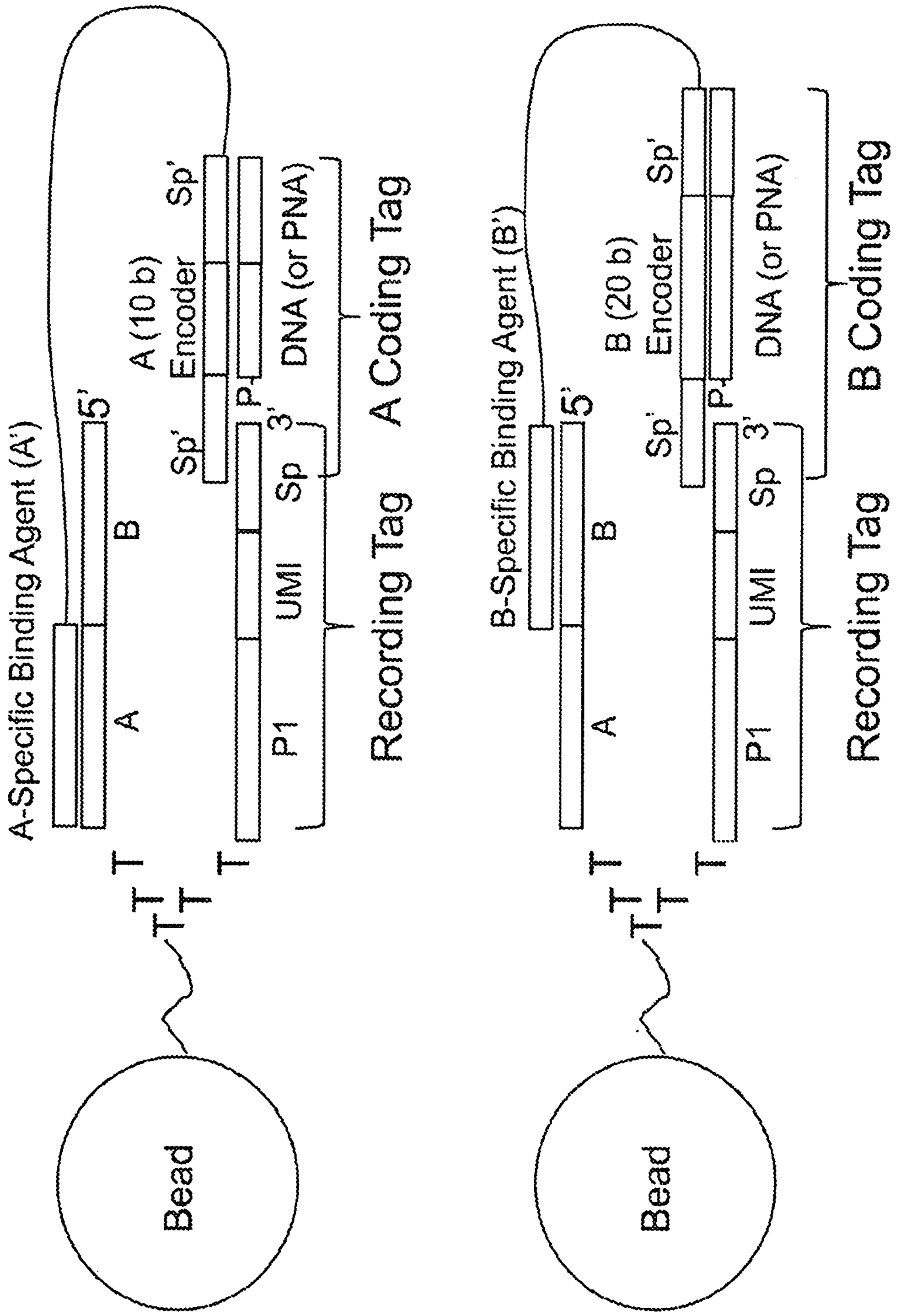


Fig. 25

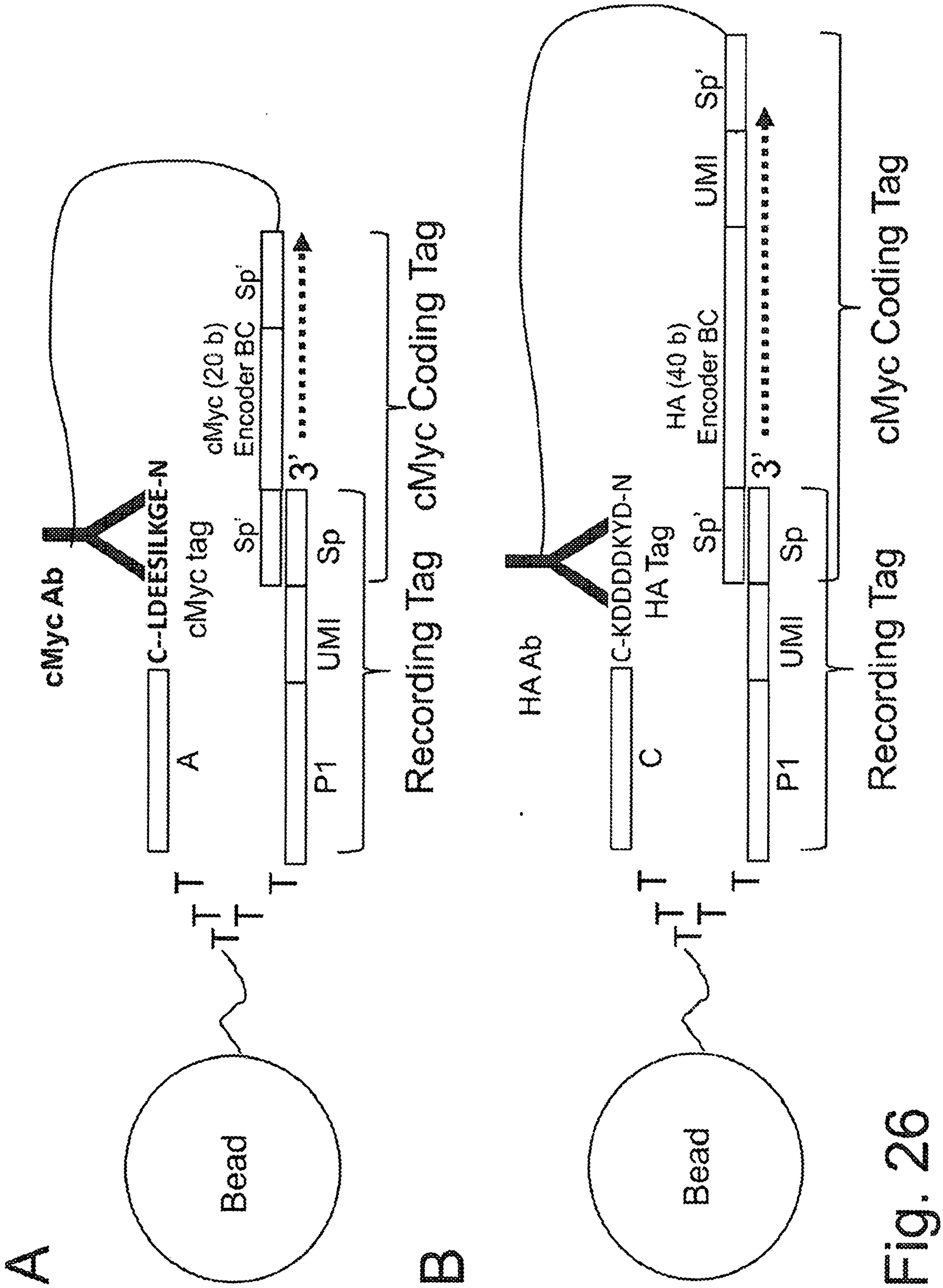

A
Bead
cMyc Ab
C--LDEESILKGE-N
cMyc tag
A
cMyc (20 b)
Encoder BC
Sp'
Sp'
3'
P1
UMI
Sp
Recording Tag
cMyc Coding Tag
B
Bead
HA Ab
C-KDDDDKYD-N
HA Tag
C
HA (40 b)
Encoder BC
UMI
Sp'
Sp'
3'
P1
UMI
Sp
Recording Tag
cMyc Coding Tag

Fig. 26

A.

ATGTCTAGCATGCCG	CCGCTTGTGATCTGG	CTGGATGGTTGTCGA	GATTCAACGGAGAAG	GCCATCAGTAAGAGA	GTCCTCGTCTATCCT	CCTTACAGGTCTGCG	GGCCAACTAAGGTGC
CCGTGTCATGTGGAA	AGATAGCGTACCGGA	ACTGCACGGTTCCAA	ACGAACCTCGCACCA	GCAACGTGAATTGAG	TAGGATTCCGTTACC	GATCATTGGCCAATT	GCACCTATTCGACAA
TAAGCCGGTATATCA	TCGAGGCTCATCATC	CGAGAGATGGTGCTT	AGGACTTCAAGAAGA	CTAAGTAGAGCCACA	TCTGACCACCGGAAG	TTCAAGGCTGAGTTG	TGGACACGATCGGCT
TTCGATATGACGGAA	GAGTACTAGAGCCAA	TCTTGAGAGACAAGA	GGTTGAATCCTCGCA	TGTCTGTTGGAAGCG	AGAGTCACCTCGTGG	TGGCTCGATTGAATC	CTATAATTCCAACGG
CGTATACGGGTTAGG	GAGCGTCAATAACGG	AATTCGCACTGTGTT	AACCAACCTCTAGCG	TTAATAGACAGGCGG	CTGATGTAGTCGAAG	GTAAGCCATCCGCTC	AACGTGGTTAGTAAG
AACTGCCGAGATTCC	GCGGTATCTACACTG	GTAGTGCCGCTAAGA	ACGCGAATATCTAAC	CGACGCTCTAACAAG	GTCGGTTGCGGATAG	ACACATGCGTAGACA	CAAGGAACGAGTGGC
TGATCTTAGCTGTGC	CTTCTCGGAAGAGAA	CCTATAGCACAATCG	GTTGAGAATTACACC	CATGGCTTATTGAGA	TCCTCCTCCTAAGAA	TGCTATGGATTCAAG	CACCAGAACGGAAGA
GAGTCGGTACCTTGA	TGAAGCCTGTGTTAA	ATCACCGAGGTTGGA	CTCTCTCTGTGAACC	ACTAGGTATGGCCGG	ATTCGGTCCACTTCA	CCACGAGGCTTAGTT	CGTACGGTCAAGCAA
							TCGGTGACAGGCTAA

Fig. 27A

B.

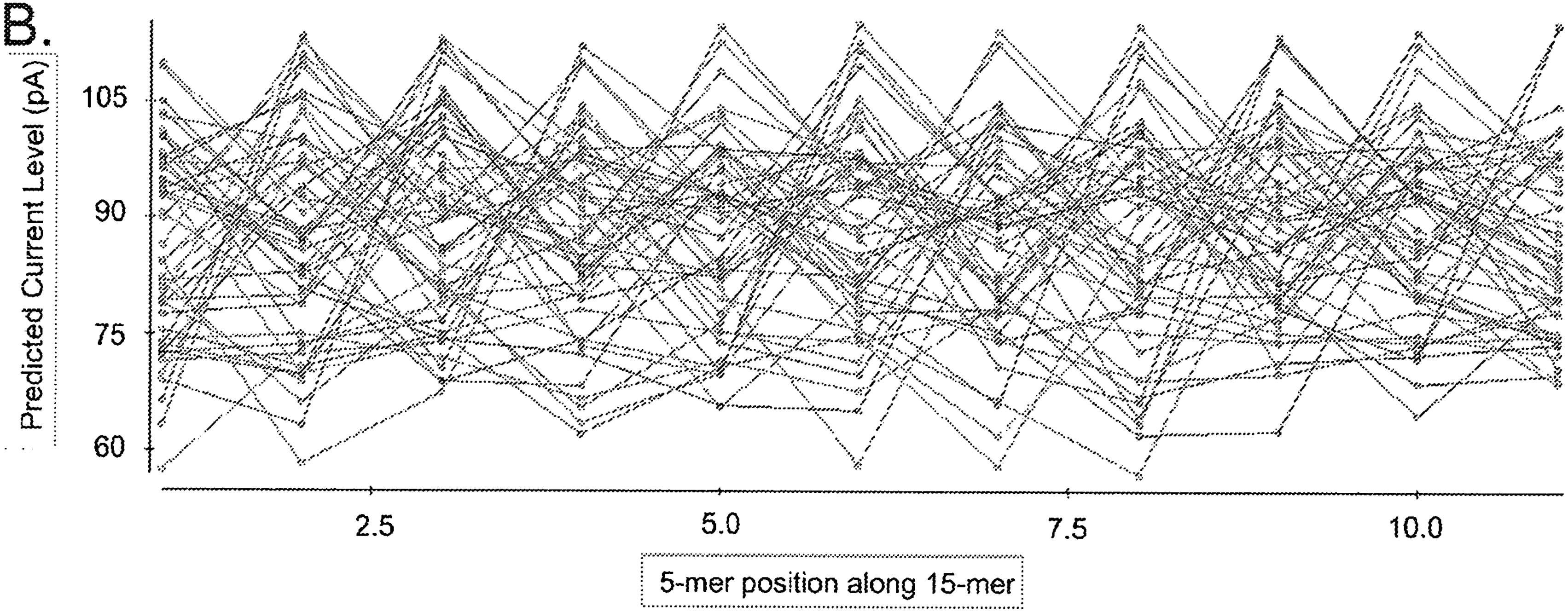


Fig. 27B

REPLACEMENT SHEET
37/60

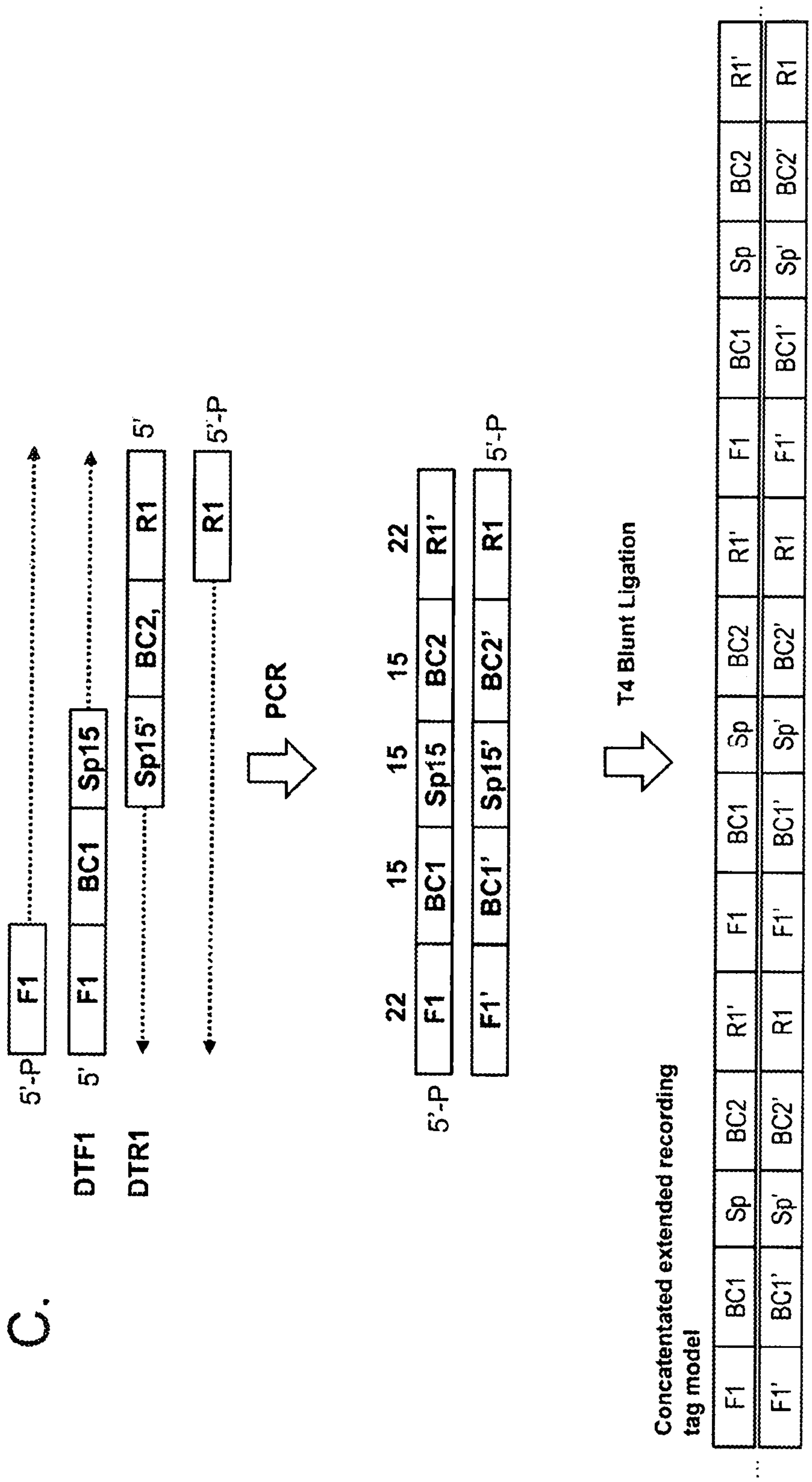


Fig. 27C

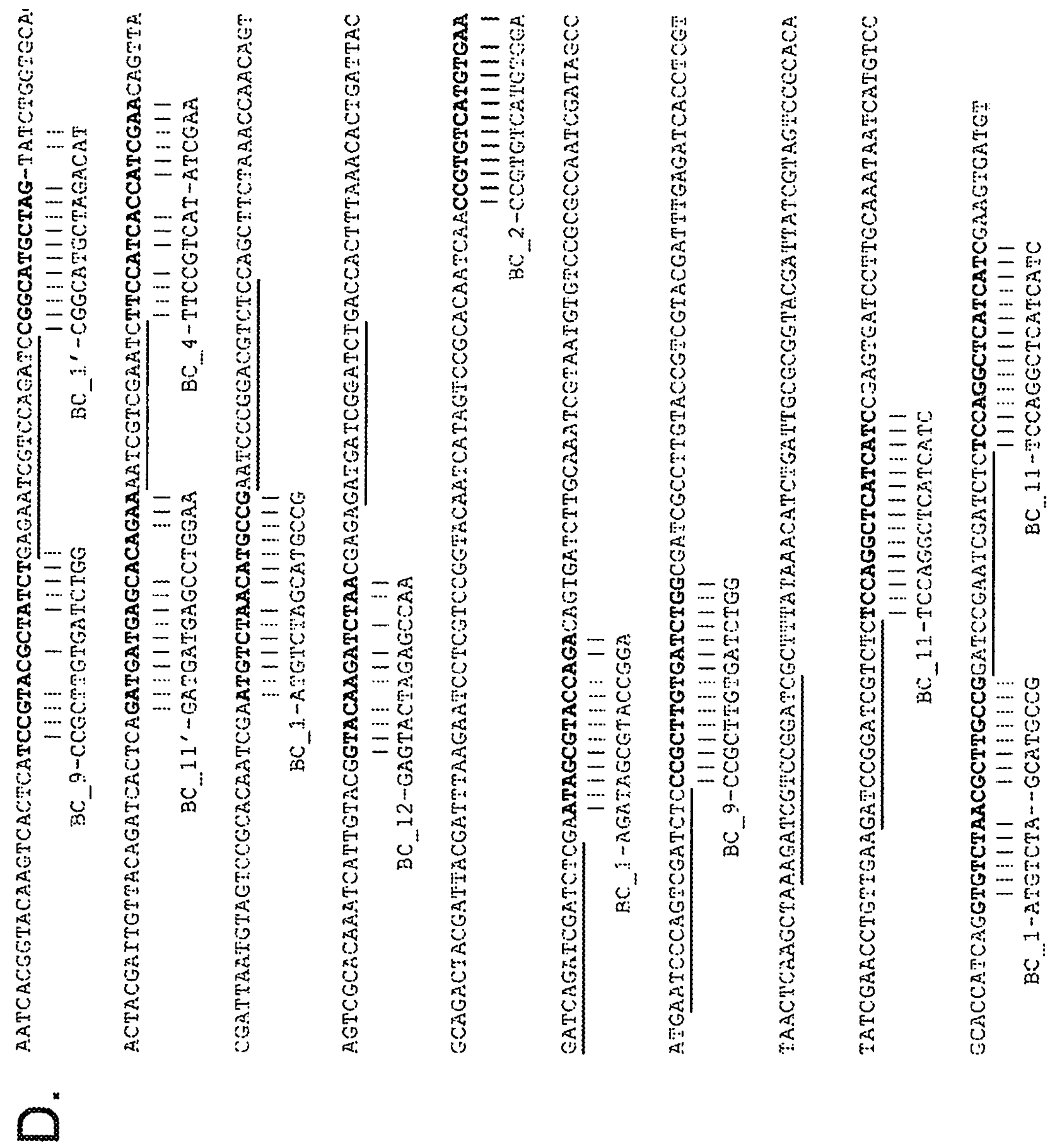


Fig. 27D

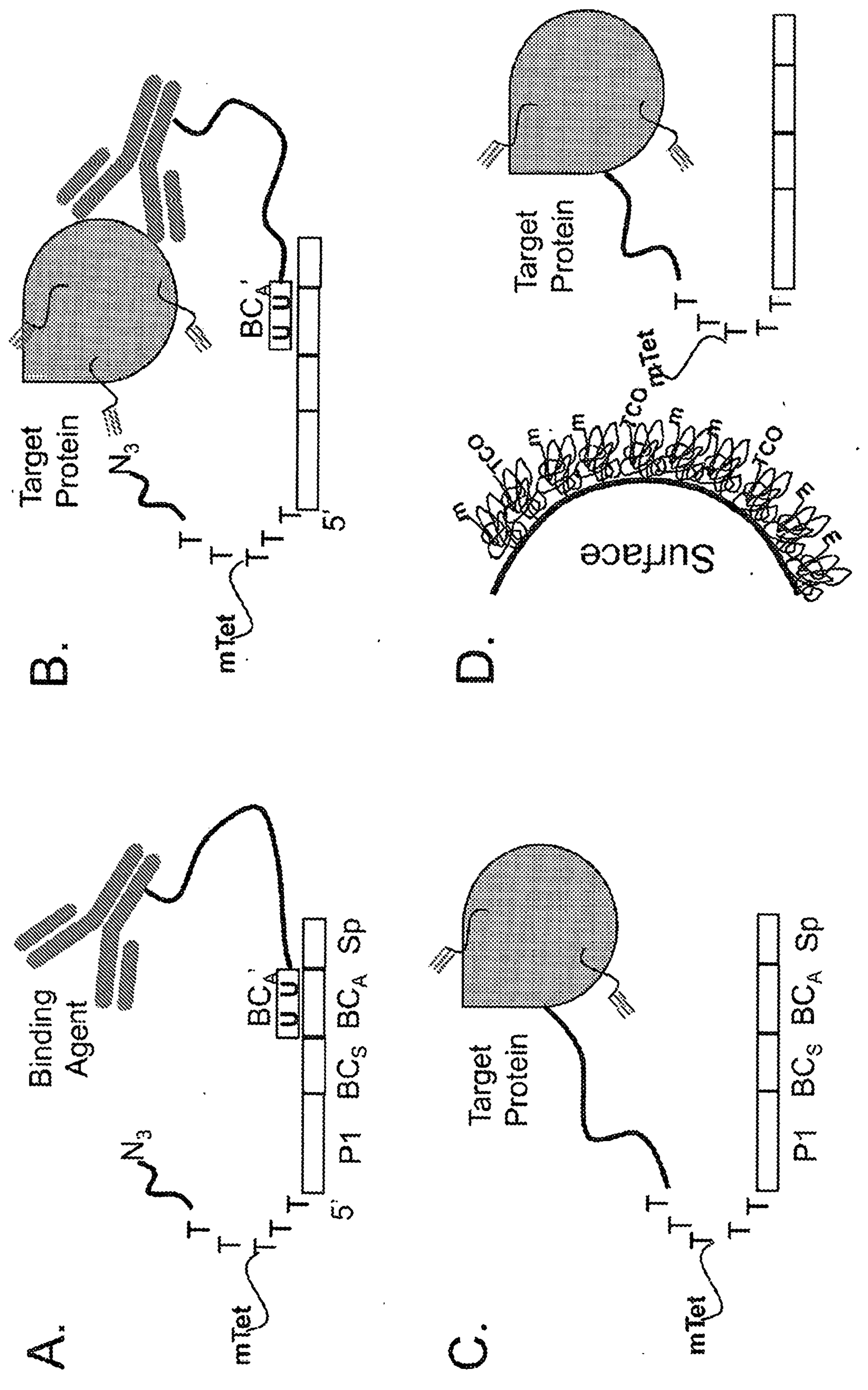

A.
Binding Agent
N3
BC_A'
U U
P1 BC_S BC_A Sp
5'
mTet
B.
Target Protein
N3
BC_A'
U U
5'
mTet
C.
Target Protein
P1 BC_S BC_A Sp
mTet
D.
Target Protein
mTet
TCO
Surface

Fig. 28

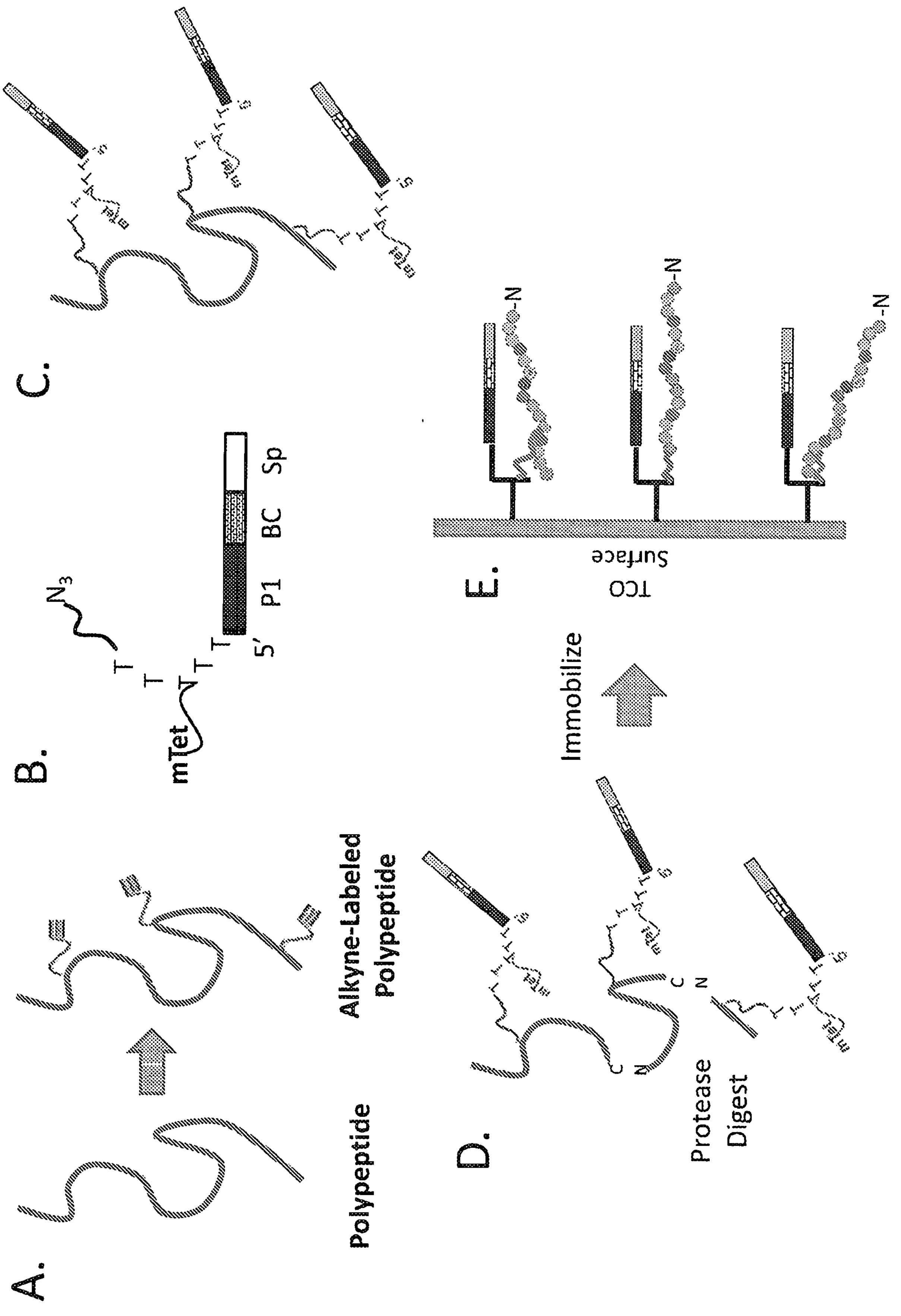

A.
Polypeptide
Alkyne-Labeled Polypeptide
B.
mTet
T
T
T
T
T
N_3
5'
P1
BC
Sp
C.
D.
Protease Digest
C
N
E.
TCO Surface
Immobilize
N-

Figure 29

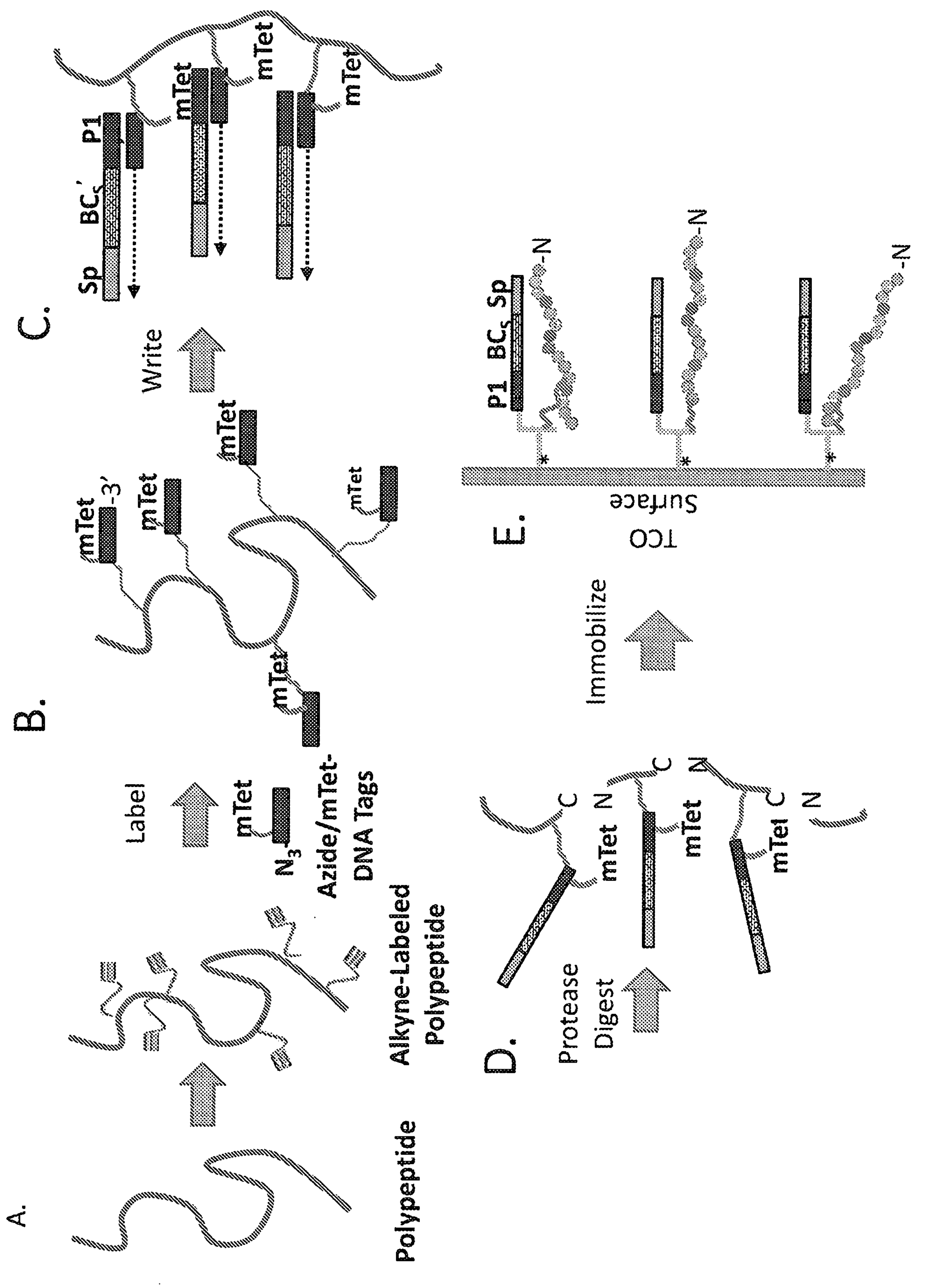

A.
Polypeptide
Alkyne-Labeled Polypeptide
Label
mTet
N3
Azide/mTet-DNA Tags
B.
mTet
-3'
Write
C.
Sp
BC5'
P1
D.
Protease Digest
C
N
Immobilize
E.
P1 BC5 Sp
N
TCO
Surface

Figure 30

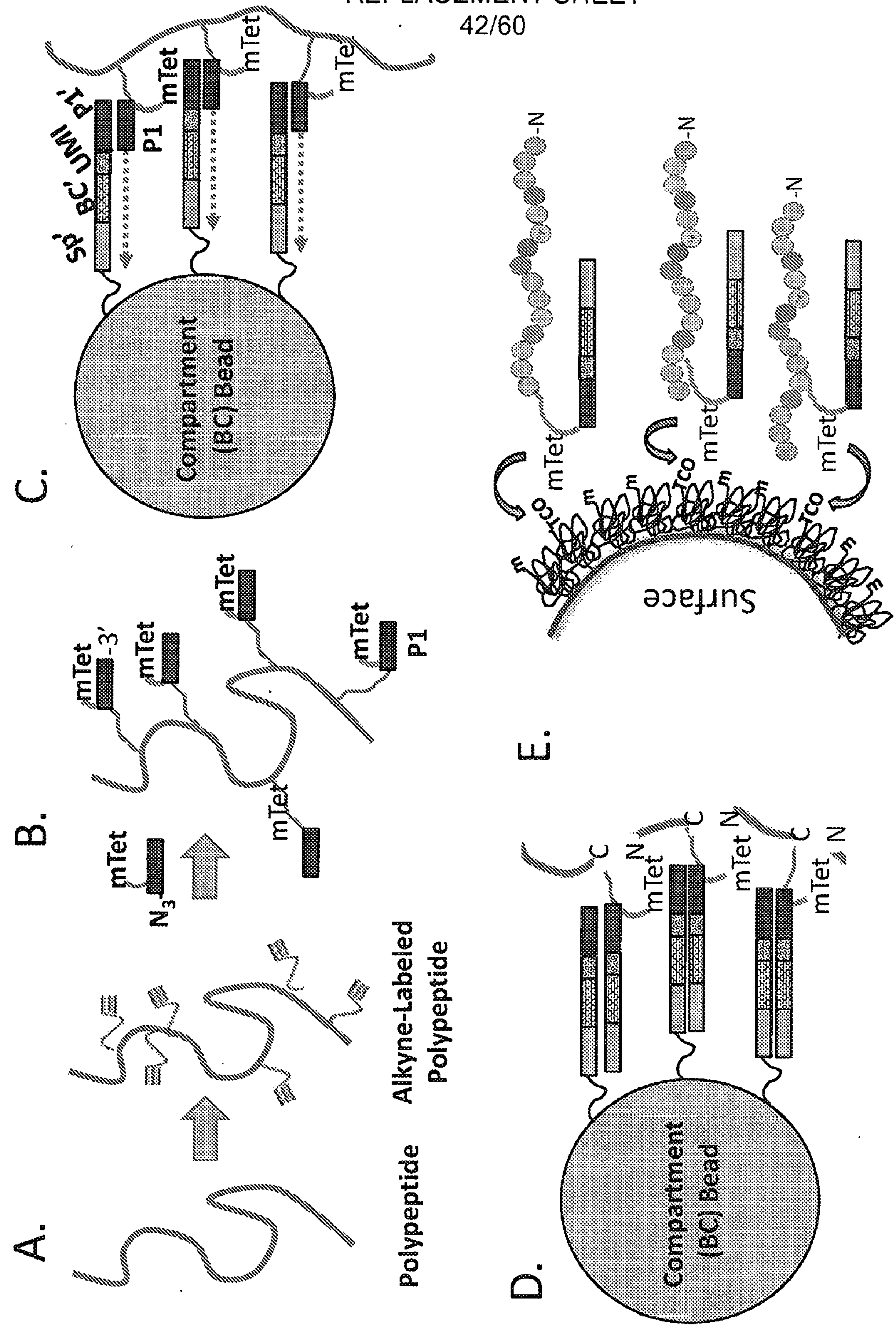

A.
Polypeptide
Alkyne-Labeled Polypeptide
B.
mTet
N3
-3'
P1
C.
Compartment (BC) Bead
SP'
BC'
UMI'
P1'
D.
C
N
E.
Surface
TCO
-N

Figure 31

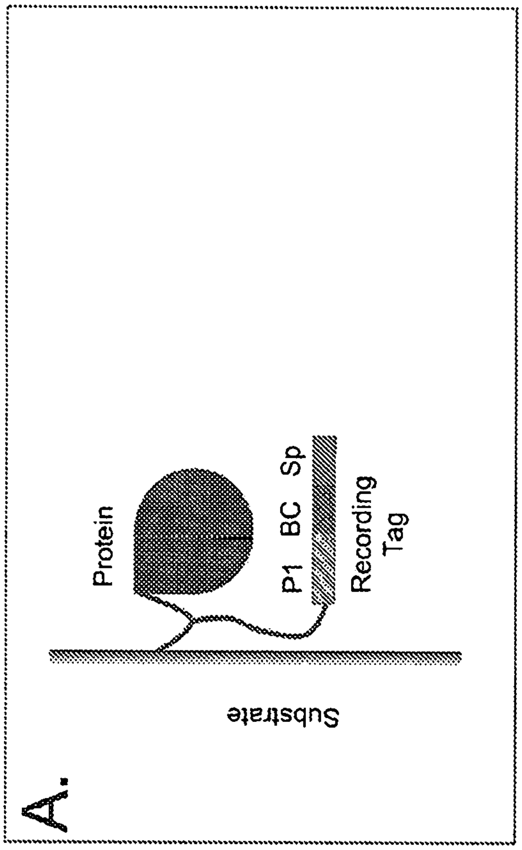


Fig. 32A

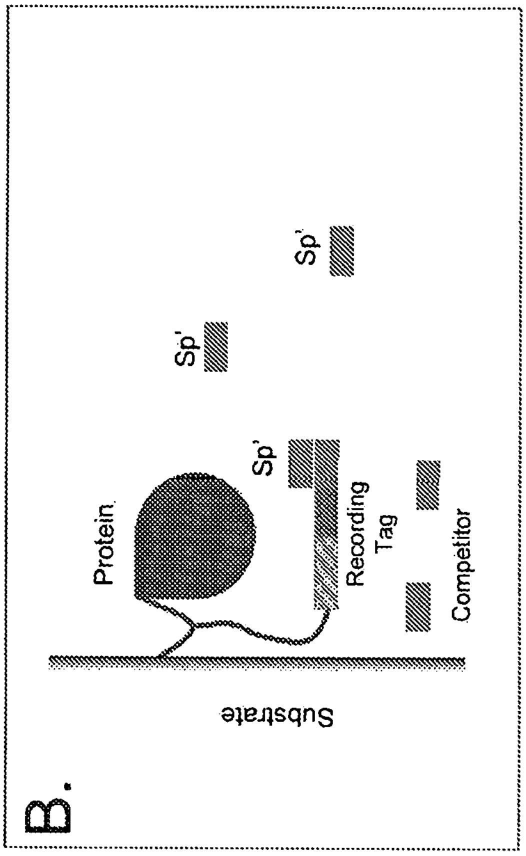


Fig. 32B

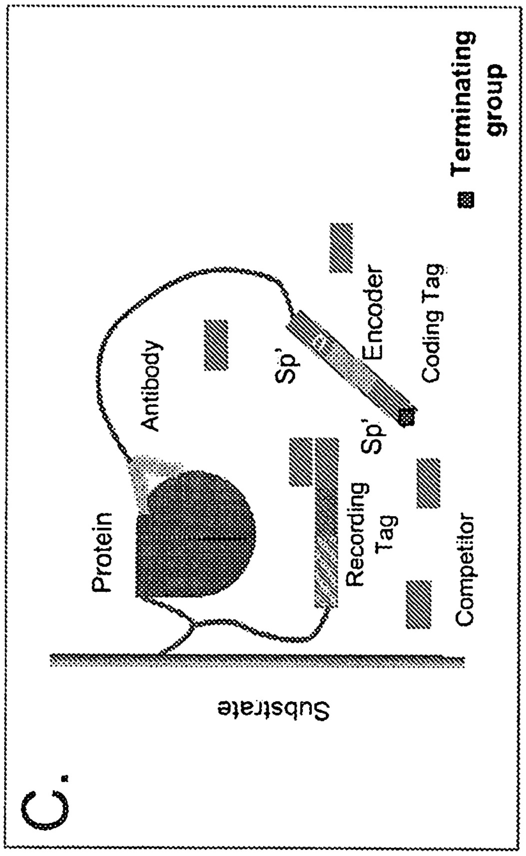


Fig. 32C

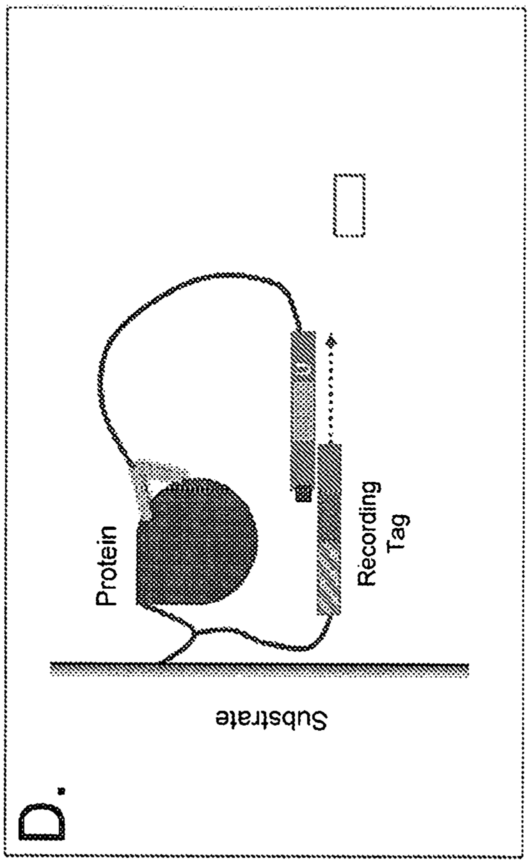


Fig. 32D

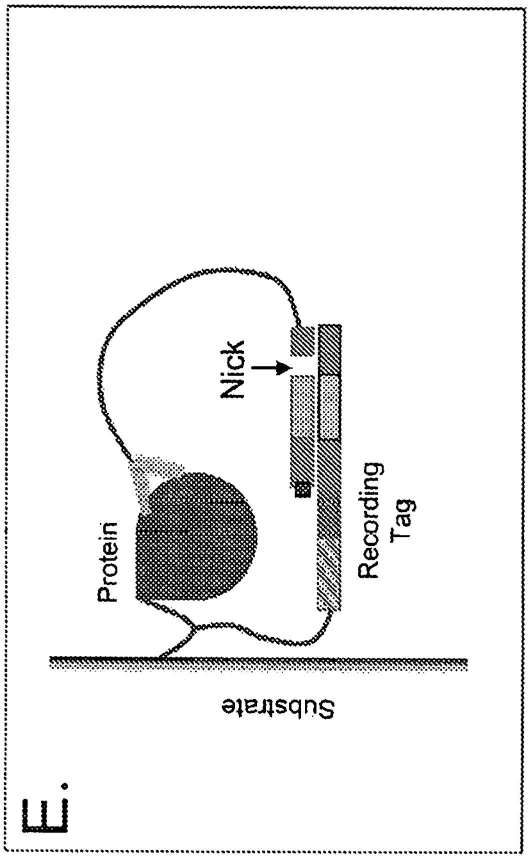


Fig. 32E

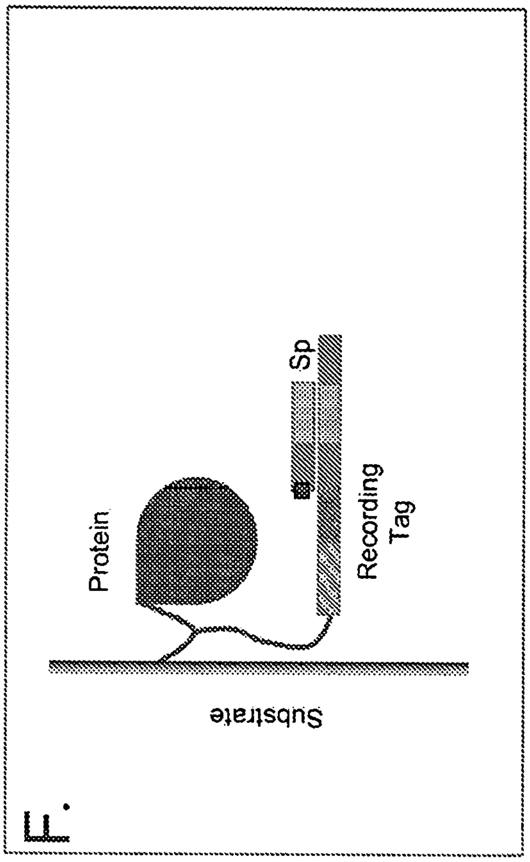


Fig. 32F

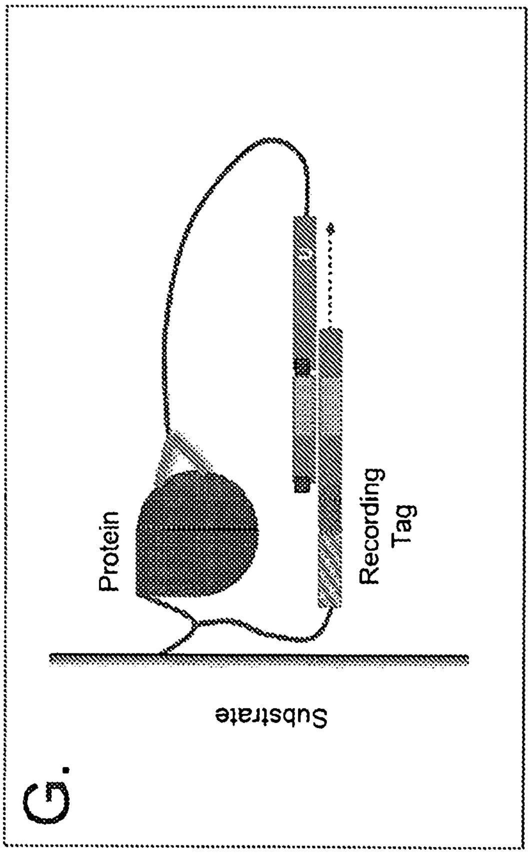


Fig. 32G

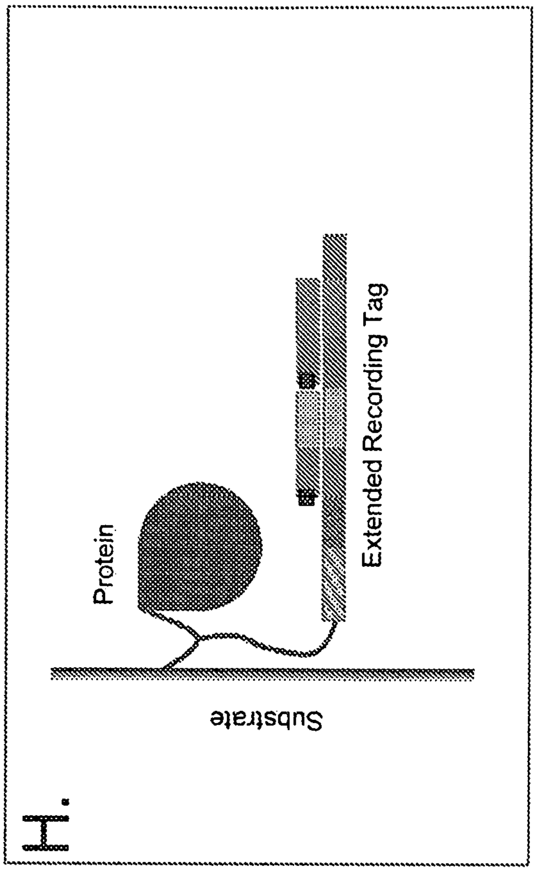


Fig. 32H

REPLACEMENT SHEET
45/60

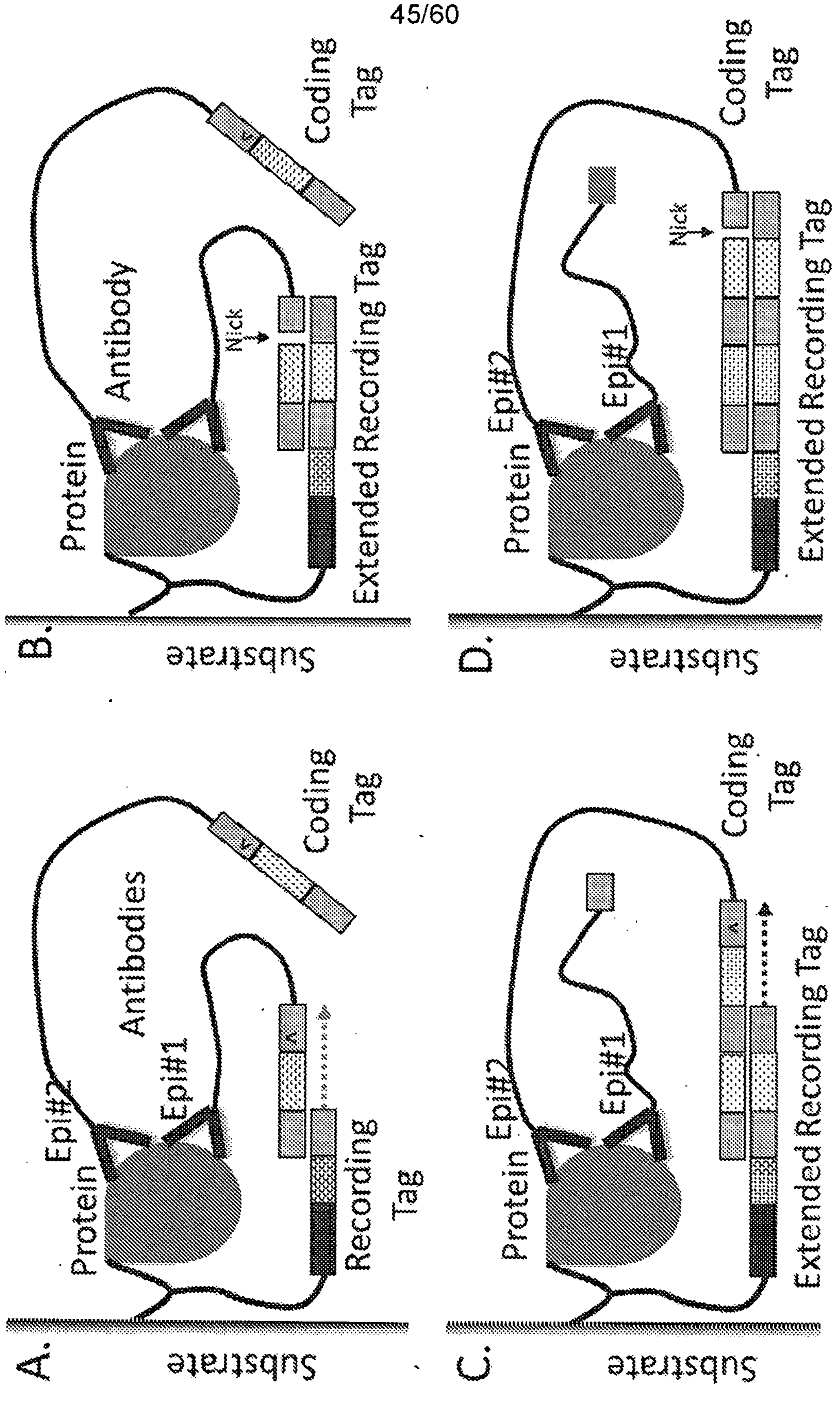


Fig. 33

REPLACEMENT SHEET
46/60

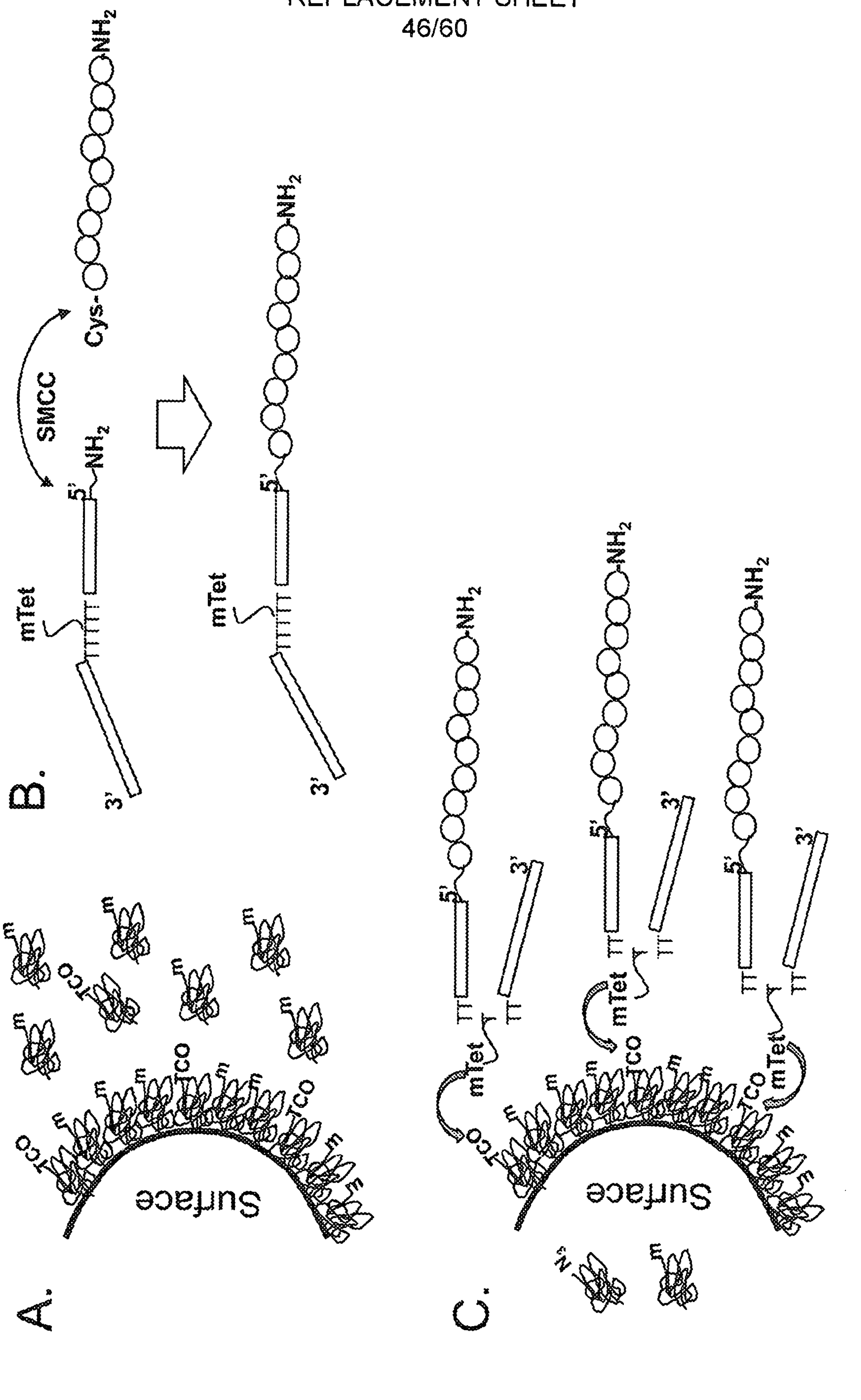


Fig. 34

REPLACEMENT SHEET
47/60

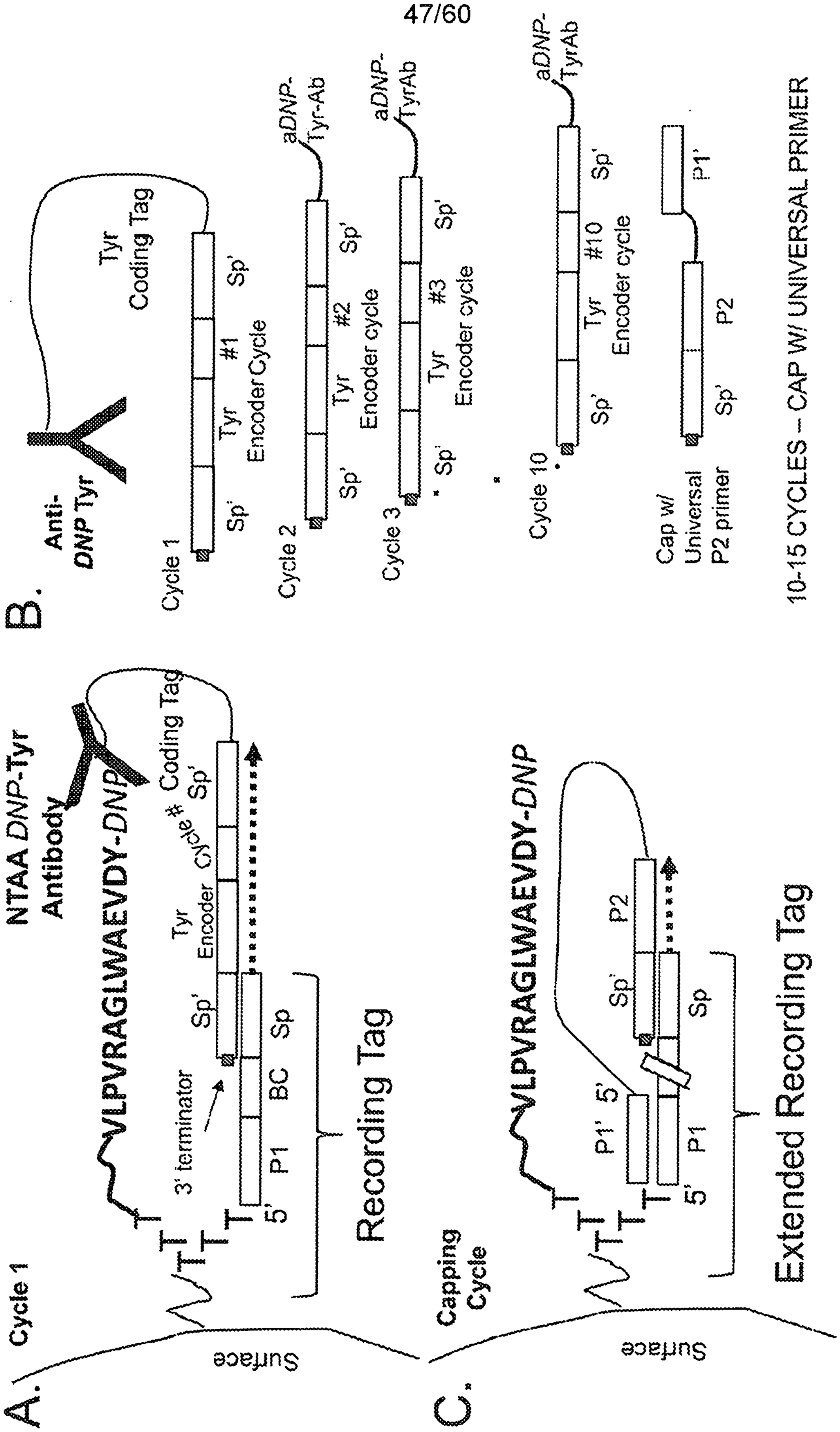


Fig. 35

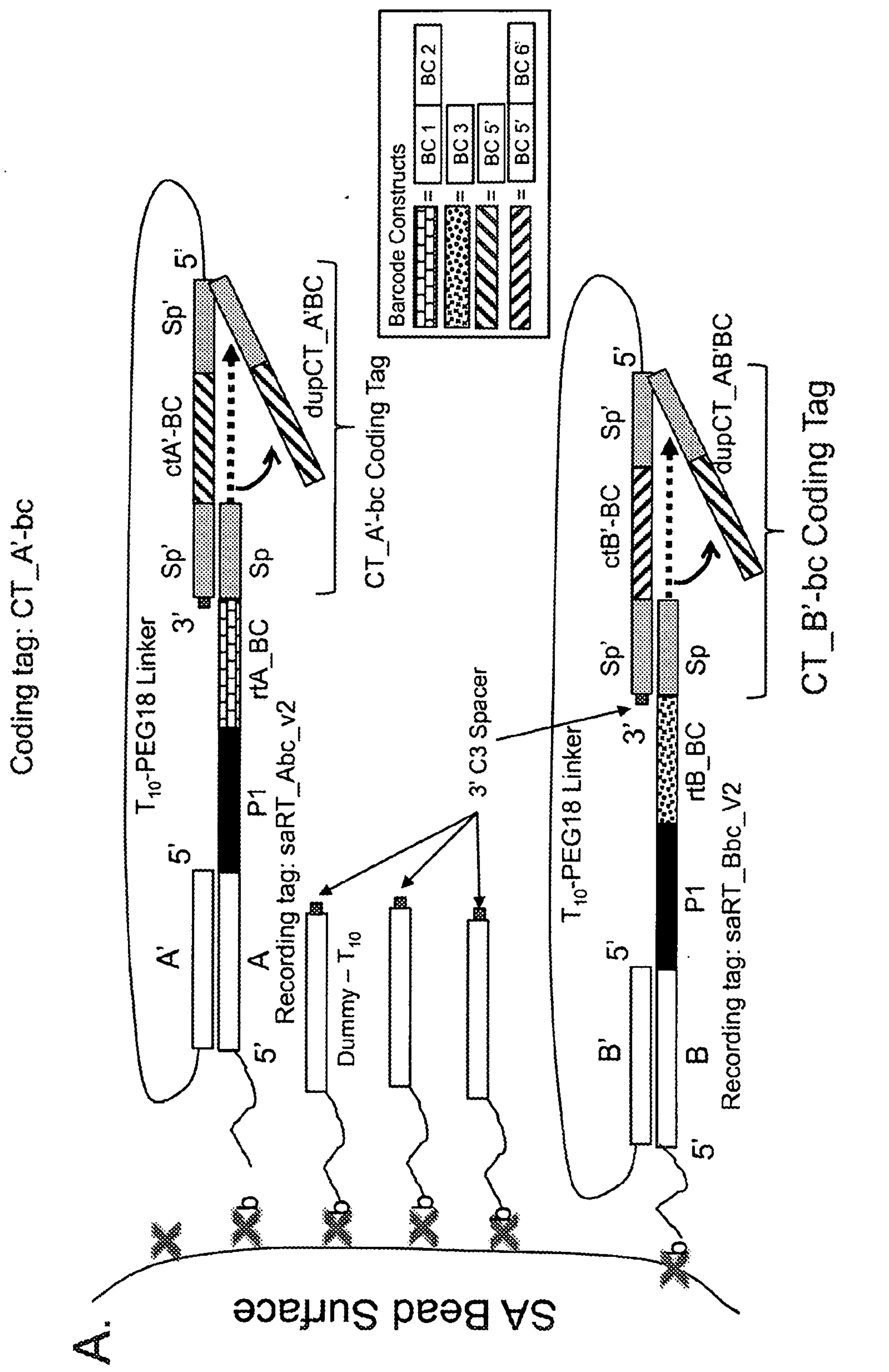

A.
SA Bead Surface
Coding tag: CT_A'-bc
T_{10}-PEG18 Linker
A'
A
5'
3'
P1
rtA_BC
Sp'
Sp
ctA'-BC
dupCT_A'BC
CT_A'-bc Coding Tag
Recording tag: saRT_Abc_v2
Dummy – T_{10}
3' C3 Spacer
Barcode Constructs
BC 1
BC 2
BC 3
BC 5'
BC 6'
B'
B
rtB_BC
ctB'-BC
dupCT_AB'BC
CT_B'-bc Coding Tag
Recording tag: saRT_Bbc_V2

Fig. 36A

B.

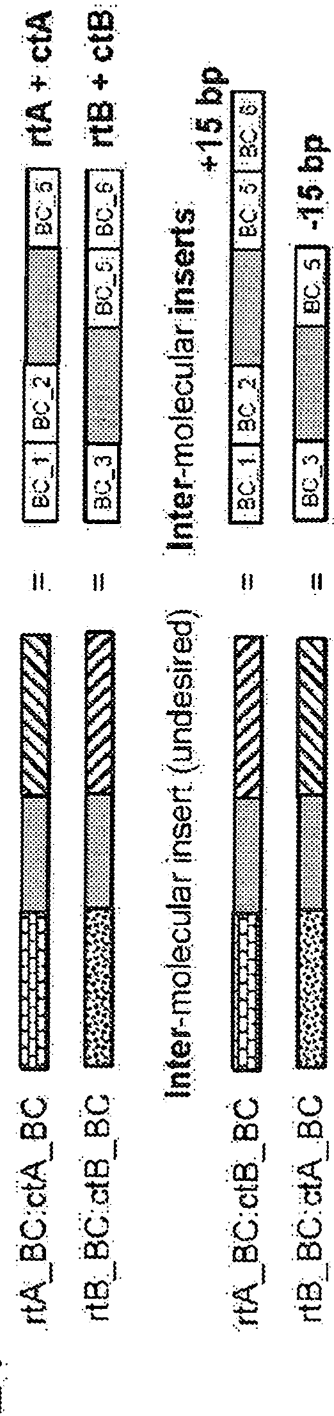


Fig. 36B

C.

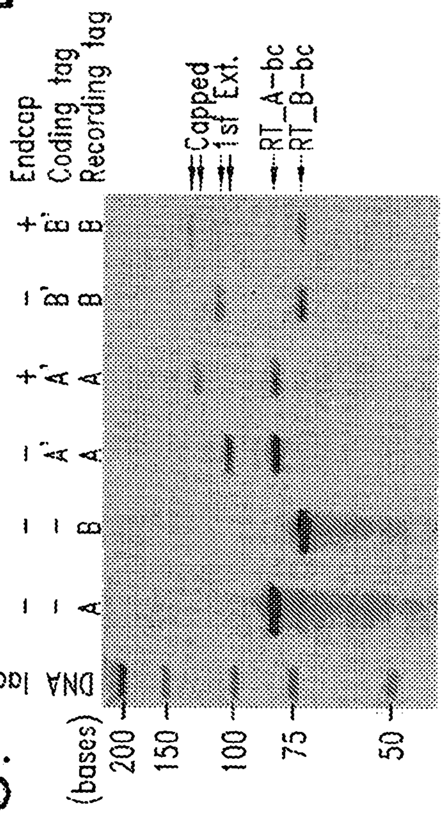


Fig. 36C

D.

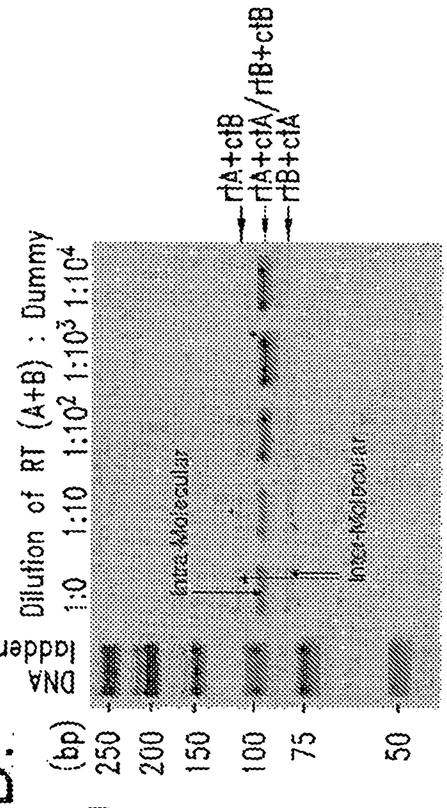


Fig. 36D

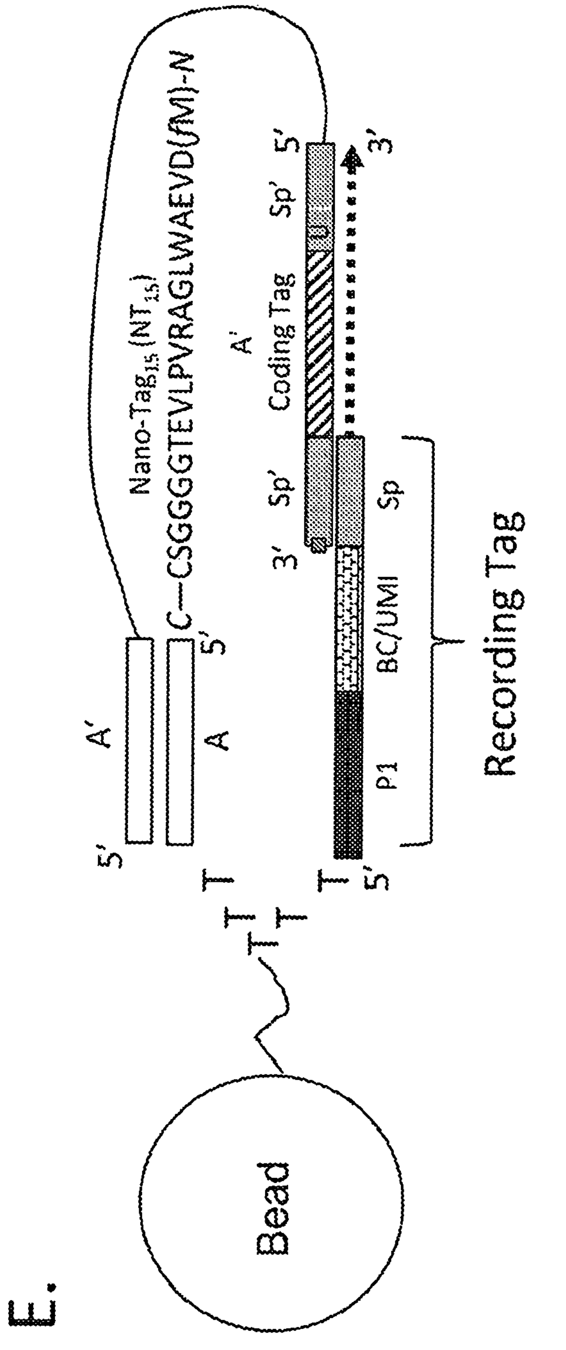


Fig. 36E

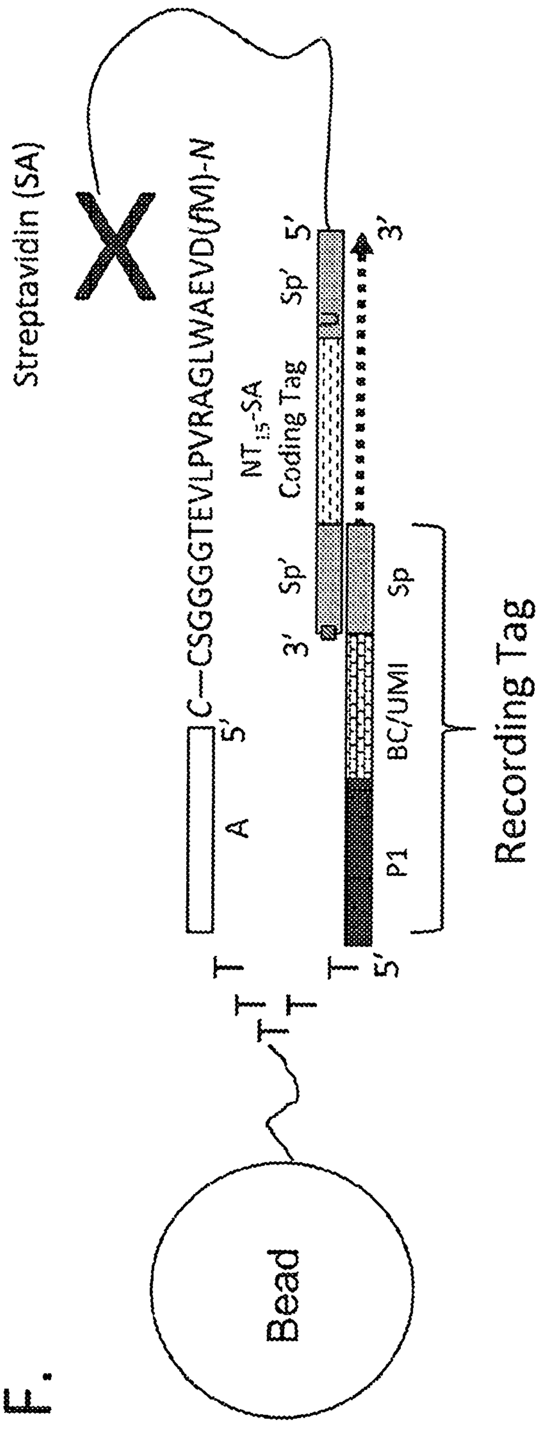


Fig. 36F

REPLACEMENT SHEET
51/60

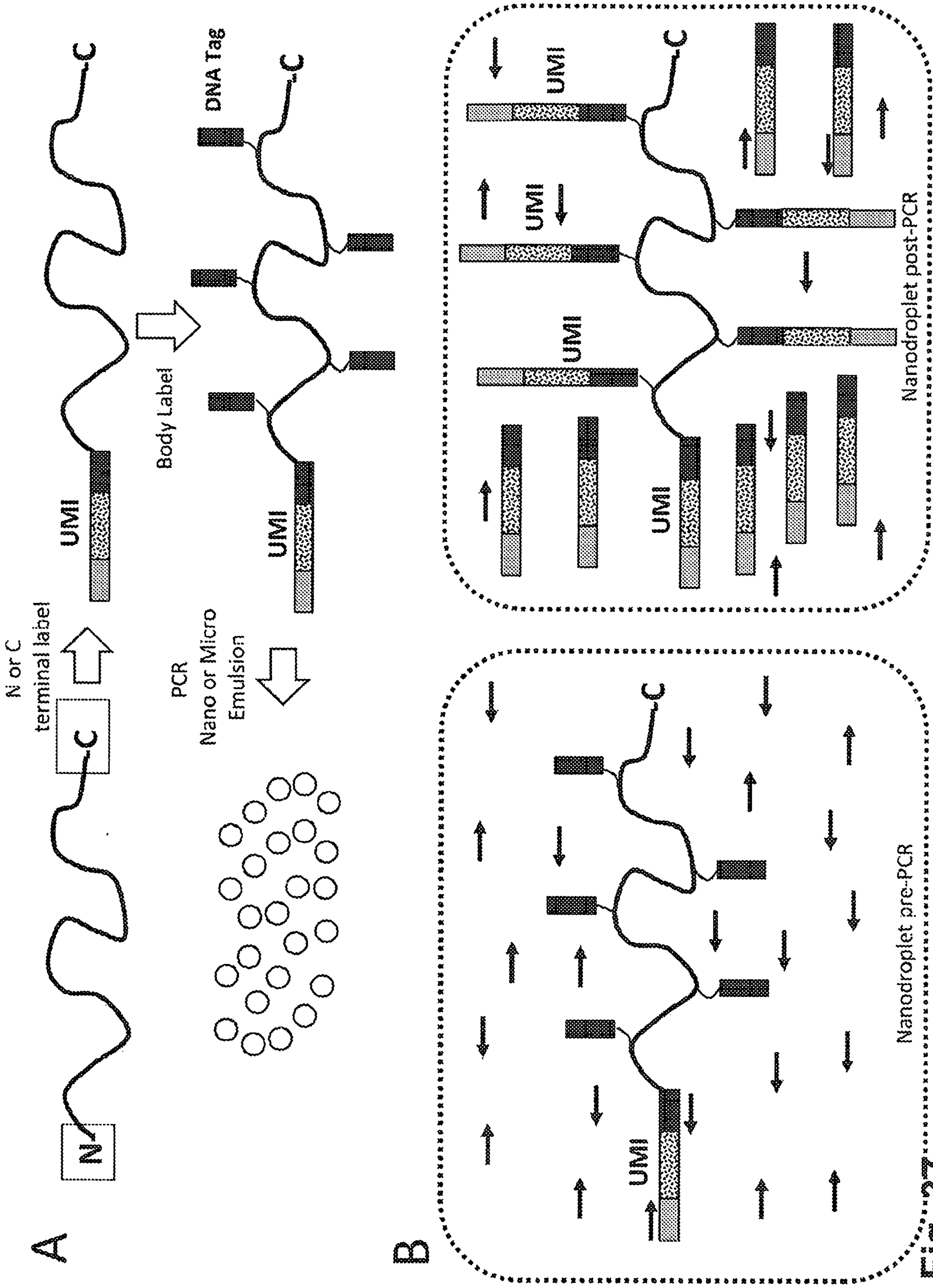


Fig. 37

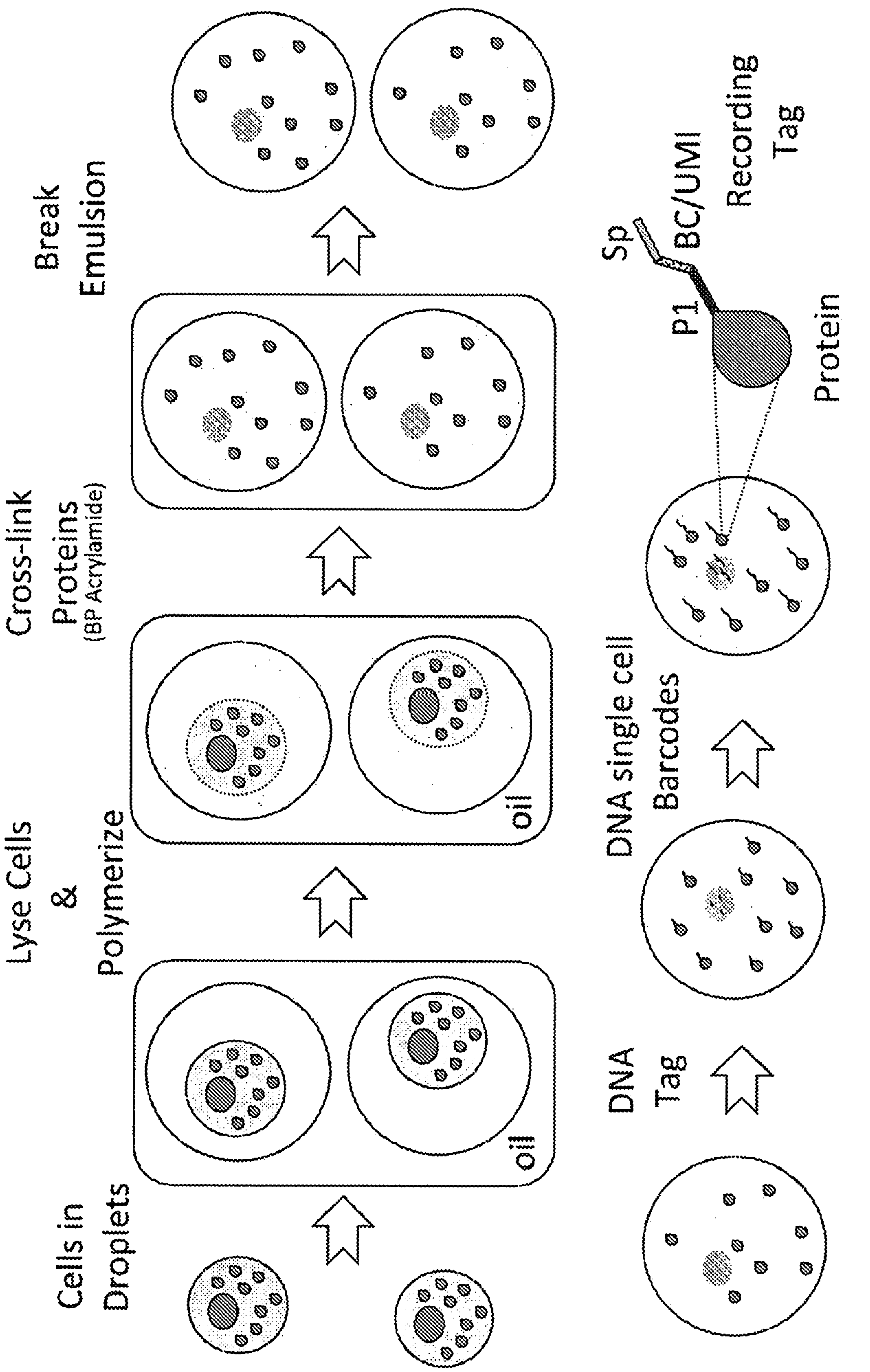

Cells in Droplets
Lyse Cells & Polymerize
oil
oil
Cross-link Proteins (BP Acrylamide)
Break Emulsion
DNA Tag
DNA single cell Barcodes
Sp
BC/UMI
Recording Tag
P1
Protein

Fig. 38

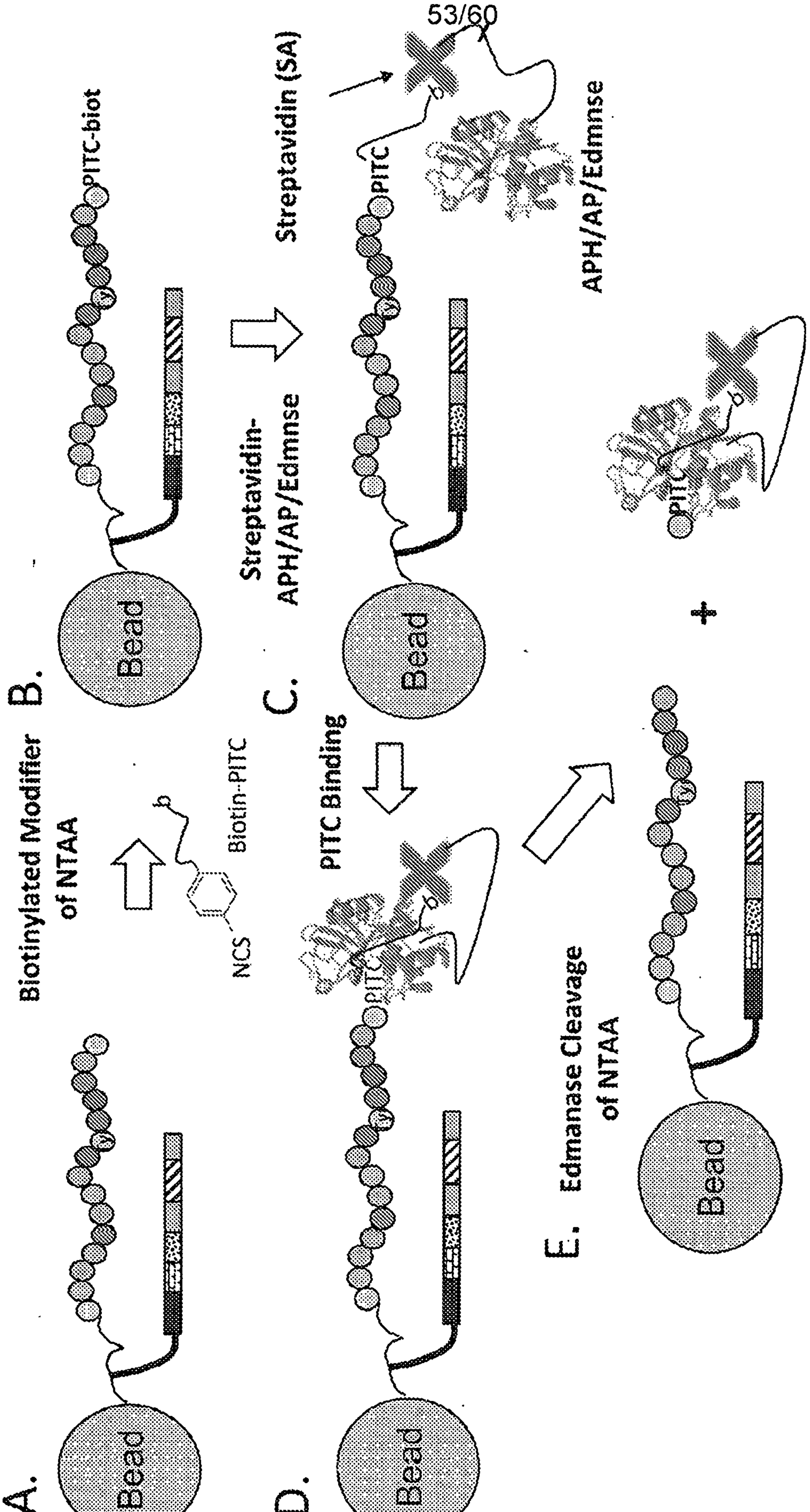
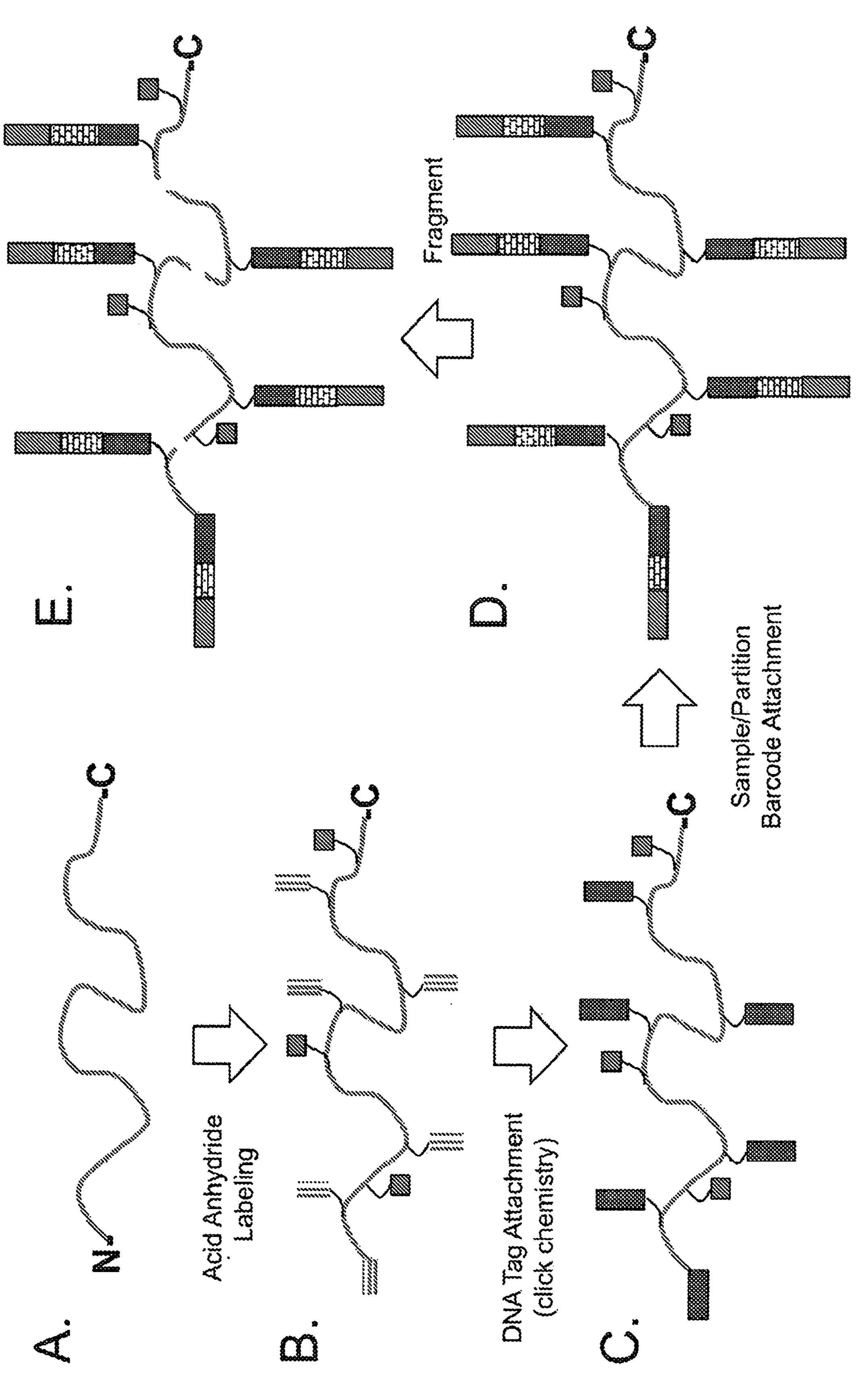
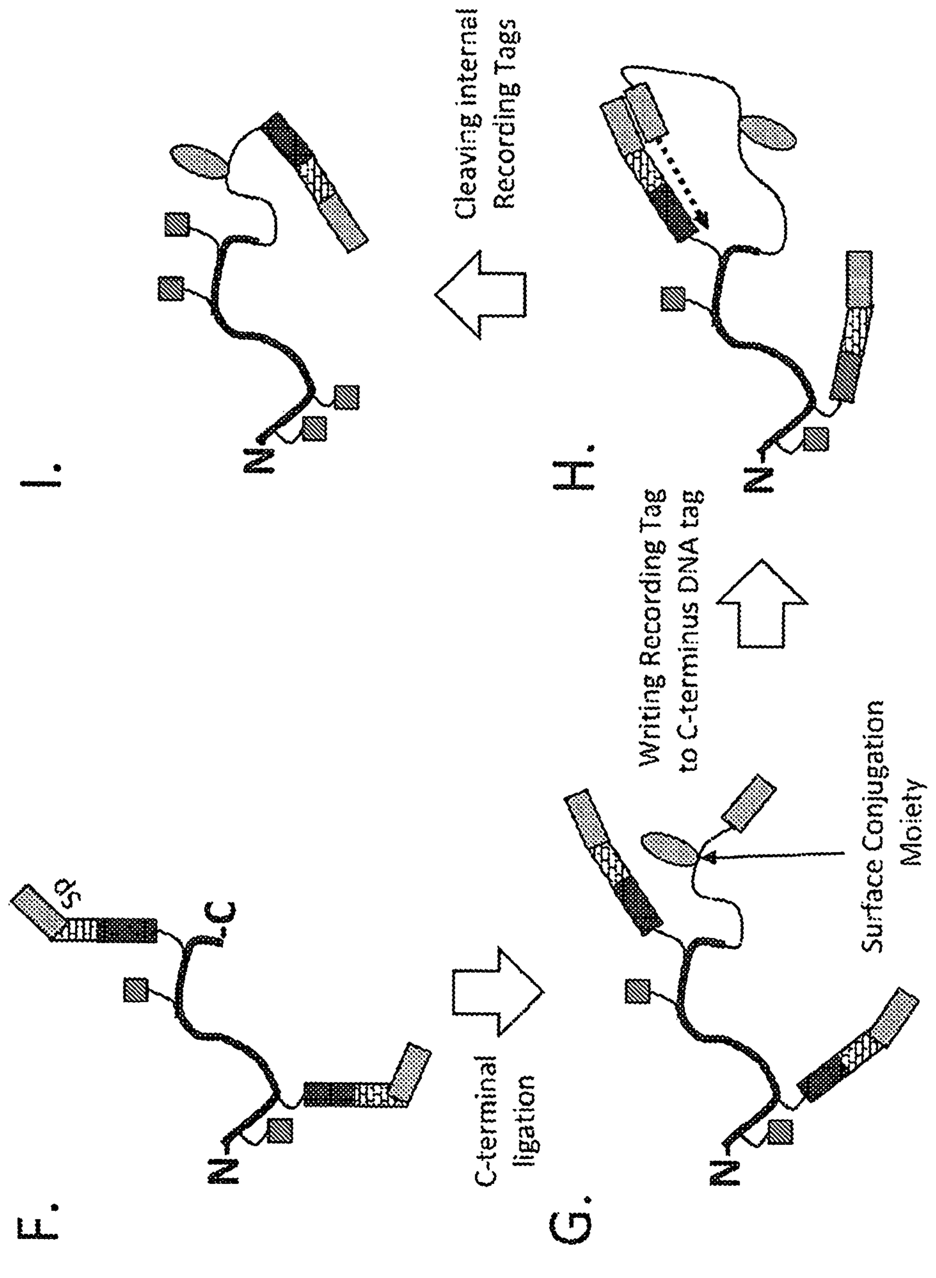

A.
Bead
Biotinylated Modifier of NTAA
NCS
Biotin-PITC
B.
Bead
PITC-biot
Streptavidin-APH/AP/Edmnse
C.
Bead
PITC
Streptavidin (SA)
APH/AP/Edmnse
PITC Binding
D.
Bead
PITC
Edmanase Cleavage of NTAA
E.
Bead
+
PITC

Fig. 39


A.
N
C
Acid Anhydride Labeling
B.
DNA Tag Attachment (click chemistry)
C.
Sample/Partition Barcode Attachment
D.
Fragment
E.

Figs. 40A-D


F.
N
C
Sp
C-terminal ligation
G.
N
Surface Conjugation Moiety
Writing Recording Tag to C-terminus DNA tag
H.
N
Cleaving internal Recording Tags
I.
N

Figs. 40F-H

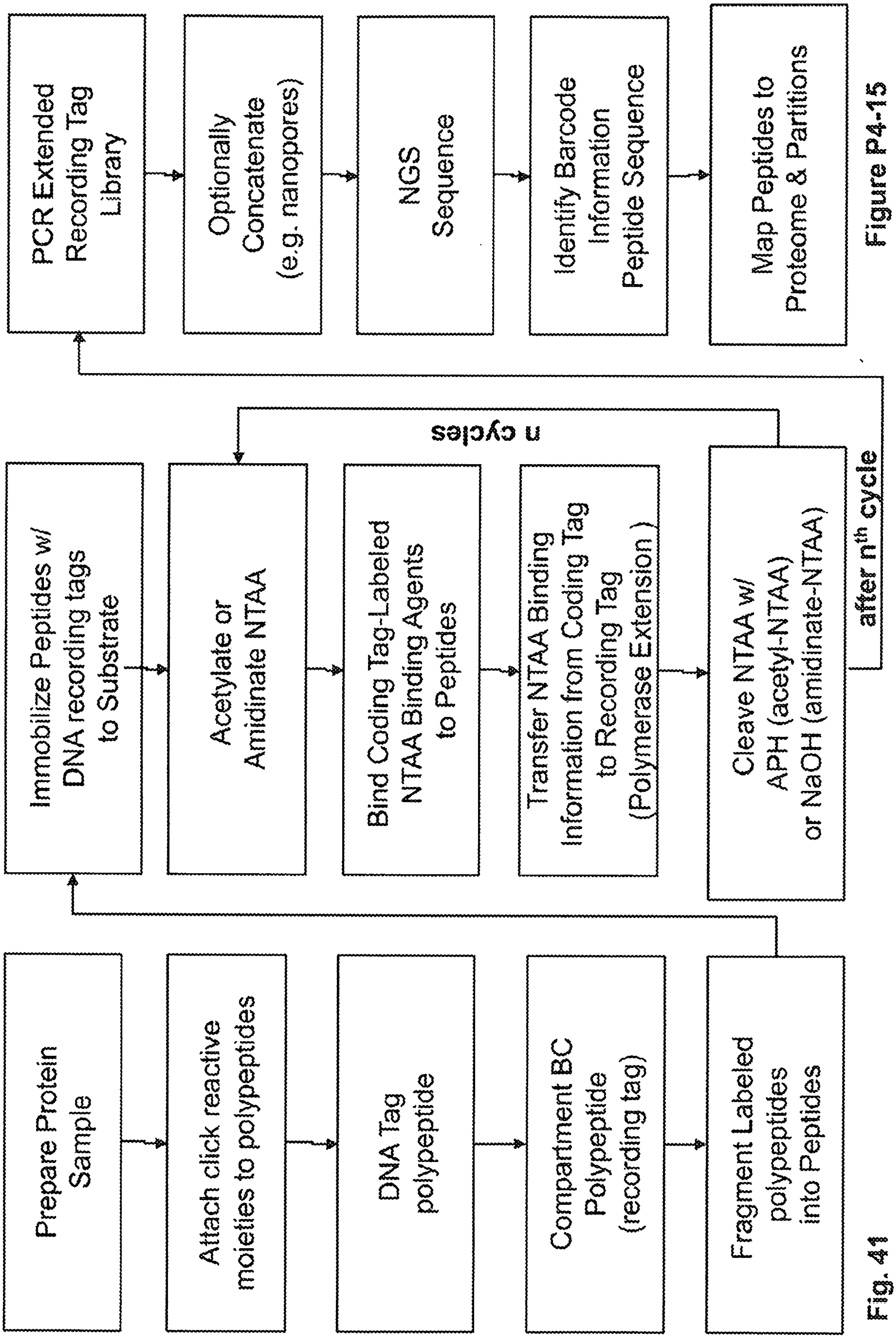


Fig. 41

Figure P4-15

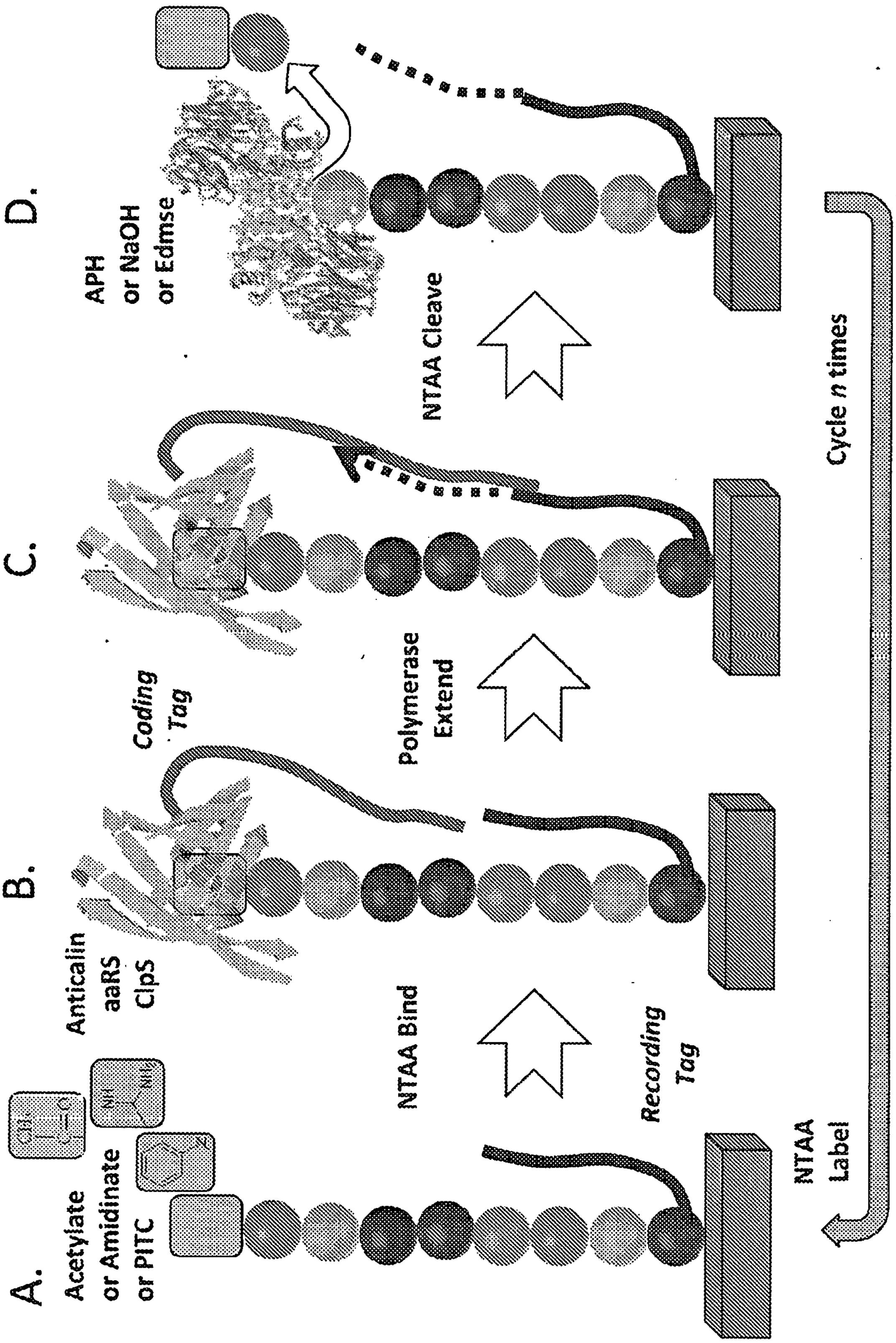

A.
Acetylate or Amidinate or PITC
B.
Anticalin aaRS ClpS
C.
D.
APH or NaOH or Edmse
NTAA Bind
Recording Tag
Coding Tag
Polymerase Extend
NTAA Cleave
NTAA Label
Cycle *n* times

Fig. 42

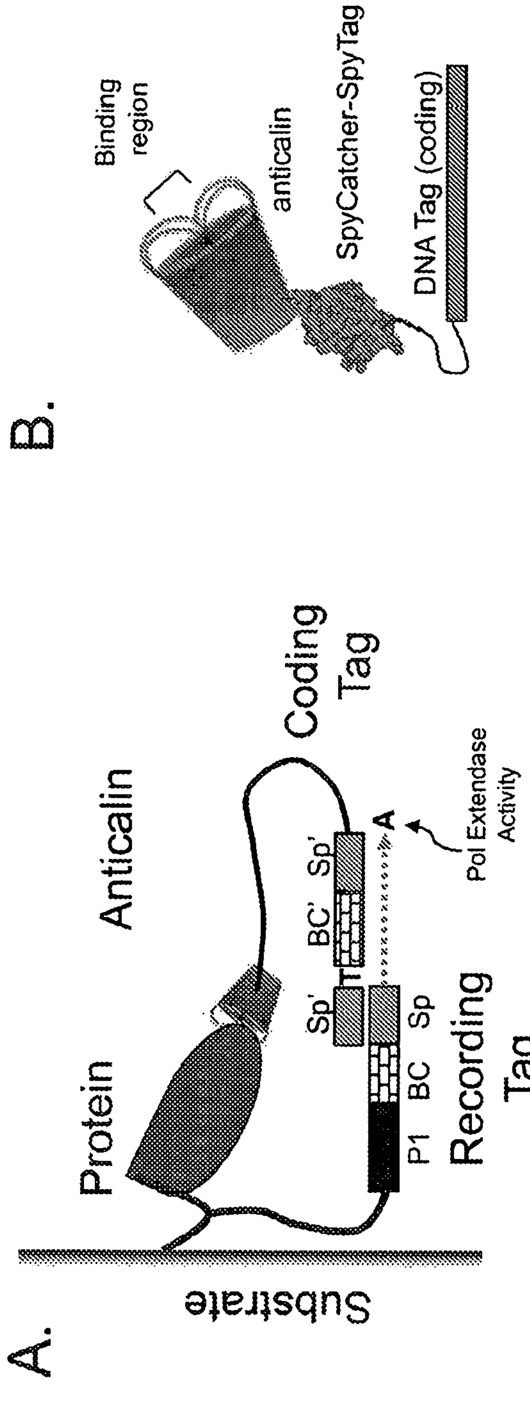

A.
Substrate
Protein
Anticalin
Coding Tag
Sp'
BC'
Sp'
A
Pol Extendase Activity
P1
BC
Sp
Recording Tag
B.
Binding region
anticalin
SpyCatcher-SpyTag
DNA Tag (coding)

Fig. 43

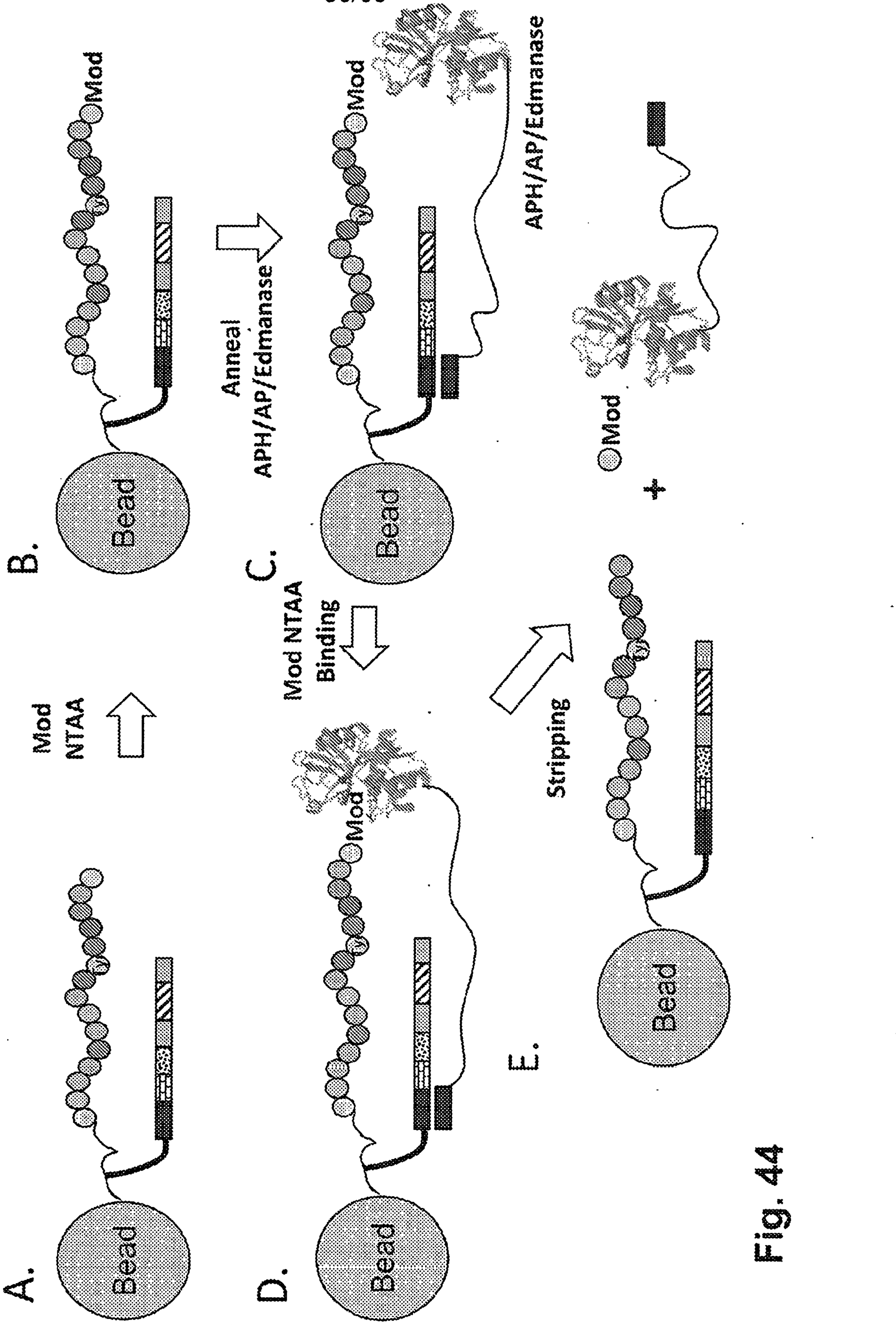

A.
Bead
Mod
NTAA
B.
Bead
Mod
Anneal
APH/AP/Edmanase
C.
Bead
Mod
APH/AP/Edmanase
Mod NTAA
Binding
D.
Bead
Mod
Stripping
E.
Bead
Mod
+

Fig. 44

REPLACEMENT SHEET
60/60

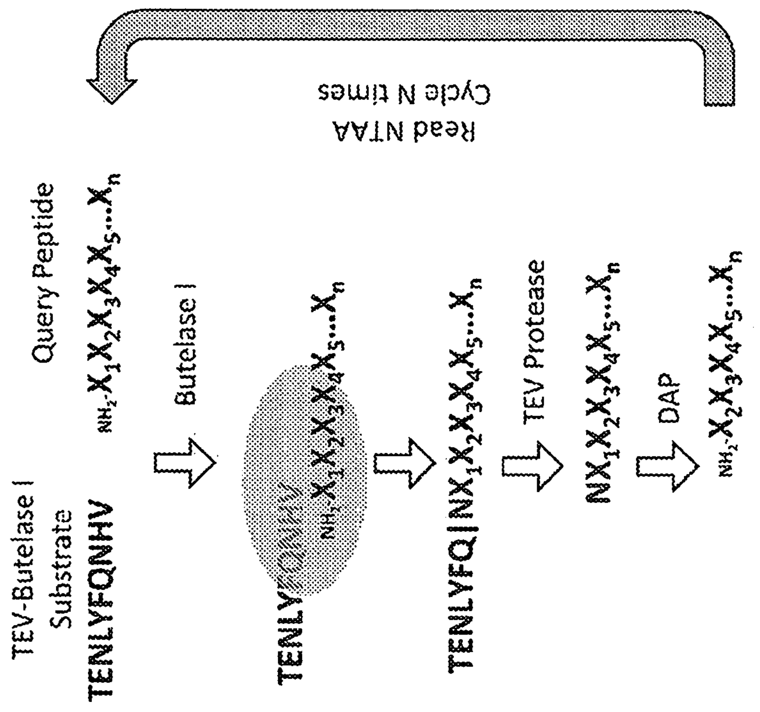


Fig. 45

SEQUENCE LISTING

<110> ProCure Life Sciences Inc.
      Chee, Mark
      Gunderson, Kevin
      Weiner, Michael Phillip

<120> MACROMOLECULE ANAYSIS EMPLOYING NUCLEIC
  ACID ENCODING

<130> 760229.401WO

<140> PCT
<141> 2017-05-02

<150> US 62/330,841
<151> 2016-05-02

<150> US 62/339,071
<151> 2016-05-19

<150> US 62/376,886
<151> 2016-08-18

<160> 180

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_1

<400> 1
atgtctagca tgccg 15

<210> 2
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_2

<400> 2
ccgtgtcatg tggaa 15

<210> 3
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_3

<400> 3
taagccggta tatca 15

<210> 4

<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_4

<400> 4
ttcgatatga cggaa 15

<210> 5
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_5

<400> 5
cgtatacgcg ttagg 15

<210> 6
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_6

<400> 6
aactgccgag attcc 15

<210> 7
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_7

<400> 7
tgatcttagc tgtgc 15

<210> 8
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_8

<400> 8
gagtcggtac cttga 15

<210> 9
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_9

<400> 9

ccgcttgtga tctgg 15

<210> 10
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_10

<400> 10
agatagcgta ccgga 15

<210> 11
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_11

<400> 11
tccaggctca tcatc 15

<210> 12
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_12

<400> 12
gagtactaga gccaa 15

<210> 13
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_13

<400> 13
gagcgtcaat aacgg 15

<210> 14
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_14

<400> 14
gcggtatcta cactg 15

<210> 15

<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_15

<400> 15
cttctccgaa gagaa 15

<210> 16
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_16

<400> 16
tgaagcctgt gttaa 15

<210> 17
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_17

<400> 17
ctggatggtt gtcga 15

<210> 18
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_18

<400> 18
actgcacggt tccaa 15

<210> 19
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_19

<400> 19
cgagagatgg tcctt 15

<210> 20
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_20

<400> 20
tcttgagaga caaga 15

<210> 21
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_21

<400> 21
aattcgcact gtgtt 15

<210> 22
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_22

<400> 22
gtagtgccgc taaga 15

<210> 23
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_23

<400> 23
cctatagcac aatcc 15

<210> 24
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_24

<400> 24
atcaccgagg ttgga 15

<210> 25
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_25

<400> 25
gattcaacgg agaag 15

<210> 26
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_26

<400> 26
acgaacctcg cacca 15

<210> 27
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_27

<400> 27
aggacttcaa gaaga 15

<210> 28
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_28

<400> 28
ggttgaatcc tcgca 15

<210> 29
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_29

<400> 29
aaccaacctc tagcg 15

<210> 30
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_30

<400> 30

acgcgaatat ctaac 15

<210> 31
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_31

<400> 31
gttgagaatt acacc 15

<210> 32
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_32

<400> 32
ctctctctgt gaacc 15

<210> 33
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_33

<400> 33
gccatcagta agaga 15

<210> 34
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_34

<400> 34
gcaacgtgaa ttgag 15

<210> 35
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_35

<400> 35
ctaagtagag ccaca 15

<210> 36

<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_36

<400> 36
tgtctgttgg aagcg 15

<210> 37
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_37

<400> 37
ttaatagaca gcgcg 15

<210> 38
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_38

<400> 38
cgacgctcta acaag 15

<210> 39
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_39

<400> 39
catggcttat tgaga 15

<210> 40
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_40

<400> 40
actaggtatg gccgg 15

<210> 41
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_41

<400> 41
gtcctcgtct atcct 15

<210> 42
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_42

<400> 42
taggattccg ttacc 15

<210> 43
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_43

<400> 43
tctgaccacc ggaag 15

<210> 44
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_44

<400> 44
agagtcacct cgtgg 15

<210> 45
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_45

<400> 45
ctgatgtagt cgaag 15

<210> 46
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_46

<400> 46
gtcggttgcg gatag 15

<210> 47
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_47

<400> 47
tcctcctcct aagaa 15

<210> 48
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_48

<400> 48
attcggtcca cttca 15

<210> 49
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_49

<400> 49
ccttacaggt ctgcg 15

<210> 50
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_50

<400> 50
gatcattggc caatt 15

<210> 51
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_51

<400> 51

ttcaaggctg agttg 15

<210> 52
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_52

<400> 52
tggctcgatt gaatc 15

<210> 53
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_53

<400> 53
gtaagccatc cgctc 15

<210> 54
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_54

<400> 54
acacatgcgt agaca 15

<210> 55
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_55

<400> 55
tgctatggat tcaag 15

<210> 56
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_56

<400> 56
ccacgaggct tagtt 15

<210> 57

<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_57

<400> 57
ggccaactaa ggtgc 15

<210> 58
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_58

<400> 58
gcacctattc gacaa 15

<210> 59
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_59

<400> 59
tggacacgat cggct 15

<210> 60
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_60

<400> 60
ctataattcc aacgg 15

<210> 61
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_61

<400> 61
aacgtggtta gtaag 15

<210> 62
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_62

<400> 62
caaggaacga gtggc 15

<210> 63
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_63

<400> 63
caccagaacg gaaga 15

<210> 64
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_64

<400> 64
cgtacggtca agcaa 15

<210> 65
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_65

<400> 65
tcggtgacag gctaa 15

<210> 66
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_1 REV

<400> 66
cggcatgcta gacat 15

<210> 67
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_2 REV

<400> 67
ttccacatga cacgg 15

<210> 68
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_3 REV

<400> 68
tgatataccg gctta 15

<210> 69
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_4 REV

<400> 69
ttccgtcata tcgaa 15

<210> 70
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_5 REV

<400> 70
cctaacgcgt atacg 15

<210> 71
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_6 REV

<400> 71
ggaatctcgg cagtt 15

<210> 72
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_7 REV

<400> 72

gcacagctaa gatca 15

<210> 73
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_8 REV

<400> 73
tcaaggtacc gactc 15

<210> 74
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_9 REV

<400> 74
ccagatcaca agcgg 15

<210> 75
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_10 REV

<400> 75
tccggtacgc tatct 15

<210> 76
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_11 REV

<400> 76
gatgatgagc ctgga 15

<210> 77
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_12 REV

<400> 77
ttggctctag tactc 15

<210> 78

<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_13 REV

<400> 78
ccgttattga cgctc 15

<210> 79
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_14 REV

<400> 79
cagtgtagat accgc 15

<210> 80
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_15 REV

<400> 80
ttctcttcgg agaag 15

<210> 81
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_16 REV

<400> 81
ttaacacagg cttca 15

<210> 82
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_17 REV

<400> 82
tcgacaacca tccag 15

<210> 83
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_18 REV

<400> 83
ttggaaccgt gcagt 15

<210> 84
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_19 REV

<400> 84
aaggaccatc tctcg 15

<210> 85
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_20 REV

<400> 85
tcttgtctct caaga 15

<210> 86
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_21 REV

<400> 86
aacacagtgc gaatt 15

<210> 87
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_22 REV

<400> 87
tcttagcggc actac 15

<210> 88
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_23 REV

<400> 88
ggattgtgct atagg   15

<210> 89
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_24 REV

<400> 89
tccaacctcg gtgat   15

<210> 90
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_25 REV

<400> 90
cttctccgtt gaatc   15

<210> 91
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_26 REV

<400> 91
tggtgcgagg ttcgt   15

<210> 92
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_27 REV

<400> 92
tcttcttgaa gtcct   15

<210> 93
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_28 REV

<400> 93

tgcgaggatt caacc 15

<210> 94
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_29 REV

<400> 94
cgctagaggt tggtt 15

<210> 95
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_30 REV

<400> 95
gttagatatt cgcgt 15

<210> 96
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_31 REV

<400> 96
ggtgtaattc tcaac 15

<210> 97
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_32 REV

<400> 97
ggttcacaga gagag 15

<210> 98
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_33 REV

<400> 98
tctcttactg atggc 15

<210> 99

<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_34 REV

<400> 99
ctcaattcac gttgc 15

<210> 100
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_35 REV

<400> 100
tgtggctcta cttag 15

<210> 101
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_36 REV

<400> 101
cgcttccaac agaca 15

<210> 102
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_37 REV

<400> 102
cgcgctgtct attaa 15

<210> 103
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_38 REV

<400> 103
cttgttagag cgtcg 15

<210> 104
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_39 REV

<400> 104
tctcaataag ccatg 15

<210> 105
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_40 REV

<400> 105
ccggccatac ctagt 15

<210> 106
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_41 REV

<400> 106
aggatagacg aggac 15

<210> 107
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_42 REV

<400> 107
ggtaacggaa tccta 15

<210> 108
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_43 REV

<400> 108
cttccggtgg tcaga 15

<210> 109
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_44 REV

<400> 109
ccacgaggtg actct 15

<210> 110
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_45 REV

<400> 110
cttcgactac atcag 15

<210> 111
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_46 REV

<400> 111
ctatccgcaa ccgac 15

<210> 112
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_47 REV

<400> 112
ttcttaggag gagga 15

<210> 113
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_48 REV

<400> 113
tgaagtggac cgaat 15

<210> 114
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_49 REV

<400> 114

cgcagacctg taagg 15

<210> 115
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_50 REV

<400> 115
aattggccaa tgatc 15

<210> 116
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_51 REV

<400> 116
caactcagcc ttgaa 15

<210> 117
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_52 REV

<400> 117
gattcaatcg agcca 15

<210> 118
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_53 REV

<400> 118
gagcggatgg cttac 15

<210> 119
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_54 REV

<400> 119
tgtctacgca tgtgt 15

<210> 120

<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_55 REV

<400> 120
cttgaatcca tagca 15

<210> 121
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_56 REV

<400> 121
aactaagcct cgtgg 15

<210> 122
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_57 REV

<400> 122
gcaccttagt tggcc 15

<210> 123
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_58 REV

<400> 123
ttgtcgaata ggtgc 15

<210> 124
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_59 REV

<400> 124
agccgatcgt gtcca 15

<210> 125
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_60 REV

<400> 125
ccgttggaat tatag 15

<210> 126
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_61 REV

<400> 126
cttactaacc acgtt 15

<210> 127
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_62 REV

<400> 127
gccactcgtt ccttg 15

<210> 128
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_63 REV

<400> 128
tcttccgttc tggtg 15

<210> 129
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_64 REV

<400> 129
ttgcttgacc gtacg 15

<210> 130
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide barcode BC_65 REV

<400> 130
ttagcctgtc accga 15

<210> 131
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> synthetic peptide

<220>
<221> MOD_RES
<222> 1
<223> formyl-Methionine

<400> 131
Met Asp Val Glu Ala Trp Leu Gly Ala Arg Val Pro Leu Val Glu Thr
1 5 10 15

<210> 132
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> synthetic peptide

<400> 132
Thr Glu Asn Leu Tyr Phe Gln Asn His Val
1 5 10

<210> 133
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide primer

<400> 133
aatgatacgg cgaccaccga 20

<210> 134
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide primer

<400> 134
caagcagaag acggcatacg agat 24

<210> 135
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide

<220>
<221> misc_feature
<222> (1)...(5)
<223> n = A,T,C or G

<400> 135
nnnnn 5

<210> 136
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> synthetic peptide

<400> 136
Asp Tyr Lys Asp Asp Asp Asp Lys
1 5

<210> 137
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> synthetic peptide

<400> 137
Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr
1 5 10

<210> 138
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> synthetic peptide

<400> 138
Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
1 5 10

<210> 139
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> synthetic peptide

<400> 139
Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1 5

<210> 140
<211> 9
<212> PRT
<213> Artificial Sequence

<220>

<223> synthetic peptide

<400> 140
Asn Trp Ser His Pro Gln Phe Glu Lys
 1   5

<210> 141
<211> 82
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<220>
<221> modified_base
<222> 1
<223> biotin

<400> 141
tttttgcaaa tggcattctg acatcccgta gtccgcgaca ctagatgtct agcatgccgc   60
cgtgtcatgt ggaaactgag tg   82

<210> 142
<211> 72
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<220>
<221> modified_base
<222> 1
<223> biotin

<400> 142
tttttttttt gactggttcc aattgacaag ccgtagtccg cgacactagt aagccggtat   60
atcaactgag tg   72

<210> 143
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<220>
<221> modified_base
<222> 1
<223> biotin

<220>
<221> modified_base
<222> 10
<223> three carbon (3C) spacer

<400> 143
tttttttttt   10

<210> 144
<211> 62

<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<220>
<221> modified_base
<222> (31)...(32)
<223> 18-atom hexa-ethyleneglycol spacer

<220>
<221> modified_base
<222> 62
<223> three carbon (3C) spacer

<400> 144
ggatgtcaga atgccatttg cttttttttt tcactcagtc ctaacgcgta tacgcactca 60
gt 62

<210> 145
<211> 63
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<220>
<221> modified_base
<222> (31)...(32)
<223> 18-atom hexa-ethyleneglycol spacer

<220>
<221> modified_base
<222> 63
<223> three carbon (3C) spacer

<400> 145
ggatgtcaga atgccatttg cttttttttt tcactcagtc ctaacgcgta tacgtcactc 60
agt 63

<210> 146
<211> 63
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<220>
<221> modified_base
<222> (31)...(32)
<223> five 18-atom hexa-ethyleneglycol spacers

<220>
<221> modified_base
<222> 63
<223> three carbon (3C) spacer

<400> 146
ggatgtcaga atgccatttg cttttttttt tcactcagtc ctaacgcgta tacgtcactc 60
agt 63

<210> 147
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<220>
<221> modified_base
<222> (25)...(26)
<223> 18-atom hexa-ethyleneglycol spacer

<220>
<221> modified_base
<222> 63
<223> three carbon (3C) spacer)

<400> 147
gcttgtcaat tggaaccagt cttttcactc agtcctaacg cgtatacggg aatctcggca 60
gttcactcag t 71

<210> 148
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<220>
<221> modified_base
<222> 44
<223> three carbon (3C) spacer

<400> 148
cgatttgcaa ggatcactcg tcactcagtc ctaacgcgta tacg 44

<210> 149
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> spacer sequence

<400> 149
actgagtg 8

<210> 150
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> spacer sequence

<400> 150
cactcagt 8

<210> 151
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide primer

<400> 151
cgtagtccgc gacactag 18

<210> 152
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide primer

<400> 152
cgatttgcaa ggatcactcg 20

<210> 153
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 153
gcaaatggca ttctgacatc c 21

<210> 154
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 154
gactggttcc aattgacaag c 21

<210> 155
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<220>
<221> modified_base
<222> 23
<223> three carbon (3C) spacer

<400> 155
cgtatacgcg ttaggactga gtg 23

<210> 156
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<220>
<221> modified_base
<222> 38
<223> three carbon (3C) spacer

<400> 156
aactgccgag attcccgtat acgcgttagg actgagtg 38

<210> 157
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide primer

<400> 157
agtccgcgca atcagatgtc tagcatgccg gatccggatc gatctc 46

<210> 158
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide primer

<400> 158
agtccgcgca atcagccgtg tcatgtggaa gatccggatc gatctc 46

<210> 159
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide primer

<400> 159
agtccgcgca atcagtaagc cggtatatca gatccggatc gatctc 46

<210> 160
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide primer

<400> 160

agtccgcgca atcagttcga tatgacggaa gatccggatc gatctc 46

<210> 161
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide primer

<400> 161
tgcaaggatc actcgccaga tcacaagcgg gagatcgatc cggatc 46

<210> 162
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide primer

<400> 162
tgcaaggatc actcgtccgg tacgctatct gagatcgatc cggatc 46

<210> 163
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide primer

<400> 163
tgcaaggatc actcggatga tgagcctgga gagatcgatc cggatc 46

<210> 164
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide primer

<400> 164
tgcaaggatc actcgttggc tctagtactc gagatcgatc cggatc 46

<210> 165
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide primer

<400> 165
aatcgtagtc cgcgcaatca g 21

<210> 166

<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide primer

<400> 166
acgatttgca aggatcactc g 21

<210> 167
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> spacer sequence

<400> 167
gatccggatc gatctc 16

<210> 168
<211> 734
<212> DNA
<213> Artificial Sequence

<220>
<223> extended recording tag construct

<400> 168
aatcacggta caagtcactc atccgtacgc tatctgagaa tcgtccagat ccggcatgct 60
agtatctggt gcagactacg attgttacag atcactcaga tgatgagcac agaaaatcgt 120
cgaatcttcc atcaccatcg aacagttacg attaatgtag tccgcacaat cgaatgtcta 180
acatgccgaa tcccggacgt ctccagcttc taaaccaaca gtagtcgcac aaatcattgt 240
acggtacaag atctaacgag agatgatcgg atctgaccac tttaaacact gattacgcag 300
actacgatta cgatttaaga atcctcgtcc ggtacaatca tagtccgcac aatcaaccgt 360
gtcatgtgaa gatcagatcg atctcgaata gcgtaccaga cagtgatctt gcaaatcgta 420
atgtgtccgc gccaatcgat agccatgaat cccagtcgat ctcccgcttg tgatctggcg 480
atcgccttgt accgtcgtac gatttgagat cacctcgtta actcaagcta aagatcgtcc 540
ggatcgcttt ataaacatct gattgcgcgg tacgattatc gtagtccgca catatcgaac 600
ctgttgaaga tccggatcgt ctctccaggc tcatcatccg agtgatcctt gcaaataatc 660
atgtccgcac catcaggtgt ctaacgcttg ccggatccga atcgatctct ccaggctcat 720
catcgaagtg atgt 734

<210> 169
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> synthetic peptide

<400> 169
Cys Pro Val Gln Leu Trp Val Asp Ser Thr
1   5   10

<210> 170
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> synthetic peptide

<220>
<221> VARIANT
<222> (1)...(10)
<223> Xaa = Any Amino Acid

<400> 170
Cys Pro Xaa Gln Xaa Trp Xaa Asp Xaa Thr
 1   5   10

<210> 171
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> FLAG epitope peptide

<400> 171
Asp Tyr Lys Asp Asp Asp Asp Lys
 1   5

<210> 172
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> V5 epitope peptide

<400> 172
Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr
 1   5   10

<210> 173
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> c-Myc epitope peptide

<400> 173
Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
 1   5   10

<210> 174
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> HA epitope peptide

<400> 174
Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
 1   5

<210> 175
<211> 14
<212> PRT

<213> Artificial Sequence

<220>
<223> V5 epitope peptide

<400> 175
Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr
1   5   10

<210> 176
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> StrepTag II peptide

<400> 176
Asn Trp Ser His Pro Gln Phe Glu Lys
1   5

<210> 177
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic nucloetide

<220>
<221> misc_feature
<222> (11)...(22)
<223> compartment bar code n = A, C, T, or G

<220>
<221> misc_feature
<222> (23)...(27)
<223> unique molecular identifier n = A, T, C or G

<400> 177
gcgcaatcag nnnnnnnnnn nnnnnnntgc aaggat 36

<210> 178
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic nucloetide

<220>
<221> misc_feature
<222> (1)...(12)
<223> Compartment barcod n = A, T, C or G

<400> 178
nnnnnnnnnn nn 12

<210> 179
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic nucloetide

<220>
<221> misc_feature
<222> (1)...(5)
<223> Unique molecular identifier; n = A, T, C or G

<400> 179
nnnnn 5

<210> 180
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> butelase I peptide substrate

<400> 180
Cys Gly Gly Ser Ser Gly Ser Asn His Val
1 5 10