



(51) International Patent Classification:

C12Q 1/48 (2006.01) G01N 33/68 (2006.01)

(21) International Application Number:

PCT/GB2019/051842

(22) International Filing Date:

28 June 2019 (28.06.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

1810697.1 29 June 2018 (29.06.2018) GB

(71) Applicant: **IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE** [GB/GB]; South Kensington Campus, Faculty Building, Exhibition Road, London SW7 2AZ (GB).

(72) Inventors: **DIMAGGIO, Peter**; c/o Imperial College of Science, Technology and Medicine, South Kensington Campus, Faculty Building, Exhibition Road, London SW7 2AZ (GB). **KARUNAKARAN, Kalesh**; c/o Imperial College of Science, Technology and Medicine, South Kensington Campus, Faculty Building, Exhibition Road, London SW7 2AZ (GB). **LUKAUSKAS, Saulius**; c/o Imperial College of Science, Technology and Medicine, South Kensington Campus, Faculty Building, Exhibition Road, London SW7 2AZ (GB).

(74) Agent: **HGF LTD**; Cavendish House, 39-41 Waterloo Street, Birmingham West Midlands B2 5PP (GB).

(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, JO, 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, 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).

Published:

— with international search report (Art. 21(3))

(54) Title: ADP-RIBOSYLATION PROFILING

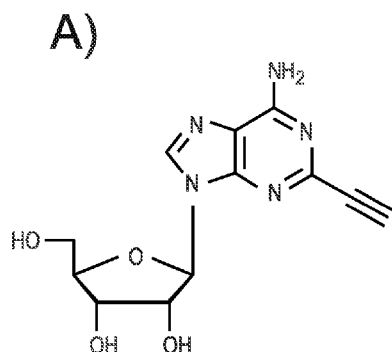


Fig. 2

(57) Abstract: The present invention relates to a method of identifying one or more adenosine diphosphate (ADP) ribosylated proteins, the method comprises applying 2-substituted adenosine, preferably 2-alkyne adenosine, to a first system for the ribosylation of proteins. The invention relates to a kit of parts and a composition comprising 2-alkyne adenosine.



## ADP-RIBOSYLATION PROFILING

### FIELD OF THE INVENTION

- 5 The present invention relates to ADP-ribosylation profiling in systems, in particular, although not exclusively, in cells.

### BACKGROUND TO THE INVENTION

Protein ADP-ribosylation is a post-translational modification (PTM) where members of a family of enzymes known as ADP-ribosyltransferases (or PARPs) covalently link ADP-ribose  
10 moieties derived from NAD<sup>+</sup> to the side chains of a diverse range amino acid residues (glutamate, aspartate, arginine, lysine, cysteine, serine) on target proteins. This PTM exists either as a monomer (mono-ADP-ribosylation), consisting of a single ADP-ribose moiety on the target residue, or polymer (poly-ADP-ribosylation) where a branched polymeric chain of ADP-ribose units originates from the target residue. Of the 17 PARPs encoded in the human  
15 genome, PARPs 1, 2 and 3 are known to have important roles in DNA regulation, particularly in the base excision repair pathway (Hottiger, 2015), and clinical PARP inhibitors have been developed as cancer therapeutics to exploit the concept of synthetic lethality in patients with germline mutations in BRCA1 or BRCA2 genes (Lord, 2017). Beyond the established roles of these nuclear PARPs in DNA damage response, the broad cellular functions for the majority  
20 of the other members of the PARP family remain elusive primarily due to the lack of technologies for the large-scale profiling of intracellular ADP-ribosylation.

Mass spectrometry (MS) proteomics studies of ADP-ribosylation have been limited due numerous technical challenges, including the lack of specificity and/or affinity when using recombinant macro domains (Jungmichel, 2013) (Martello, 2016), cis-diol binding boronate  
25 beads (Zhang, 2013), antibodies (Gagne, 2008) or adapted phosphoproteome enrichment strategies (Daniels, 2014) to enrich for this low abundance PTM against complex protein backgrounds. Additionally, most methods reported to date require *a priori* knockdown of poly-ADP-ribose glycohydrolase (PARG) activity to sufficiently increase the baseline level of poly-ADP-ribosylation for MS detection. However, this approach is confounded by PARG  
30 knockdown inducing physiological changes in cells (Min, 2010). A further limitation to many previous studies is that they have been performed under stress induction which induces PARP1 activation and thus potentially masks the underlying activities of the other PARP enzymes.<sup>3-7</sup> Lastly, although chemical biology approaches based on the incorporation of a “clickable” handle onto the NAD<sup>+</sup> cofactor have emerged as a promising alternative for

selective and bioorthogonal enrichment of ADP-ribosylated proteins (Gibson, 2016), the lack of cell permeability of NAD<sup>+</sup> has restricted these studies to cell lysates.

Recently, it was demonstrated that “clickable” NAD<sup>+</sup> can be metabolically produced by treating cells with the corresponding “clickable” adenosine precursor (Westcott, 2017). Specifically, 6-alkyne adenosine (6YnAd), a compound that was previously demonstrated to be suitable for labelling poly(A) tails of mRNAs in mammalian cells (Grammel, 2012), was utilised for highly sensitive fluorescence detection and Label Free Quantification (LFQ) proteomic profiling of ADP-ribosylated proteins in live cells under stress induction<sup>10</sup>. Although a few proteomic studies have reported ADP-ribosylation on the scale of several hundred cellular protein targets<sup>3-7</sup>, all such studies were performed under stress induction which induces PARP1 activation and thus potentially masks the underlying activities of the other PARP enzymes.

It is an object of this invention to enable the quantitative profiling of ADP-ribosylation for several thousands of endogenous proteins in an unbiased manner, and preferably without stress induction.

## SUMMARY OF THE INVENTION

Accordingly, the present invention provides a method of identifying one or more adenosine diphosphate (ADP) ribosylated proteins comprising applying 2-substituted adenosine, for example 2-alkynyl adenosine, say 2-alkyne adenosine, to a first system for the ribosylation of proteins. Preferably, the first system for the ribosylation of proteins is a cell, such as a cell *in vitro*.

Preferably the methods of the invention further comprise applying a 6-substituted adenosine, for example 6-alkynyl adenosine, say 6-alkyne adenosine, to a second system for the ribosylation of proteins. Preferably, the second system for the ribosylation of proteins is a cell, such as a cell *in vitro*. The 6-substituted adenosine, e.g. 6-alkynyl adenosine such as 6-propargyl adenosine, may, for example, be applied separately, simultaneously or sequentially with the 2-substituted adenosine, e.g. 2-alkynyl adenosine such as 2-ethynyl adenosine, to the same system, for example to the same collection of cells *in vitro*, or the 6-substituted adenosine may, for example, be applied in a parallel or duplicate system to the 2-substituted adenosine, for example to a collection of cells *in vitro* that have been obtained, prepared and cultured under the same conditions as the cells to which the 2-substituted adenosine is applied. Alternatively, the 6-substituted adenosine, may, for example, be applied to a different system to the 2-substituted adenosine, for example the first system can be a particular cancer cell line *in vitro* and the second system can be a different cancer cell line *in vitro* or the same cancer cell line *in vitro* in a different state (e.g. different passage, different stage of growth, etc).

The adenosine analogues used in the invention preferably include alkynyl groups, and therefore allow the use of Click chemistry to manipulate, process and identify the proteins in which the analogues are incorporated following ribosylation. The substituted adenosines used in the invention are preferably 2-alkynyl adenosine and/or 6-alkynyl adenosine. The alkynyl groups of the 2- and 6- substituted adenosines are preferably independently selected from a C2 – C10 alkynyl, and further preferably independently selected from a C2-C5 alkynyl. For example, the alkynyl groups of the 2- and 6- substituted adenosines may be independently selected from ethynyl, butynyl (e.g. 1- or 3-butynyl), 1-propynyl, propargyl, and pentynyl. Preferably the 2-substituted adenosine will have a terminal alkyne moiety at the 2-position (e.g. 2-propynyl, 3-butynyl etc.). The 2-substituted adenosine may be 2-ethynyl adenosine (see Figure 2A) and the 6-substituted adenosine may be 6-propargyl adenosine (see Figure 2B). Other alkynyl substituents may be deployed at the 2 and 6 positions respectively.

Therefore, the methods of the invention may further comprise using a capture reagent to selectively enrich for the ribosylated proteins, for example to enrich for or extract the ribosylated proteins from other system components. The capture reagent may be added to the system after the 2-substituted adenosine, such as 2-alkynyl adenosine, and/or the 6-substituted adenosine, such as 6-alkynyl adenosine, and the capture reagent may be specifically added to ribosylated proteins through the use of click chemistry and the alkyne group of the adenosine analogues.

The capture reagent will have an affinity for a or the capture matrix. The capture reagent preferably comprises a clickable group, for example an azide moiety. The capture reagent may be a di- or tri-functional capture reagent. The capture reagent may comprise one or more of biotin, a biotin analogue (for example desthiobiotin) or an immunotherapy tag (e.g. an immunotherapy fluorescent tag). The capture reagent may comprise 5-carboxytetramethylrhodamine (TAMRA) and/or polyethylene glycol. The capture reagent may be selected from 5/6-TAMRA-LysAzide Biotin, 5/6-TAMRA-Azide Biotin (both available from Jena BioScience of Jena Germany, Cat. No. CLK 1048) or Azide-PEG3-Biotin (CAS Number 875770-34-6).

Preferably the enrichment, isolation or extraction of the ribosylated proteins occurs through the use of an interactor molecule, a molecule that specifically interacts with, preferably binds, the capture reagent. For example, in particularly preferred methods of the invention the enrichment or extraction occurs through the use of, for example, as an interactor molecule one or more of avidin, NeutrAvidin or streptavidin, which is able to interact with and/or bind to, say, biotin present in the capture reagent, such as through the use of interactor molecule-coated stationary medium, such as beads, which stationary medium will consequently bind, say, biotin that is attached to the ribosylated proteins due to the inclusion of the capture

reagent in the proteins; the non-ribosylated proteins, which do not incorporate the adenosine analogue, can be washed away from the stationary medium leaving the ribosylated proteins selectively attached.

Preferred methods of the invention further comprise reducing the size of ADP-ribose modifications added to the proteins. Said reduction of the size may occur, for example, through cleavage by chemical treatment or enzyme treatment. The enzyme treatment may be Nudix treatment or PARG treatment. Once the size of the added group has been reduced, the site of ADP-ribosylation, or location within the protein, of one or more of the ribosylated proteins may be identified.

Generally, one or more labels may be added to the ADP-ribosylated proteins in the methods of the invention. The labels may be added to the ribosylated proteins via the alkyne group of the incorporated adenosine analogues, using click chemistry, and/or the labels may be added by other means, for example following enrichment, isolation or extraction of the ribosylated proteins. In particularly preferred methods of the invention, the labels added to the ribosylated proteins are Tandem Mass Tag (TMT) labels and/or SILAC labels. Alternatively, other quantitative proteomic methodology may be employed, for example iTRAQ, dimethyl labelling and so on. We prefer to use TMT labels because that methodology provides a signal enhancement effect in MS2 that uniquely enables detection of even low-abundant modified proteins.

In particularly preferred methods of the invention, one or more of the ADP ribosylated proteins are characterized or identified using mass spectrometry (MS) and/or liquid chromatography–mass spectrometry (LC-MS).

Methods of the invention may further comprise using a debiased linear analysis to analyse the data output from the LC-MS/MS.

Aspects of the invention may provide a method of identifying one or more adenosine diphosphate (ADP) ribosylated proteins comprising applying a first adenosine analogue and a second, different, adenosine analogue to a system, *e.g.* a cell *in vitro* or *in vivo*. The first and second adenosine analogues may comprise adenosine with substituents at different positions. In particular, the first adenosine analogue may be a 2-substituted adenosine, *e.g.* 2-alkynyl adenosine, for example 2-alkyne adenosine, and the second adenosine analogue may be a 6-substituted adenosine, *e.g.* 6-alkynyl adenosine, such as 6-alkyne adenosine.

Further aspects of the invention provide a composition for identifying one or more adenosine diphosphate (ADP) ribosylated proteins comprising a 2-substituted adenosine, *e.g.* 2-alkyne adenosine and a kit of parts for identifying one or more adenosine diphosphate (ADP) ribosylated proteins comprising a first composition comprising a 2-substituted adenosine, *e.g.*

2-alkynyl adenosine or 2-alkyne adenosine, and a second composition comprising a 6-substituted adenosine, e.g. 6-alkynyl adenosine e.g. 6-alkyne adenosine. A further composition of the invention includes two differently substituted adenosines for identifying ADP ribosylated proteins. The compositions of the invention and kit of parts of the invention may be for separate, simultaneous or sequential administration of different adenosines, e.g. 2-substituted adenosine and 6-substituted adenosine, to first and second systems, for example in accordance with the methods of the invention.

### **BRIEF DESCRIPTION OF THE FIGURES**

The invention will now be described by way of example only and with reference to the accompanying figures, in which:

**Figure 1** shows an overview of ADP-ribosylation by PARP;

A) A summary of the ADP-ribosylation of target protein by PARP enzymes;

B) Nomenclature of the PARP family of proteins and their cell activities.

**Figure 2** shows further details of structures involved in the methods of the invention;

A) Chemical structure of the NAD precursor 2Yn-Ad (2-alkyne adenosine).

B) Chemical structure of the NAD precursor 6YnAd (6-alkyne adenosine).

C) Schematic representation of Poly[ADP-Ribosylation] of a protein;

**Figure 3** shows a schematic overview of a method in accordance with the invention;

**Figure 4** shows general structures of labelled ribosylated proteins generated in methods of the invention, using the 2YnAd probe;

A) Labelled Mono[ADP-Ribosylated] protein;

B) Labelled Poly[ADP-Ribosylated] protein;

**Figure 5** shows general structures of labelled ribosylated proteins generated in methods of the invention, using the 6YnAd probe.

A) Labelled Mono[ADP-Ribosylated] protein following metabolic incorporation of 6YnAd probe and click chemistry.

B) Labelled Poly[ADP-Ribosylated] protein following metabolic incorporation of 6YnAd probe and click chemistry;

**Figure 6** shows chemical structures of click capture reagents that may be used in methods of the invention;

**Figure 7** shows metabolic labelling of MDA-MB-231 breast cancer cells by 6Yn-Ad and 2Yn-Ad.

A) Fluorescence labelling of ADP-ribosylated proteins.

B) Coomassie blue staining of total proteins;

5 **Figure 8** shows a schematic overview of quantitative profiling of global ADP-ribosylome in MDA-MB-231 breast cancer cell line using a combination of metabolic incorporation of the clickable NAD precursors (6YnAd and 2YnAd) and TMTsixplex isobaric mass tagging.

10 A) MDA-MB-231 cells were metabolically labelled with 3 different concentrations of 6YnAd and 2YnAd in parallel. After cell lysis and click reaction with a biotin-containing capture reagent, the labelled proteins are affinity enriched on NeutrAvidin-Agarose resins. Stringent washing at this stage removes non-modified proteins from the beads. After on-bead reduction, alkylation and tryptic digestion, the eluted peptides are desalted and each sample is labelled using a unique TMTsixplex channel.

15 B) After quenching the TMT reactions, the 6 samples are mixed together, cleaned up, fractionated and subjected to Nanoflow LC-MS/MS analysis. Quantification of reporter ions from the MS<sup>2</sup> spectra provides relative quantification of labelling efficiencies of the two compounds at all three tested concentrations;

**Figure 9** shows a comparison of metabolic labelling of 6YnAd and 2YnAd, according to the schematic overview of Figure 8.

20 A) Heatmap indicating mass spectrometry intensities across all three concentrations after subtracting bias terms (e.g. replicate bias and intercept).

25 B) Volcano plot between LIMMA coefficient values (positive values on the x axis indicate preference for 2YnAd and negative values indicate preference for 6YnAd). Adjusted p values (-log<sub>10</sub> scale) from LIMMA analysis are plotted on Y axis. The proteasomal subunits and DNA replication licensing factors specifically enriched for the 2YnAd in the biological pathway analysis are highlighted in the plot and

30 **Figure 10** shows experimental workflow for the dual metabolic labelling method, with a schematic illustrating the workflow involving parallel treatment of PARP inhibitors Olaparib and Rucaparib followed by dual metabolic labelling of 6YnAd and 2YnAd, click chemistry, affinity enrichment of the labelled proteins on NeutrAvidin-Agarose beads and TMT10plex-based quantification of global changes in the cellular ADP-ribosylome.

## DETAILED DESCRIPTION

Figure 1A shows the ADP-ribosylation of a target protein wherein adenosine binds to a target protein by use of a PARP. More than 15 different PARP enzymes have been identified (Figure 1B), although the majority are poorly characterized and their targets unknown.

The present invention provides simple methods to profile post translational modification (PTM) events that are taking place in a particular system. In particular, the use of the adenosine analogue 2-alkyne adenosine (Figure 2A) provides the ability to profile live cells to identify the ADP-ribosylation of proteins (Figure 1A). Using adenosine analogues, which are then incorporated into NAD<sup>+</sup> within the cell, surprisingly overcomes the difficulties of introducing NAD<sup>+</sup> analogues into cells. Therefore, the invention provides the ability to generate a profile of the ADP-ribosylation that is occurring, for example in different cells types and/or under different conditions. As well as enabling further elucidation of normal cellular pathways and their control, this allows, in particular, the assessment and identification of new disease-associated pathways and drug targets.

We have found that tolerance for NAD<sup>+</sup> analogues varies such that whilst some proteins will be ribosylated regardless of the analogue used, whether and/or the extent to which others are ribosylated depends on which particular adenosine analogue is present in the system (e.g. 2-alkynyl (e.g. alkyne) adenosine or 6-alkynyl (e.g. alkyne) adenosine). Without wishing to be bound by any theory, it is believed that this is due to preferences among the PARP enzymes for both particular NAD<sup>+</sup> analogues and particular proteins. Therefore, methods in which say 2-alkynyl adenosine is used provide information about additional ADP-ribosylated proteins, compared to information from methods in which, say, 6-alkynyl adenosine is used in the same experimental circumstances and conditions. Also, by providing methods in which both 2-substituted adenosine and 6-substituted adenosine are used, the invention enables the profiling of a wider group of ribosylated proteins in one series of experiments, compared to just using one of the adenosine analogues on their own, for example 6-alkynyl adenosine. Additionally, by comparing the ribosylation that occurs when one adenosine analogue is present (e.g. 2-alkynyl adenosine) to that which occurs when a different adenosine analogue is present (e.g. 6-alkynyl adenosine), it is possible to interrogate the analogue preferences of the system and the underlying causes thereof. Thus, in methods in which both 2-substituted adenosine and 6-substituted adenosine are added to a system, the different analogues may be added to the same system, e.g. cells, at the same time, or they may be added to the same system, e.g. cells, at different times, or they may be added to a parallel system, for example to duplicate cells that have been obtained, prepared and cultured under the same conditions as the cells to which the other adenosine is applied. Alternatively, the 2-substituted adenosine and 6-substituted adenosine may be added to different systems, for example, the first system may be cells from a particular cancer cell line and the second system may be cells from a

different cancer cell line or the same cancer cell line in a different state (e.g. different passage, different stage of growth, etc).

The combination of the dual metabolic incorporation strategy with TMT-based isobaric labelling approach enables comprehensive profiling of cellular ADP-ribosylation. This novel combination provides unprecedented enhancement in the sensitivity of detection of the PTM and provides an opportunity to derive more information about many ADP-ribosylated proteins. Additionally, the TMT-labelling approach is able to be used to increase the sensitivity of detection. We have also been able to demonstrate that it is possible to undergo characterization without induction of cellular stress and knockdown of PARP hydrolyzing enzymes (e.g. PARG) prior to the proteomic profiling of the ADP-ribosylation. PARG knockdown (which is a requirement for existing methods to overcome poor sensitivity) is known to cause alterations in cellular physiology. Also, induction of cellular stress (which is a requirement of the existing methods) leads to preferential in-cell activation of PARP1 which masks activity of other members of the PARP family and severely undermine their study. The invention, because of its high sensitivity, requires neither PARG knockdown/knockout nor induction of cellular stress and as such provides a way for an unbiased comprehensive quantitative profiling of endogenous ADP-ribosylated proteins.

In order to assist the understanding of the present invention, certain terms used herein will now be further defined, and more generally further details of the invention will be given, in the following paragraphs.

### Adenosine Analogues

The invention provides methods comprising adding 2-substituted adenosine, such as 2-alkynyl adenosine (Figure 2A), and optionally 6-substituted adenosine, such as 6-alkynyl adenosine (Figure 2B) to systems. These adenosine analogues can be synthesised using methods known to those skilled in the art; see, for example (Jiang, 2009) and (Wang, 2014), which are incorporated herein by reference.

The adenosine analogues are metabolically incorporated into Nicotinamide Adenosine Dinucleotide (NAD) in the systems used in the methods of the invention, for example in the live cells, such that NAD<sup>+</sup> analogues are produced in the systems, which have the alkynyl group, or “clickable handle” of the corresponding adenosine analogue precursor molecule. The NAD<sup>+</sup> analogues can then be used in the ADP-ribosylation of target protein residues by PARP enzymes (Figures 1A and 1B), resulting in ADP-ribosylated proteins with the alkynyl group “clickable handle” (see Figure 3).

The PARP enzymes may add mono ADP-ribose modifications to their protein target(s), or they may add poly ADP-ribose modifications to their protein target(s) (Figure 1A). The currently understood preferences with regard to individual PARP enzyme ability to add mono or poly groups, is shown in Figure 1B. Therefore, the methods of the invention may comprise the  
5 addition of mono ADP-ribose groups having a 2-alkynyl handle or 6-alkynyl handle, leading to corresponding labelled proteins (see Figures 4A and 5A, respectively, for example), and/or may comprise the addition of poly ADP-ribose groups having a 2-alkynyl handle or 6-alkynyl handle, leading to corresponding labelled proteins (see Figures 4B and 5B, respectively, for example).

10 The alkyne bonds of the analogues allow the use of click chemistry in methods of the invention. "Click chemistry" is a term generally understood in the art. The alkyne bonds provide that the ADP-ribosylation added during the methods of the invention can be conjugated to desirable labels, such as capture reagents, in particular using a Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC).

15 Although we prefer CuAAC Click chemistry, other Click chemistry could be deployed, for example SPAAC, SPANC and strained alkene reactions could each be utilised. In deploying other techniques due notice will be taken of steric issues and the tolerance of the system (for example cellular enzymes within the system) to steric clashes.

### **A system for the ribosylation of proteins**

20 In the methods of the invention, 2-substituted adenosine is added to a first system for the ribosylation of proteins and, optionally, 6-substituted adenosine is added to a second system for the ribosylation of proteins. A system for the ribosylation of proteins comprises the elements necessary to incorporate one of the substituted adenosine analogues of the methods of the invention into NAD<sup>+</sup>, and for one or more PARP enzymes to use the NAD<sup>+</sup> to ADP-  
25 ribosylate one or more target proteins. In particularly preferred methods of the invention, the first, and where present, second systems are cells, preferably live cells. The methods of the invention are preferably carried out *in vitro*.

In methods in which both 2-substituted adenosine and 6-substituted adenosine are added to systems, they may be added to the same systems such that the first and second systems are  
30 one and the same. For example, the two adenosine analogues may be added to the same tissue culture of cells. In this way, the addition of both analogues allows the detection of a wider range of ADP-ribosylated proteins than would be detected by the use of just one of the analogues on their own, in otherwise similar circumstances.

Alternatively, the first and second systems may be different but deliberately similar, such that the only differences identified in the systems using the method are due to whether it was the 2-substituted adenosine, such as 2-alkynyl adenosine, or 6-substituted adenosine, such as 6-alkynyl adenosine, that was added to the system. For example, the first and second systems  
5 may be cultures of the same type(s) of cell, such as HeLa or MB-231 breast cancer cells, where the culture has and is being grown under identical conditions, so that any difference in ribosylation pattern can be associated with a preference within the system for either the 2-substituted adenosine or 6-substituted adenosine analogue.

Of course, as is routine in scientific endeavour, the methods may comprise performing  
10 replicates, such that the 2-substituted adenosine and/or 6-substituted adenosine are added to multiple systems, and some of the systems do not differ with respect to the conditions of the systems, including which and how much analogue is added to them (see Figure 8). Alternatively, or additionally, experiments may be carried out to investigate the effect of different conditions on ADP-ribosylation patterns, so that most of the conditions of the systems  
15 are kept the same, but particular conditions, such as the concentration of the analogue or a particular drug or reagent in the systems, are deliberately altered in a measured way (see Figures 8 and 9). Conditions that may be assessed in this way, to analyse the effects on ADP-ribosylation, include PARP inhibitor treatments (see Figure 10), oxidative stress induction, cytotoxic/chemotherapy agent treatment, viral/bacterial infection, and hypoxia induction.

## 20 Identifying adenosine diphosphate (ADP) ribosylated proteins

The methods of the invention are for identifying ADP-ribosylated proteins. Such “identification” should be interpreted broadly in the context of the invention. Therefore, for example, identification may simply comprise noting the presence of ribosylated proteins, e.g. seeing proteins in which the alkyne handle is incorporated following the addition of an adenosine  
25 analogue (Figure 7A). Alternatively, identification in accordance with the invention may comprise analyzing and characterizing specific ribosylated proteins, for example by assigning the correct name to the proteins (Figure 9).

## Labels

30 Methods of the invention may comprise adding one or more labels or tags to the ADP-ribosylated proteins. The one or more labels or tags may be any suitable label or tag, for example for visualising, identifying, or capturing one or more of the ADP-ribosylated proteins.

In preferred methods of the invention, a label or tag is added to the ADP-ribosylated protein using click chemistry and, say, an alkyne clickable handle.

The label or tag may be added to the proteins through the incorporation of a capture reagent. For example, click enrichment may be carried out in methods of the invention by the addition  
5 of di- or tri-functional capture reagent carrying azide chemical functionality and biotin (and a fluorescent dye unit in the case of the tri-functional capture reagent) to the treated cells. The click reaction ensures chemical tagging/labelling of the ADP-ribosylated proteins/peptides with the capture reagent, which can be selectively enriched (i.e. purified) using its biotin group on modified Avidin-based beads (exploiting the strong biotin-Avidin interaction so that extensive  
10 wash steps can remove unmodified proteins). Suitable capture reagents therefore include 5/6-TAMRA-LysAzide-Biotin, 5/6-TAMRA-Azide-Biotin, and Azide-PEG3-Biotin (Figure 6).

Methods of the invention may comprise the addition of Tandem Mass Tag (TMT) labels and/or the use of Stable Isotope Labeling with Amino acids in Cell culture (SILAC) (Harsha, 2008). A preferred method using TMT labels in accordance with the invention is illustrated in Figure  
15 8.

More than one tag or label may be added to the ADP-ribosylated proteins in the methods of the invention, and different labels may be added at different stages of methods of the invention, according to need. For example, a capture reagent may be added to the proteins using click chemistry, in order to extract or enrich for the ribosylated proteins, and the capture  
20 reagent may be removed following the isolation or enrichment and a TMT label added instead to assist with characterisation of the ribosylated proteins by MS and/or LC-MS.

The label or tag may be added to the proteins after preprocessing to provide additional information about the protein and its ribosylation. For example, the system used in the methods of the invention may be a live cell and the label or tag, for example a TMT label, may  
25 be added after cell fractionation of the system(s). In this way, it may be possible to identify the locations of the ADP-ribosylated protein(s) in the live cell.

### **Enrichment, and Extraction**

The methods of the invention may comprise extracting the ADP-ribosylated proteins. The  
30 extraction may be absolutely specific or near absolutely specific, so that substantially no non-ribosylated proteins remain with the ADP-ribosylated proteins, or the extraction may be relative, leading to an enrichment of the ADP-ribosylated proteins, so that the concentration of the ADP-ribosylated proteins significantly increases but some non-ribosylated proteins remain.

Preferably, the extraction occurs using a capture reagent, as discussed above. Thus, specific affinity enrichment can be carried out.

### **Reducing the size of ADP-ribose modifications**

5 It may be preferred in methods of the invention to reduce the size of the ADP-ribose modification, particularly where a poly ADP-ribose chain has been added to the proteins. This may assist with subsequent processing of the protein(s), such as with the analysis of the protein and/or the site of its modification using LC-MS and/or MS.

10 Suitable methods for reducing the size of the modification(s) by cleavage include chemical cleavage, for example using the method of (Zhang, 2013) , or enzymatic cleavage, for example using Nudix treatment or PARG treatment. In particularly preferred methods, all three cleavage strategies (chemical, Nudix and PARG) may be used to ensure comprehensive cleavage.

15 The cleavage may occur immediately following extraction of the ADP-ribosylated proteins, for example whilst the proteins are still attached to avidin beads used for extraction so that the cleavage occurs “on-bead”.

The cleavage results in the presence of modified residues in the proteins, which residues can be detected and identified using MS, so that the site of addition of the ADP-ribose chain can be identified within one or more of the proteins.

### **Mass Spectrometry (MS)**

20 In preferred methods of the invention, MS, and in particular quantitative liquid chromatography-tandem mass-spectrometry (LC-MS/MS) is used to identify the ADP-ribosylated proteins and their sites of modification.

### **Data Analysis**

25 Detecting significant changes in the levels of protein ADP-ribosylation under non-stressed conditions requires a statistical test that is both powerful and robust. We have adapted the empirical Bayes method LIMMA for this purpose, as it utilizes the data across all proteins to shrink the sample variance to a pooled estimate. We have used a method similar to that referred to in the literature (Kammers, 2010). Our linear models include a term to account for bias between different replicates arising from the stochastic nature of MS/MS sampling by the mass  
30 spectrometer so this bias could be excluded from the analysis. These linear models can also be customized to directly capture the response arising from different features in the experimental

design, such as the comparison between two or more enzyme inhibitors over multiple concentrations, time points or varying biological conditions of the samples.

## **EXPERIMENTAL RESULTS**

### **5    1.        Materials and Methods**

#### **Cell culture**

MDA-MB-231 cells (ATCC) were grown in Dulbecco's Modified Eagle Medium (DMEM) (Gibco) supplemented with 10% foetal bovine serum (FBS) (Gibco) without antibiotics in a humidified incubator with 5% CO<sub>2</sub> at 37°C and passages at every 2-3 days at 80-90%  
10    confluency.

#### **Metabolic labelling and preparation of cell extracts.**

In a 6-well plate, 4 x 10<sup>5</sup> cells were treated with 1 ml of fresh media (with 10% FBS) containing 2 µl of appropriate concentration of 6YnAd or 2YnAd (stock solutions in DMSO) or DMSO control. For labelling experiments in 10 cm plates, 5 ml of fresh media with 10% FBS and 10  
15    µl of the clickable adenosine analogue stock solutions were used. The cells were incubated at 37 °C and 5% CO<sub>2</sub> for 1 h. The cells were then washed with PBS three times and lysed using 4% SDS lysis buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 4% SDS and 500 U Benzonase) and whole-cell lysates were collected using a cell scraper. Lysates were cleared at 16,000g for 5 min and protein concentrations were estimated using Bio-Rad DC™ protein  
20    assay.

#### **Click chemistry and in-gel fluorescence detection.**

The cell lysates at 1 mg/ml concentration were treated with freshly premixed click chemistry reaction cocktail (100 µM capture reagent, 1 mM CuSO<sub>4</sub> solution, 1 mM tris(2-carboxyethyl) phosphine hydrochloride) (TCEP), 100 µM [tris(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA)) for 3 h at room temperature. The following stock solutions of the click chemistry reagents were used. Capture reagent (Azido-TAMRA-Biotin or Azide-PEG3-Biotin (Jena Bioscience)): 10 mM in DMSO; CuSO<sub>4</sub>: 50 mM in MilliQ water; TCEP: 50 mM in MilliQ water; TBTA: 10 mM in DMSO. Proteins were precipitated by adding ice-cold methanol (4 volumes), chloroform (1.5 volumes) and water (3 volumes) and the precipitates were collected by  
25    centrifugation at 16,000g for 5 min. The protein precipitates were then washed twice with ice-cold methanol (10 volumes) and the supernatants were discarded. The protein pellets were  
30

air dried for 20 min and suspended in resuspension buffer (4% SDS, 50 mM HEPES pH 7.4, 150 mM NaCl) to 3 mg/ml concentration. Equal volumes of 2X SDS loading buffer (Bio-Rad) containing 10%  $\beta$ -mercaptoethanol were added and the samples were boiled at 95 °C for 10 min and allowed to cool to room temperature. 30  $\mu$ g of protein was loaded per gel lane (12% SDS Tris-HCl gels) and resolved by SDS-PAGE. The gels were scanned for fluorescence labelling using GE typhoon 5400 gel scanner.

#### **Affinity enrichment on NeutrAvidin-Agarose.**

The dried protein pellets obtained after click chemistry and protein precipitation were suspended in resuspension buffer (4% SDS, 50 mM HEPES pH 7.4, 150 mM NaCl) to 10 mg/ml concentration. In a typical experiment, 500  $\mu$ g of whole-cell lysate after click chemistry and protein precipitation was resuspended in 50  $\mu$ l of the resuspension buffer. The samples were subsequently diluted 20 fold with HEPES buffer (50 mM HEPES pH 7.5, 150 mM NaCl) so that the final SDS amount is 0.2%. 50  $\mu$ l of NeutrAvidin-Agarose beads (Thermo scientific) were washed three times with 0.2% SDS buffer (0.2% SDS 50 mM HEPES pH 7.5, 150 mM NaCl) and added to the samples. The mixtures were rotated on an end-over-end rotating shaker for 1.5 h at room temperature. The beads were then washed 4 times with 1% SDS buffer (1% SDS, 50 mM HEPES pH 7.5, 150 mM NaCl), 4 times with 5M urea buffer (5M urea in 50 mM HEPES pH 7.5, 150 mM NaCl) and 4 times with 50 mM freshly prepared ammonium bicarbonate buffer. Each washing was performed with 20 volumes of the washing solutions and centrifugation of the beads between washing steps were carried out at 2,000g for 1 min.

#### **On-bead reduction, alkylation, tryptic digestion and desalting.**

The beads after affinity enrichment were resuspended in 100  $\mu$ l of 50 mM ammonium bicarbonate buffer and treated with 3 mM dithiothreitol (DTT) (100 mM stock solution in water) solution for 30 min at room temperature. The beads were washed once with 50 mM ammonium bicarbonate buffer and resuspended in 100  $\mu$ l of 50 mM ammonium bicarbonate buffer and treated with 10 mM iodoacetamide (IAA) (500 mM stock solution in water) in dark for 30 min at room temperature. The beads were again washed with 50 mM ammonium bicarbonate buffer and resuspended in 100  $\mu$ l of fresh 50 mM ammonium bicarbonate buffer and treated with 2  $\mu$ g of trypsin (Promega, Product Code V5113) at 37 °C overnight. The samples were acidified to pH 3 using formic acid and allowed to stand for 5 min, centrifuged and collected the supernatant. The beads were washed with 0.1% formic acid solution in water, centrifuged and the supernatant were mixed with the previous supernatant. The collected tryptic peptides were desalted on C18 Empore disks on micropipette tip format and the desalted peptide mixture were dried on a speedvac.

**TMTsixplex labelling.**

MDA-MB-231 cells were fed with three different concentrations (1, 0.5 and 0.25 mM) of 6YnAd and 2YnAd in parallel. After 1 h of incubation, whole-cell lysates were prepared and proteins were quantified. 500 µg of proteins from each condition were subjected to click chemistry using the capture reagent Azido-TAMRA-Biotin. Proteins were precipitated, re-solubilised and subjected to affinity enrichment on NeutrAvidin-Agarose beads. After extensive washings, the beads were subjected to DTT treatment, IAA treatment and overnight trypsinisation. The peptide digests were collected, desalted, dried and re-dissolved in 100 µl of 100 mM triethylammonium bicarbonate (TEAB) buffer (pH 8.5). Each sample was subjected to TMT labelling with a unique TMTsixplex reagent for 1 h at room temperature. After quenching the TMT reactions with hydroxylamine (10 µL of 5% hydroxylamine per reaction for 15 min at room temperature) treatment, the six samples were mixed together, concentrated to complete dryness and re-dissolved in 0.1% formic acid. The sample was desalted on C18 Empore disks on micropipette tip format and the desalted sample was dried on a speedvac.

**Mass-spectrometry analysis of TMT labelled samples**

Peptides were reconstituted in 0.1% trifluoroacetic acid and chromatographically resolved using an Ultimate 3000 RSLCnano (Dionex) UHPLC. Peptides were first loaded onto an Acclaim PepMap 100 C18, 3 µm particle size, 100Å pore size, 20 mm x 75 µm ID (Thermo Scientific) trap column heated to 40°C, using a loading buffer (2% acetonitrile (MeCN) and 0.05 % trifluoroacetic acid in 97.05 % H<sub>2</sub>O) with a flow rate of 7 µL/minute. Chromatographic separation was achieved using an EASY-Spray column, PepMap C18, 2 µm particles, 100 Å pore size, 500 mm x 75 µm ID (Thermo Scientific) heated to 40 °C. The gradient utilised a flow of 0.3 µL/minute, starting at 98 % mobile A (0.1 % formic acid, 5% dimethyl sulfoxide (DMSO) in H<sub>2</sub>O) and 2 % mobile B (0.1 % formic acid, 75 % acetonitrile, 5% DMSO and 19.9 % H<sub>2</sub>O). After 5.5 minutes mobile B was increased to 8 % over 0.5 minutes, to 30 % over 27 minutes, further increased to 40 % in 10 minutes, and finally up to 95 % in 2 minutes and held for 4 minutes. Mobile B was reduced back to 5 % over 1 minute for the rest of the acquisition.

Data was acquired in real time over 60 minutes using an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific). Eluents were ionised using 2.1kV spray voltage and a capillary temperature of 275 °C. For all samples MS1 spectra were measured with a resolution of 120000 at 200 *m/z* and a mass range of 375 to 1500 *m/z*, with a 50 ms maximum injection time, and 4e5 AGC target.

High-resolution HCD MS2 spectra were generated in positive mode for precursor masses using a normalised collision energy of 38 %, within a 0.7 *m/z* isolation window, using

quadrupole isolation, in top speed mode over a 3 second cycle time, at 50000 resolution in profile mode, 100 ms maximum injection time and 5e4 AGC target, with an instrument determined scan range beginning at 100 *m/z*.

- 5 To encourage quality peptide fragmentation, a number of filters were utilised, including; peptide mono isotopic precursor selection, minimum intensity exclusion of 5e4, and charge state inclusion of 2-5 ions. To prevent repeat sampling a dynamic exclusion with exclusion count 1, exclusion duration of 30 seconds, +/- 10 ppm tolerance window, with isotope exclusion and single charge state selection was used.

### Mass-spectrometry data processing

- 10 All raw files were analysed using MaxQuant software (version 1.5.3.30) queried against the Uniprot complete human database (October 2017). The following search parameters were used: Reporter ion MS2 with multiplicity 6plex TMT for the TMTsixplex experiments and 10plex TMT for the TMT10plex experiments, trypsin digestion with maximum 2 missed cleavages, oxidation of methionine and acetylation of protein N-termini as variable modifications, carbamidomethylation of cysteine as a fixed modification, minimum peptide length of 6, protein FDR 0.01. Appropriate correction factors for the individual TMT channels for both peptide N-terminal labelling and lysine side-chain labelling as per the TMT kits used (Thermo scientific) were configured into the integrated Andromeda search prior to the MaxQuant search.

### PARP inhibitor treatment, dual metabolic incorporation and TMT10plex labelling

- 20 MDA-MB-231 cells grown to 80-90% confluency in 10 cm plates were treated with 25, 5, 1, 0.2 and 0 (DMSO control)  $\mu$ M concentrations of the PARP inhibitors Olaparib (10 mM stock solution in DMSO) and Rucaparib (10 mM stock solution in DMSO) in parallel. All treatments were performed in 10 ml DMEM with 10% FBS and the cells were placed in an incubator with 5% CO<sub>2</sub> and 37 °C for 1 h. After 1 h, the media were removed and the cells were supplemented with fresh complete media (5 ml) containing 0.25 mM concentrations of both 6YnAd (500 mM stock solution in DMSO) and 2YnAd (500 mM stock solution in DMSO) continued incubation for one more hour. The cells were washed with PBS three times and whole-cell lysates were prepared and proteins were quantified. 500  $\mu$ g protein per each condition were subjected to click reactions using Azide-PEG3-Biotin capture reagent. Proteins were precipitated using the methanol/chloroform/water system and air-dried precipitates were resuspended in the pull-down buffer (50 mM HEPES, pH 7.4, 0.2% SDS, 150 mM NaCl) and subjected to affinity enrichment on NeutrAvidin-Agarose beads. After extensive washings, the beads were subjected to DTT treatment, IAA treatment and overnight trypsinisation. The peptide digests from the 10 conditions were collected, desalted, dried and re-dissolved in 100  $\mu$ l of 100 mM

triethylammonium bicarbonate (TEAB) buffer (pH 8.5). Each sample was subjected to TMT labelling with a unique TMT10plex reagent for 1 h at room temperature. After quenching the TMT reactions with hydroxylamine treatment, the 10 samples were mixed together, concentrated to complete dryness and re-dissolved in 0.1% formic acid solution in water. The sample was desalted on C18 Empore disks on micropipette tip format and the desalted sample was dried on a speedvac. The sample was dissolved in 300 µl of 0.1% TFA solution in water and subjected to fractionation using High pH Reverse-Phase Peptide Fractionation Kit (Thermo scientific) to 8 fractions and dried on a speedvac.

### Preprocessing of MaxQuant output for data analysis

Columns corresponding to the corrected reporter intensity were extracted from the proteinGroups.txt files in MaxQuant output. Potential contaminants, proteins only identified by site, reverse proteins and proteins containing less than two peptides were discarded. Out of the remaining proteins, only proteins that have non-zero intensities across all reporters and replicates were kept (1604 proteins in 2YnAd-6YnAd dataset, 2334 in the inhibitor dataset). Intensities of the remaining proteins were normalised using TMM method (Robinson M., 2010) and converted to log2-per-million values. The resulting data was then compared with previously reported annotations of ADP-ribosylated proteins available in ADPriboDB (v1.0) (Vivelo, 2017), which was augmented to include annotations reported since 2015 (v2.0; unpublished), after prefiltering the database to contain only annotations derived from human datasets. LIMMA linear model design

Both cofactor preference and inhibitor treatment datasets were analysed using LIMMA method. The linear methods for each of these datasets were designed separately and described below.

#### *2YnAD/6YnAD*

The normalised intensities are assumed to be modelled by the following effects in a format described below:

$$\log_2(\text{normalised intensity}) \sim [\text{Cofactor effect, Sum encoding}] + [\log(\text{concentration}) \text{ effect}] + [\text{Replicate effect}] + \text{Intercept}.$$

We used the following encoding for the effects.

1. **Cofactor effect** (i.e. the main measured effect). Encoded by +1 in reporters corresponding to 2YnAd and -1 in reporters corresponding to 6YnAd (e.g. sum-to-one encoding). Positive coefficients to this factor therefore indicate 2YnAd preference, and negative values indicate 6YnAd preference.
2. **log(concentration) effect**: encoded as log values of the concentrations in mM. Effect is assumed to be the same for both cofactors. Higher values indicate stronger response.

3. **Replicate:** 1-of-K encoding (minus one redundant term absorbed in intercept). This term accounts for technical-replicate associated bias in protein intensities.

The complete design matrix is available as Table 1 below.

| Headers  |           |                  | Design matrix |           |                           |                   |                          |
|----------|-----------|------------------|---------------|-----------|---------------------------|-------------------|--------------------------|
| Reporter | Replicate | Concentration_mM | Cofactor      | Intercept | C(Cofactor, Sum)[S.2YnAd] | C(Replicate)[T.2] | np.log(Concentration_mM) |
| 0        | 1         | 1                | 6YnAd         | 1         | -1                        | 0                 | 0                        |
|          | 2         | 1                | 6YnAd         | 1         | -1                        | 1                 | 0                        |
| 1        | 1         | 0.5              | 6YnAd         | 1         | -1                        | 0                 | -0.693147                |
|          | 2         | 0.5              | 6YnAd         | 1         | -1                        | 1                 | -0.693147                |
| 2        | 1         | 0.25             | 6YnAd         | 1         | -1                        | 0                 | -1.386294                |
|          | 2         | 0.25             | 6YnAd         | 1         | -1                        | 1                 | -1.386294                |
| 3        | 1         | 1                | 2YnAd         | 1         | 1                         | 0                 | 0                        |
|          | 2         | 1                | 2YnAd         | 1         | 1                         | 1                 | 0                        |
| 4        | 1         | 0.5              | 2YnAd         | 1         | 1                         | 0                 | -0.693147                |
|          | 2         | 0.5              | 2YnAd         | 1         | 1                         | 1                 | -0.693147                |
| 5        | 1         | 0.25             | 2YnAd         | 1         | 1                         | 0                 | -1.386294                |
|          | 2         | 0.25             | 2YnAd         | 1         | 1                         | 1                 | -1.386294                |

**Table 1.** Design matrix for 2YnAd/6YnAd dataset

*Inhibitor dataset* For this dataset we have applied the linear modelling only on the datapoints corresponding to 25µM of inhibitor and 0 µM (reporters 0, 4, 5 and 9) only, not taking the data for intermediate concentrations into account. The following linear model was assumed.

$$\text{Log}(\text{intensity}) \sim [\text{Inhibitor preference effect}] + [\text{Inhibition effect}] + [\text{Replicate bias}] + [\text{Intercept}].$$

5 The three factors have been encoded as follows:

1. **Inhibitor preference effect.** One of the two major effects in this experiment. This factor is designed to capture the effect of a particular inhibitor on the intensity of the protein. We encoded this using the following map: -1 = Olaparib, +1 = Rucaparib and 0 = No inhibitor (i.e. concentration = 0 µM). In this way, positive values correspond to protein having to higher intensity upon Rucaparib treatment and negative values corresponding to higher intensity upon Olaparib treatment.

- 2. **Inhibition effect.** This term corresponds to the second major effect in the experiment, which captures the general effect of inhibition to the protein that does not depend on the inhibitor. It was encoded as 1 (in cases where inhibitor is present) and 0 otherwise. This way values below than zero correspond to decrease in intensity upon inhibition.
- 3. **Replicate bias:** This term accounts for technical-replicate associated bias in protein intensities. This was encoded with one-in-K encoding with one redundant term absorbed to Intercept.

The complete design matrix is available in Table 2 below.

| Headers  |           |           | Design matrix    |           |                  |                 |                                        |                             |                    |                    |
|----------|-----------|-----------|------------------|-----------|------------------|-----------------|----------------------------------------|-----------------------------|--------------------|--------------------|
| Reporter | Replicate | Intercept | Concentration_uM | Inhibitor | Inhibitor_binary | Inhibitor_added | C(Inhibitor_binary, contrast)[custom0] | C(Inhibitor_added) [T.True] | C(Replicate) [T.2] | C(Replicate) [T.3] |
| 0        | 1         | 1         | 25               | Olaparib  | Olaparib         | TRUE            | -1                                     | 1                           | 0                  | 0                  |
|          | 2         | 1         | 25               | Olaparib  | Olaparib         | TRUE            | -1                                     | 1                           | 1                  | 0                  |
|          | 3         | 1         | 25               | Olaparib  | Olaparib         | TRUE            | -1                                     | 1                           | 0                  | 1                  |
| 4        | 1         | 1         | 0                | Olaparib  | NA               | FALSE           | 0                                      | 0                           | 0                  | 0                  |
|          | 2         | 1         | 0                | Olaparib  | NA               | FALSE           | 0                                      | 0                           | 1                  | 0                  |
|          | 3         | 1         | 0                | Olaparib  | NA               | FALSE           | 0                                      | 0                           | 0                  | 1                  |
| 5        | 1         | 1         | 25               | Rucaparib | Rucaparib        | TRUE            | 1                                      | 1                           | 0                  | 0                  |
|          | 2         | 1         | 25               | Rucaparib | Rucaparib        | TRUE            | 1                                      | 1                           | 1                  | 0                  |
|          | 3         | 1         | 25               | Rucaparib | Rucaparib        | TRUE            | 1                                      | 1                           | 0                  | 1                  |
| 9        | 1         | 1         | 0                | Rucaparib | NA               | FALSE           | 0                                      | 0                           | 0                  | 0                  |
|          | 2         | 1         | 0                | Rucaparib | NA               | FALSE           | 0                                      | 0                           | 1                  | 0                  |
|          | 3         | 1         | 0                | Rucaparib | NA               | FALSE           | 0                                      | 0                           | 0                  | 1                  |

Table 2. Design matrix for inhibitor dataset.

## LIMMA processing and statistical testing

We have used a similar procedure to the one described in (Kammers, 2010), which is incorporated herein by reference. The linear models were encoded as described above. LIMMA analysis was run with trend parameter set to TRUE (commonly referred to as limma-trend method). After the model has been fit, we have performed statistical testing on the effect coefficients in the model using the equations provided by (Kammers, 2010). Specifically, we have tested the null hypothesis that cofactor effect coefficient is zero in 2YnAd/6YnAd dataset, and the null hypotheses that inhibitor preference effect and inhibition effect coefficients are not zero (separately) in inhibitor treatment dataset. The p values were adjusted with Benjamini-Hochberg method. Any adjusted p values below 0.01 were deemed significant and annotated in the plots and output. In order to generate debiased plots (e.g. Figure 9a), we have subtracted the values for bias coefficients (e.g. Replicate bias and Intercept) from the normalised log2 intensities.

## 2. A dual metabolic labelling strategy for tagging intracellular ADP-ribosylation.

We aimed to further develop the metabolic labelling strategy using clickable NAD<sup>+</sup> precursors since this approach has precedence in the study of mRNA using cell models and facilitates labelling of the ADP-ribosylome directly in the native cellular environment. Although successful metabolic labelling of the ADP-ribosylome has been demonstrated using 6YnAd, previous studies have shown that the corresponding clickable NAD<sup>+</sup> cofactor (6-alkyne-NAD) is a slightly weaker substrate of PARP1 than endogenous NAD<sup>+</sup> (Jiang, 2010). This raises the question as to whether the placement of the alkyne group on the NAD<sup>+</sup> cofactor analogue affects binding and utilisation by the different PARP enzymes.

We decided to investigate if complementary labelling of the ADP-ribosylome could be achieved using a combination of 2YnAd and 6YnAd. We observed via fluorescent detection that 2YnAd is also efficiently utilised for the labelling of ADP-ribosylated proteins with similar qualitative profiles to that of the 6YnAd (Fig. 7). To assess how the alkyne handle on the 2 or 6 position of NAD<sup>+</sup> potentially affects PARP utilisation and hence downstream coverage of protein targets, we performed a proteomic comparison of the differential labelling between 2YnAd and 6YnAd using quantitative mass spectrometry based on tandem mass tags (TMT)-based isobaric tagging (Dayon, 2008). The TMT labelling strategy offers increased sample multiplexing to enable the head-to-head comparison of the labelling between 2YnAd and 6YnAd at different concentrations or conditions, and the isobaric nature of the tag results in the pooling of peptide signal across all experimental conditions, which increases the sensitivity

in detecting lower abundance ADP-ribosylated proteins relative to any non-specific background. Using the MDA-MB-231 breast cancer cell line as a model system, we developed a workflow (Fig. 8) that combines (i) metabolic incorporation of 2YnAd or 6YnAd, (ii) bioorthogonal tagging of the ADP-ribosylated proteome using Copper (I) catalysed azide-alkyne cycloaddition (CuAAC) click reaction, (iii) affinity enrichment of the labelled proteome using the robust biotin-NeutrAvidin pull-down system, (iv) TMTsixplex isobaric labelling and (v) LC-MS/MS analysis.

This chemical proteomics workflow for profiling intracellular ADP-ribosylation events identified over 1600 protein targets, representing a significant improvement in proteomic coverage compared to other published studies despite requiring less protein input (500µg total protein input before affinity enrichment per condition). Importantly, statistical analysis of the replicate data sets using LIMMA (Ritchie, 2015) revealed that specific subsets of protein targets were preferentially labelled with either 2YnAd or 6YnAd treatment (Fig. 9a and b), thus supporting our hypothesis that endogenous PARP enzymes may elicit differential structural preference for the NAD<sup>+</sup> cofactor analogue depending upon the position of the clickable alkyne handle.

### 3. The differential response of the ADP-ribosylome to PARP inhibition.

The observed quantitative differences in protein targets labelled by either 2YnAd or 6YnAd can be exploited for more comprehensive profiling of the cellular ADP-ribosylome by simultaneously treating cells with both compounds. As several PARP inhibitors have recently emerged as promising cancer therapeutics, we utilised this dual metabolic labelling strategy to comprehensively and quantitatively profile global changes to the cellular ADP-ribosylome upon PARP inhibitor treatment (Fig. 10). Specifically, we treated MDA-MB-231 cells with equimolar concentrations of 2YnAd and 6YnAd to investigate differential labelling of protein ADP-ribosylation for five different concentrations (0, 0.2, 1, 5, 25 µM) of the clinical PARP1 inhibitors Olaparib and Rucaparib. The samples were combined in a TMT10plex experiment and over 2300 proteins were identified which, we believe, is the largest reported data set representing the cellular ADP-ribosylome to date.

Statistical analysis of the replicate data sets using LIMMA revealed a strongly correlative response in the levels of ADP-ribosylation with respect to Olaparib and Rucaparib treatment, which indicates a congruent set of PARP1 substrates. Importantly, the majority of the proteins identified were not found to change significantly after the 1 hour treatment with either PARP inhibitor, thus highlighting the robustness and reproducibility of this chemical proteomics workflow.

#### 4. Conclusion

In summary, we have developed a robust and sensitive workflow for the large-scale quantitative profiling of cellular ADP-ribosylation. The lack of a sensitive proteomic method for profiling cellular ADP-ribosylome under non-stress conditions has hampered the functional studies of ADP-ribosylation beyond the established roles of the PTM in DNA regulation. The dual metabolic incorporation approach in combination with TMT isobaric tagging provided the much-awaited technological platform for comprehensive quantitative proteomic profiling of cellular ADP-ribosylome under non-stress conditions. Our method revealed that ADP-ribosylation is widespread among cellular proteins than previously thought. While the use of NeutrAvidin-Agarose beads enabled robust and reproducible enrichment of ADP-ribosylated proteins, an alternative peptide-level enrichment workflow using antibiotin antibody was also developed for characterisation of modifications sites of ADP-ribosylation. We envision the use of the integrated chemical proteomic technology described herein for functional studies of ADP-ribosylation in various contexts such as regulation of tissue factor mRNAs stabilities during inflammation, cancer research, bacterial infections, and viral infections.

## References

1. Hottiger, M. O. *Annu. Rev. Biochem.* **84**, 227-263 (2015).
2. Lord, C. J. & Ashworth, A. *Science* **355**, 1152-1158 (2017).
3. Jungmichel, S. *et al. Mol. Cell* **52**, 272-285 (2013).
4. Martello, R. *et al. Nat. Commun.* **7**, 12917 (2016).
5. Zhang, Y., Wang, J., Ding, M. & Yu, Y. *Nat. Methods* **10**, 981-984 (2013).
6. Gagne, J. P. *et al. Nucleic Acids Res.* **36**, 6959-6976 (2008).
7. Daniels, C. M., Ong, S. E. & Leung, A. K. *J. Proteome Res.* **13**, 3510-3522 (2014).
8. Min, W., Cortes, U., Herceg, Z., Tong, W. M. & Wang, Z. Q. *Carcinogenesis* **31**, 2058-2065 (2010).
9. Gibson, B. A. *et al. Science* **353**, 45-50 (2016).
10. Westcott, N. P., Fernandez, J. P., Molina, H. & Hang, H. C. *Nat. Chem. Biol.* **13**, 302-308 (2017).
11. Grammel, M., Hang, H. & Conrad, N. K. *ChemBioChem* **13**, 1112-1115 (2012).
12. Jiang, H., Kim, J. H., Frizzell, K. M., Kraus, W. L. & Lin, H. *J. Am. Chem. Soc.* **132**, 9363-9372 (2010).
13. Dayon, L. *et al. Anal. Chem.* **80**, 2921-2931 (2008).
14. Ritchie, M. E. *et al. Nucleic Acids Res.* **43**, e47 (2015).
15. Cho-Park, P. F. & Steller, H. *Cell* **153**, 614-627 (2013).
16. Kakarougkas, A. *et al. Nucleic Acids Res.* **41**, 10298-10311 (2013).
17. Haikarainen, T., Krauss, S. & Lehtio, L. *Curr. Pharm. Des.* **20**, 6472-6488 (2014).
18. Knezevic, C. E. *et al. Cell Chem Biol* **23**, 1490-1503 (2016).
19. Bartolomei, G., Leutert, M., Manzo, M., Baubec, T. & Hottiger, M. O. *Mol. Cell* **61**, 474-485 (2016).
20. Thorsell, A. G. *et al. J Med Chem* **60**, 1262-1271 (2017).
21. Robinson, M. D. & Oshlack, A. *Genome Biol.* **11**, R25 (2010).
22. Vivello, C. A., Wat, R., Agrawal, C., Tee, H. Y. & Leung, A. K. *Nucleic Acids Res.* **45**, D204-D209 (2017).
23. Kammers, K., Cole, R. N., Tiengwe, C. & Ruczinski, I. *EuPA Open Proteom* **7**, 11-19 (2015).
24. Fabregat, A. *et al. Nucleic Acids Res.* **46**, D649-D655 (2018).
25. Yu, G. & He, Q. Y. *Mol. BioSyst.* **12**, 477-479 (2016).
26. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. *OMICS.* **16**, 284-287 (2012).
27. H. Jiang *et al. J. Am. Chem. Soc.*, **131**, 1658-1659 (2009).
28. Y. Wang *et al. Angew. Chem. Int. Ed.*, **53**, 8159-8162 (2014).
29. Harsha HC *et al. Nature Protocols.* **3** (3): 505-16 (2008).

**CLAIMS**

1. A method of identifying one or more adenosine diphosphate (ADP) ribosylated proteins comprising applying 2-substituted adenosine to a first system for the ribosylation of proteins.
2. A method according to claim 1, wherein the 2-substituted adenosine is a 2-alkynyl adenosine, and preferably wherein the 2-substituted adenosine is a 2-alkyne adenosine.
3. A method according to claim 1 or claim 2, wherein the method further comprises applying 6-substituted adenosine, preferably a 6-alkynyl adenosine, and further preferably a 6-alkyne adenosine, to a second system for the ribosylation of proteins.
4. A method according to claim 3, comprising applying the 2-substituted adenosine to cells *in vitro* as the first system and comprising applying the 6-substituted adenosine to cells *in vitro* as the second system.
5. A method according to claim 3 or 4, wherein the first and second systems are the same systems or duplicates of the same system.
6. A method according to any preceding claim, further comprising adding a capture reagent for selectively enriching the ribosylated proteins.
7. A method according to claim 6, wherein the capture reagent is a di- or tri-functional capture reagent.
8. A method according to claim 6 or claim 7, wherein the capture reagent is selected from 5/6-TAMRA-LysAzide Biotin, 5/6-TAMRA-Azide Biotin or Azide-PEG3-Biotin.
9. A method according to any of claims 6 to 8, further comprising using the added capture reagent to extract the ribosylated proteins.
10. A method according to claim 9, wherein the extraction occurs through the use of a molecule that binds the capture reagent.
11. A method according to claim 10, wherein the extraction occurs through the use of one or more of avidin, streptavidin, NeutrAvidin which binds to biotin present in the capture reagent, for example through the use of avidin coated beads.
12. A method according to any preceding claim, further comprising reducing the size of the ADP-ribose modifications added to the proteins.
13. A method according to claim 12, wherein said reduction in size occurs through cleavage by chemical treatment or enzyme treatment.
14. A method according to claim 13, wherein said enzyme treatment is Nudix treatment or PARG treatment.

15. A method according to any of claims 12 to 14, further comprising identifying the site of ADP-ribosylation of one or more of the ribosylated proteins.
16. A method according to any preceding claim, further comprising adding one or more labels to the ADP ribosylated proteins.
- 5 17. A method according to claim 16 wherein the labels are Tandem Mass Tag (TMT) labels and/or SILAC labels.
18. A method according to any preceding claim further comprising identifying one or more of the ADP ribosylated proteins using mass spectrometry (MS) and/or liquid chromatography–mass spectrometry (LC-MS).
- 10 19. A method according to claim 18, further comprising using a debiased linear analysis to analyse the data output from the LC-MS/MS.
20. A composition for identifying one or more adenosine diphosphate (ADP) ribosylated proteins, the composition comprising 2-alkyne adenosine, preferably 2-alkynyl adenosine, and 6-alkyne adenosine, preferably 6-alkynyl adenosine, for separate, simultaneous or sequential  
15 administration to first and second systems.
21. A kit of parts for identifying one or more adenosine diphosphate (ADP) ribosylated proteins, the kit comprising a first composition comprising 2-alkynyl adenosine, preferably 2-alkyne adenosine, and a second composition comprising 6-alkynyl adenosine, preferably 6-alkyne adenosine.

1/11

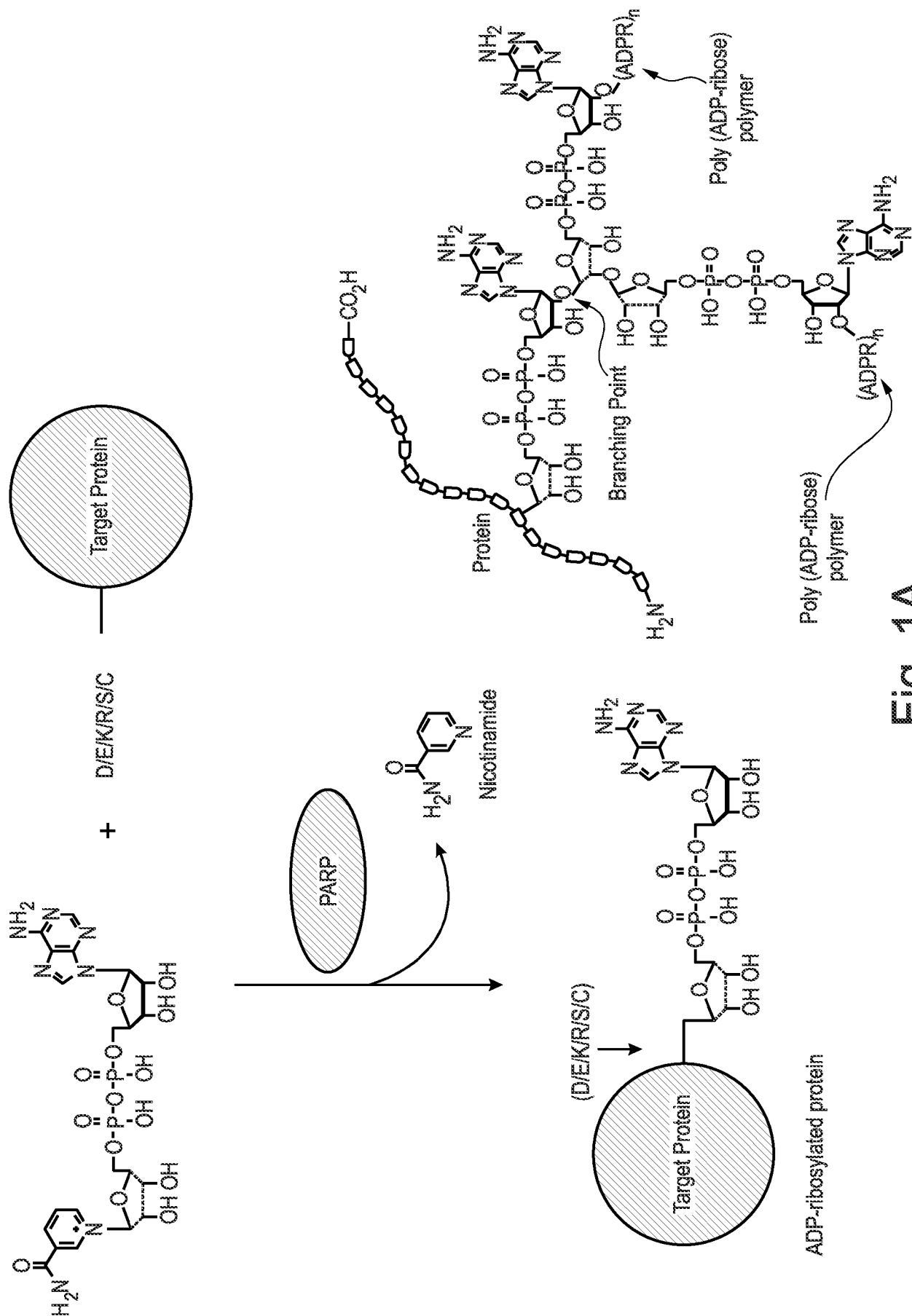


Fig. 1A

2/11

| <u>Gene name</u> | <u>Alternative<br/>name(s)</u> | <u>Enzymatic<br/>activity</u> |
|------------------|--------------------------------|-------------------------------|
| PARP1            | ARTD1                          | Poly-                         |
| PARP2            | ARTD2                          | Poly-                         |
| PARP3            | ARTD3                          | Mono-                         |
| PARP4            | ARTD4, vaultPARP               | Poly-                         |
| PARP5a           | ARTD5, Tankyrase 1             | Poly-                         |
| PARP5b           | ARTD6, Tankyrase 2             | Poly-                         |
| PARP6            | ARTD17                         | Mono-                         |
| PARP7            | ARTD14, TIPARP                 | Mono-                         |
| PARP8            | ARTD16                         | Mono-                         |
| PARP9            | ARTD9, BAL1                    | No activity detected          |
| PARP10           | ARTD10                         | Mono-                         |
| PARP11           | ARTD11                         | Mono-                         |
| PARP12           | ARTD12                         | Mono-                         |
| PARP13           | ARTD13, ZAP                    | No activity detected          |
| PARP14           | ARTD8, BAL2                    | Mono-                         |
| PARP15           | ARTD7, BAL3                    | Mono-                         |
| PARP16           | ARTD15                         | Mono-                         |

Fig. 1B

3/11

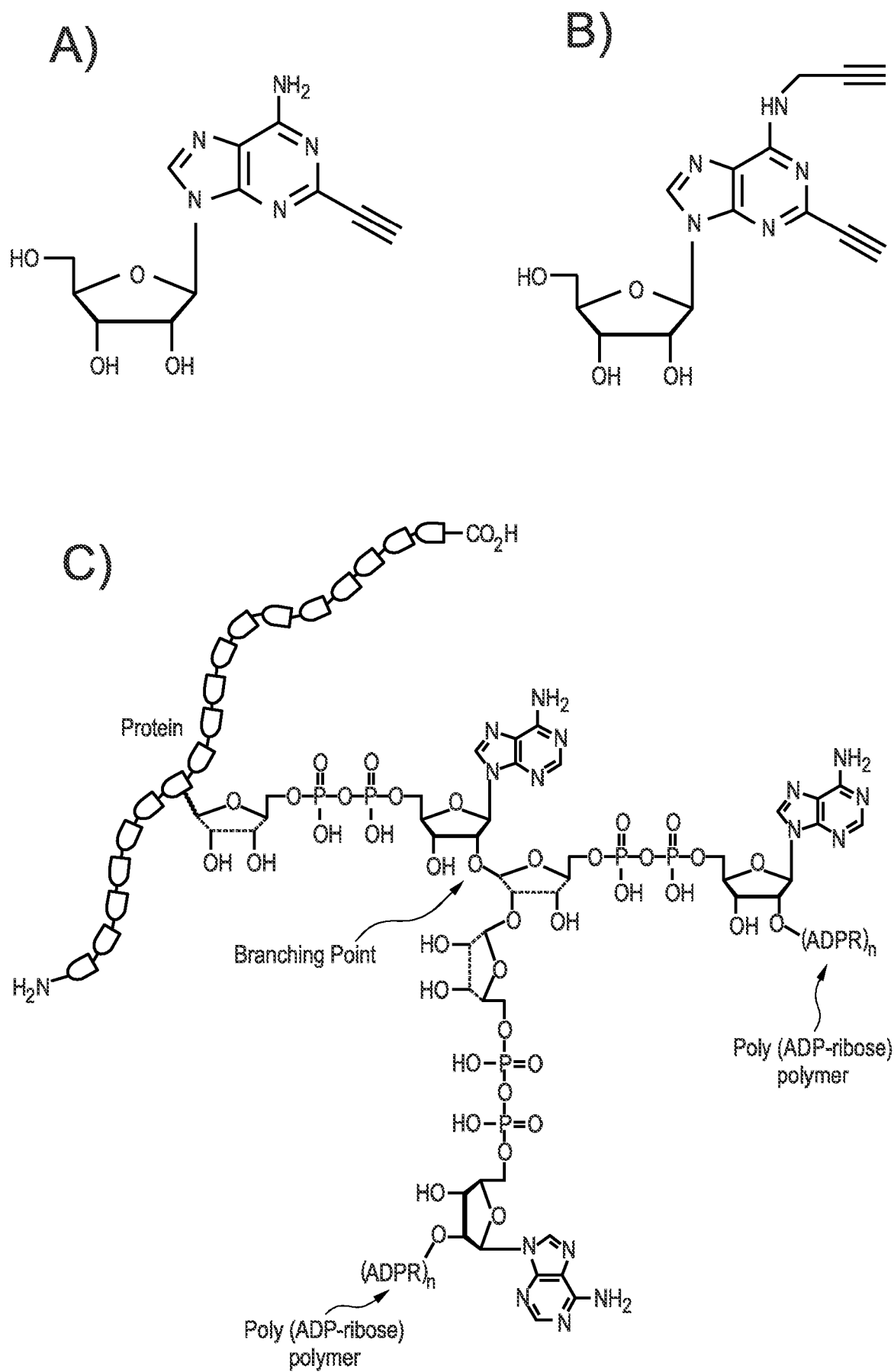


Fig. 2

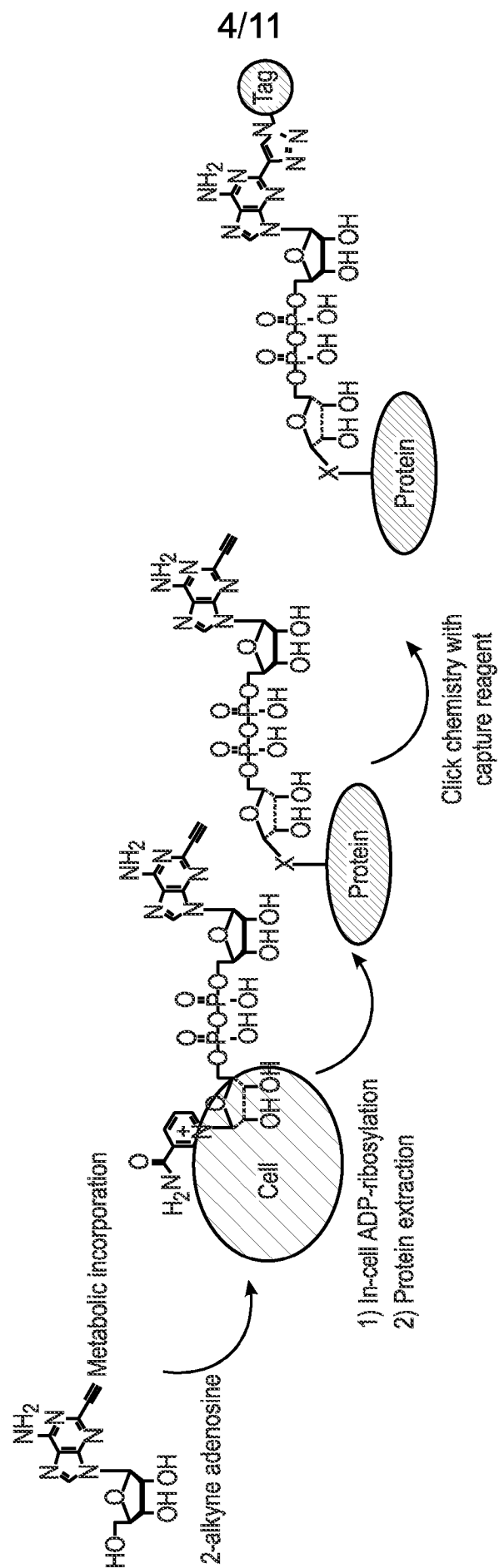


Fig. 3

5/11

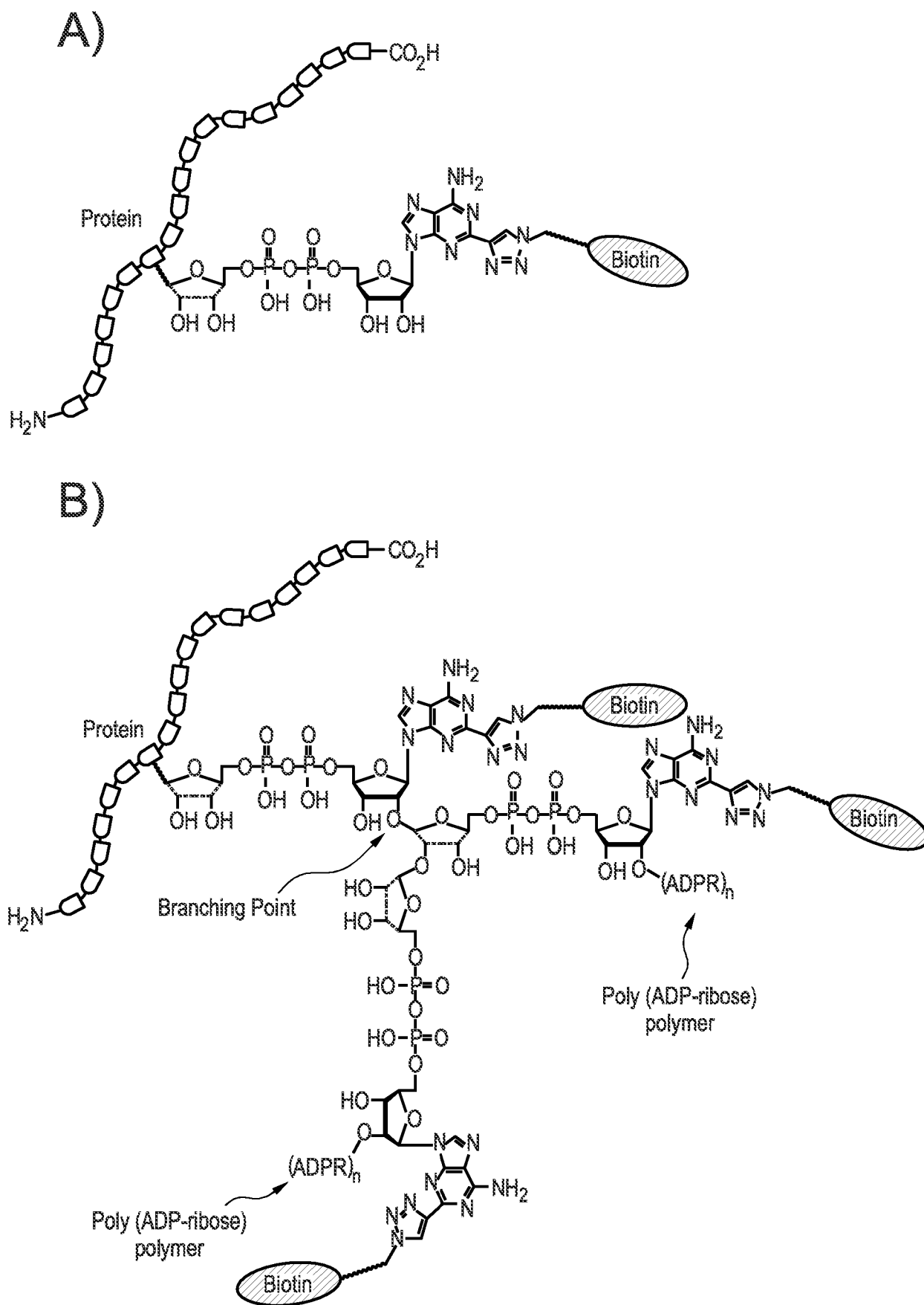


Fig. 4

6/11

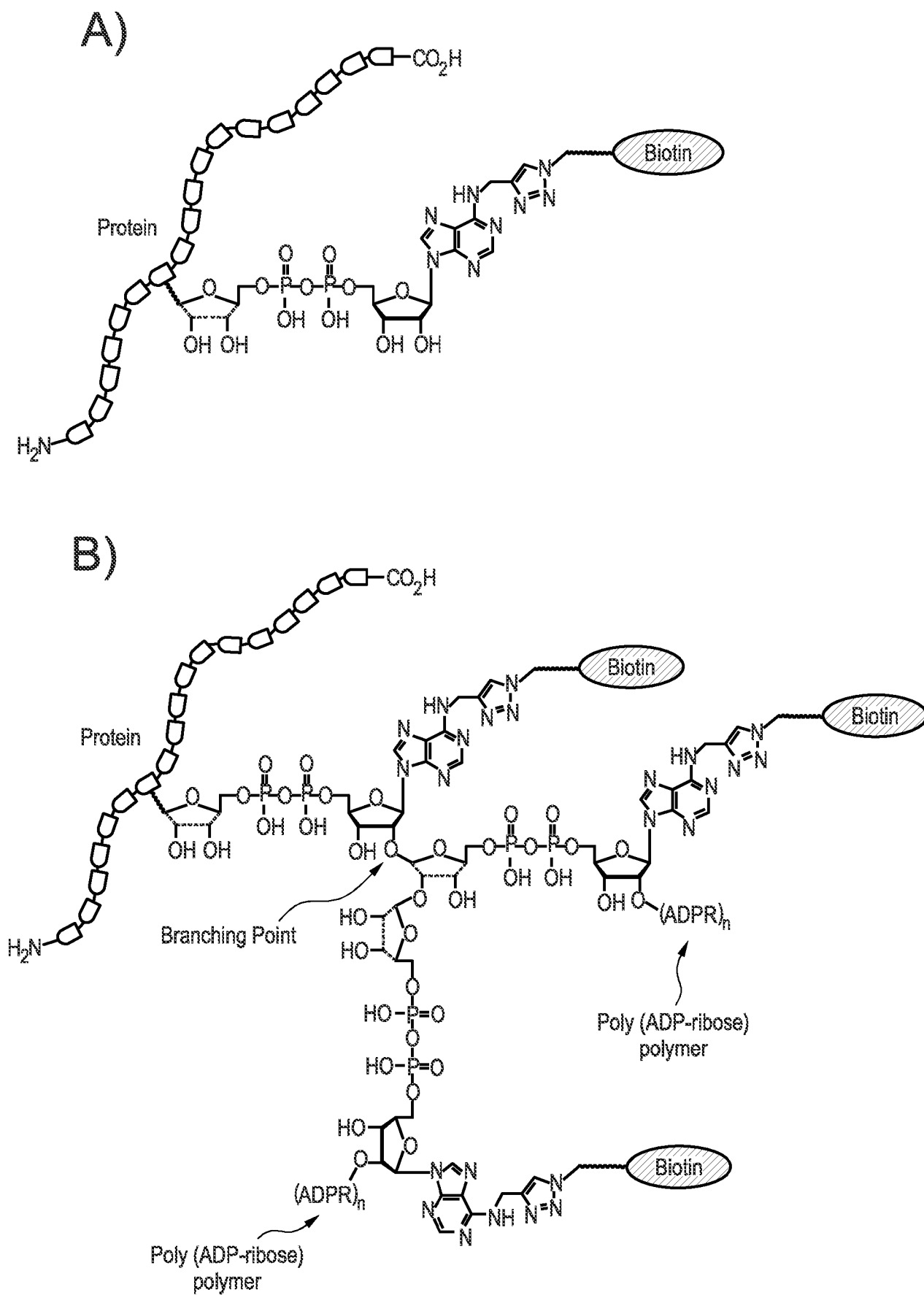


Fig. 5

7/11

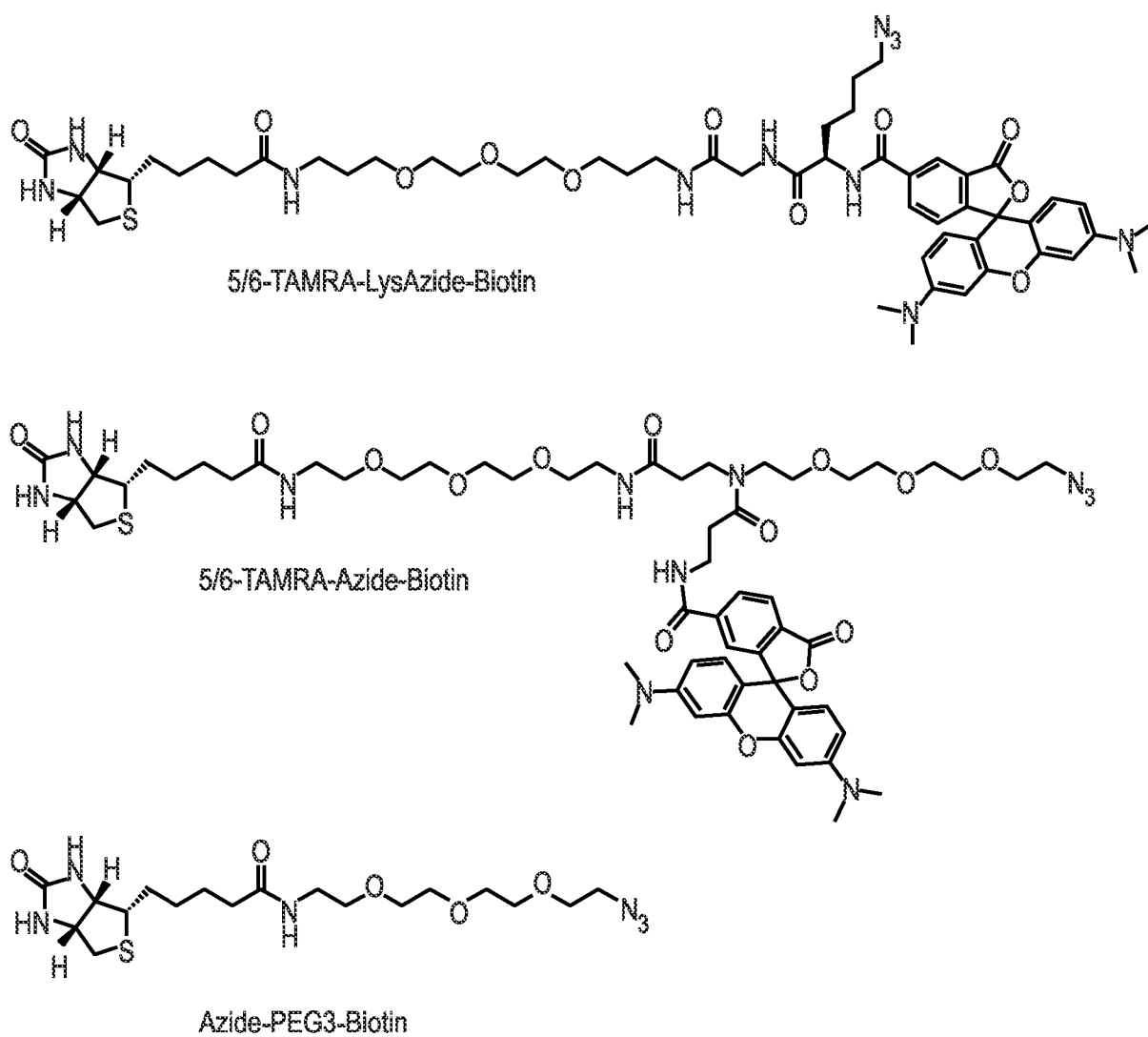


Fig. 6

8/11

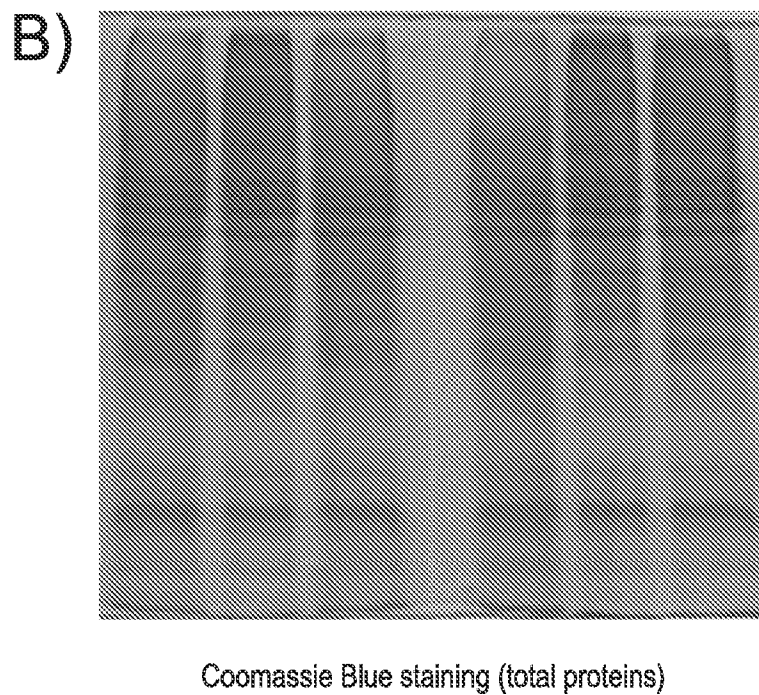
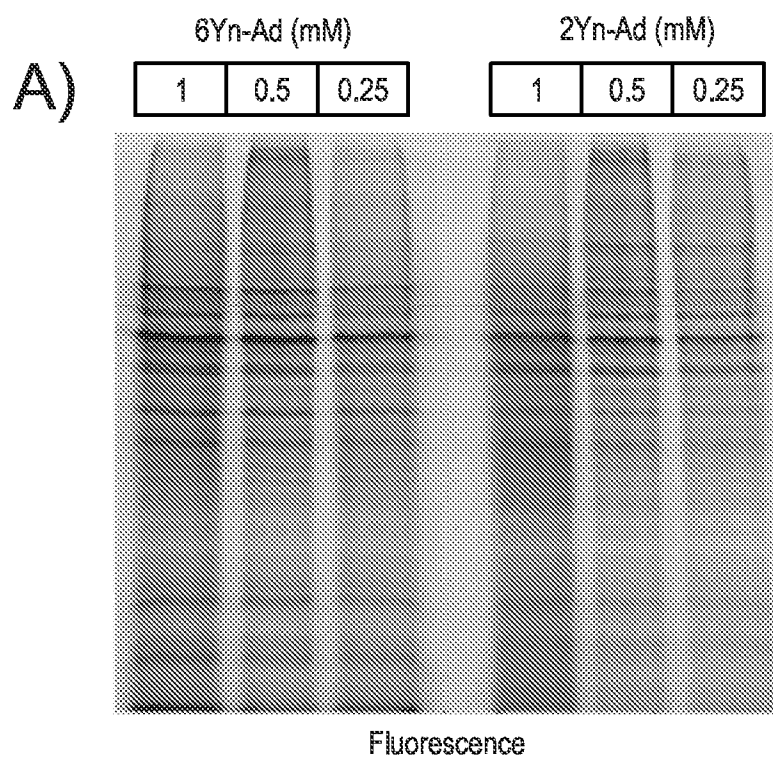


Fig. 7

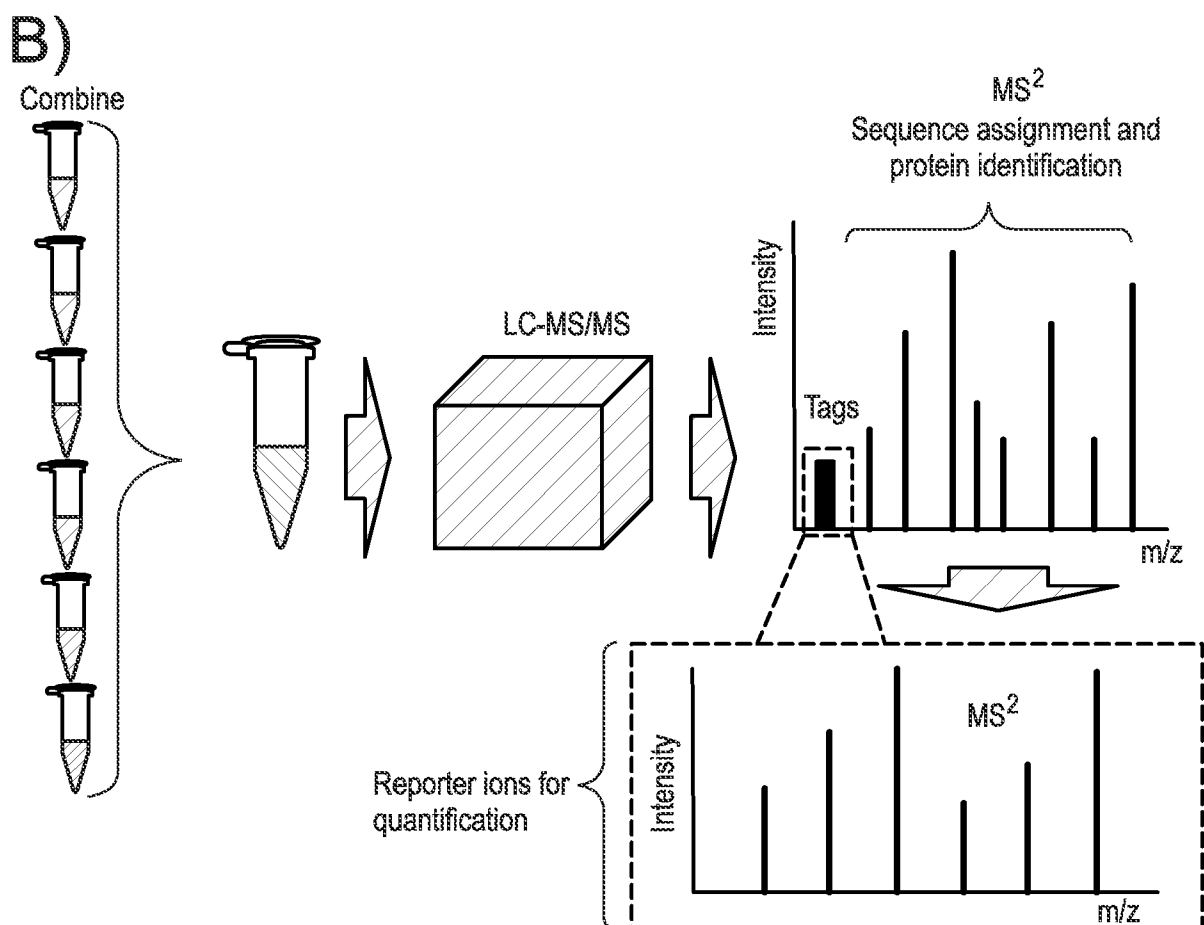
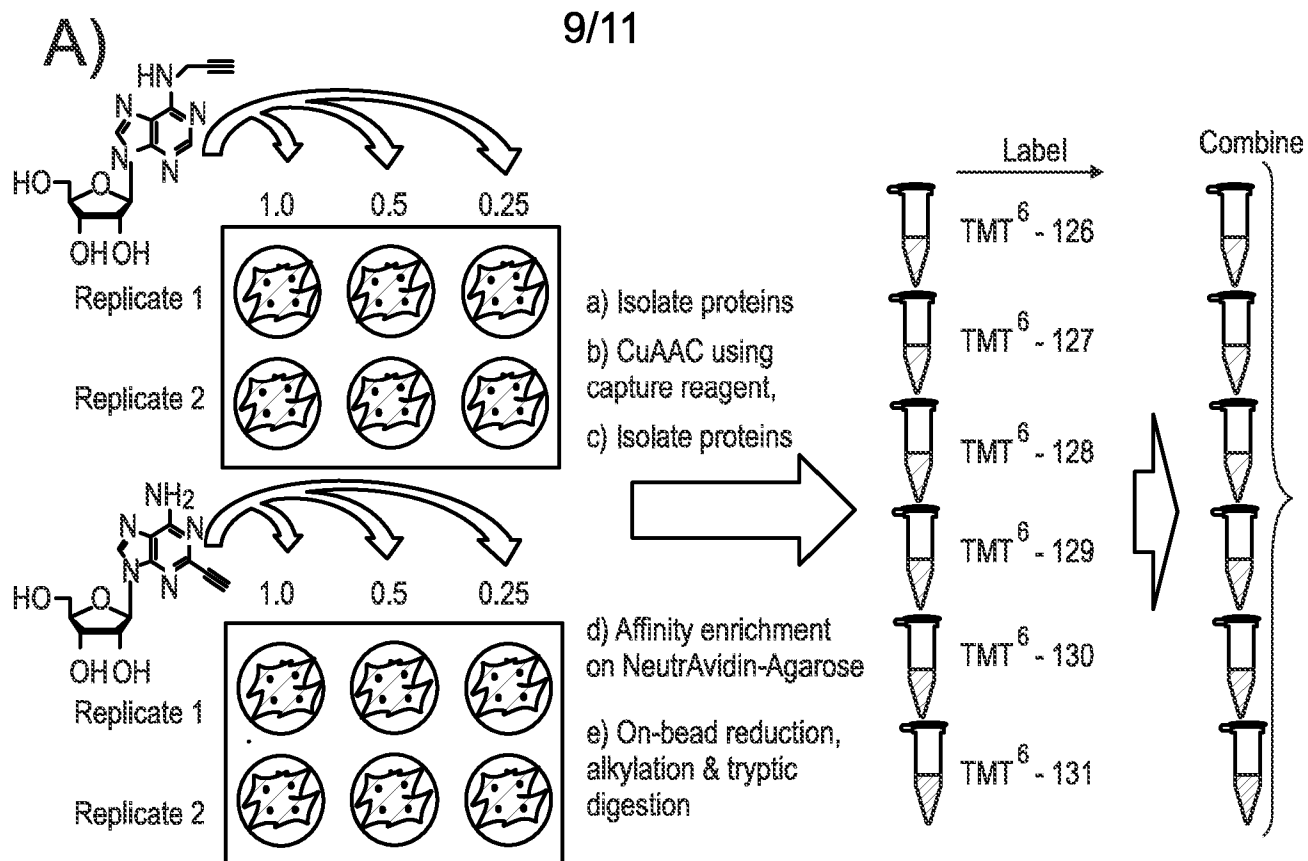


Fig. 8

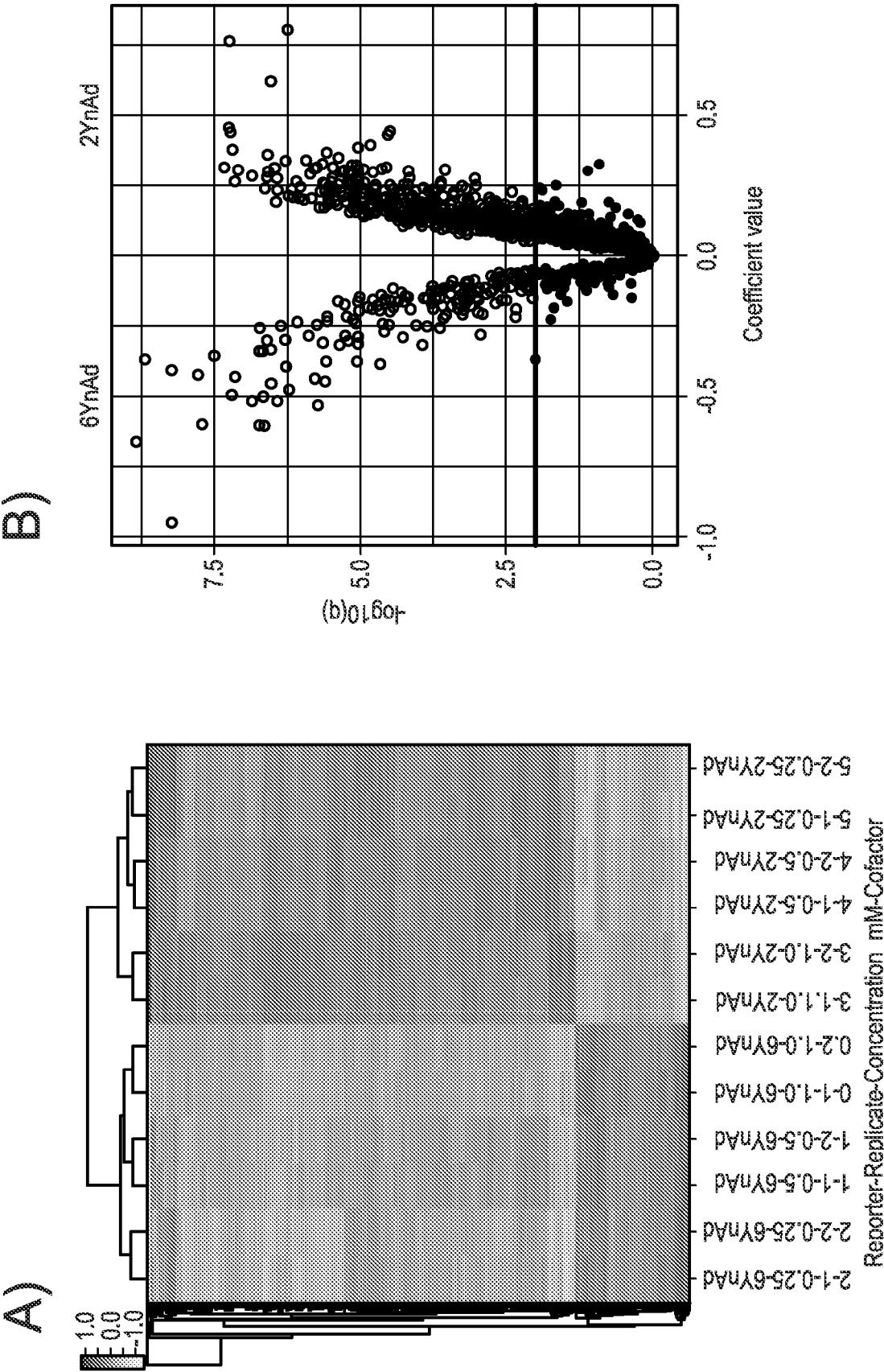


Fig. 9

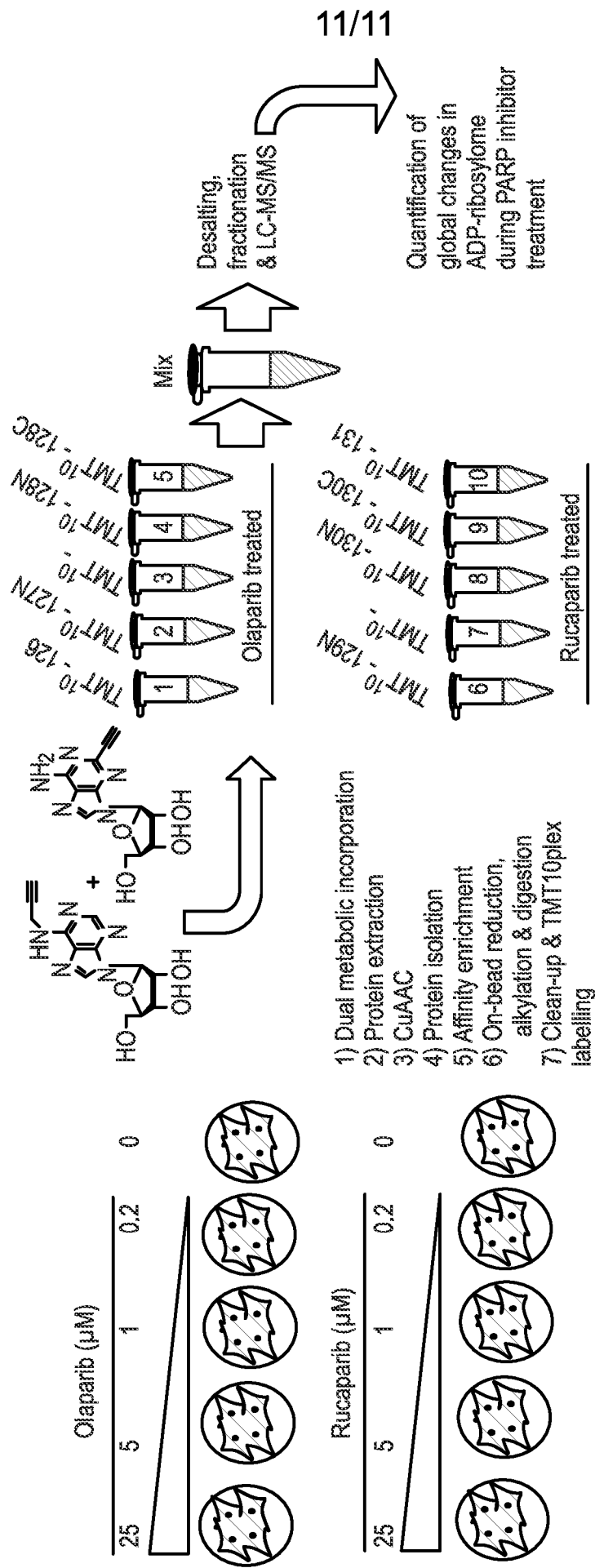


Fig. 10

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/GB2019/051842

A. CLASSIFICATION OF SUBJECT MATTER  
INV. C12Q1/48 G01N33/68  
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
C12Q G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, WPI Data

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                                                                                                    | Relevant to claim No. |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| A         | <p>WESTCOTT N.P. ET AL.: "Chemical proteomics reveals ADP-ribosylation of small GTPases during oxidative stress", NATURE CHEM. BIOL., vol. 13, no. 3, March 2017 (2017-03), pages 302-308, XP055611292, cited in the application abstract<br/>page 302, column 2, last paragraph<br/>figure 1<br/>page 304, column 1, last paragraph - column 2, paragraph 1<br/>page 307, column 1, last paragraph - column 2, last paragraph<br/>-----<br/>-/--</p> | 1-21                  |



Further documents are listed in the continuation of Box C.



See patent family annex.

\* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

7 August 2019

Date of mailing of the international search report

23/08/2019

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040,  
Fax: (+31-70) 340-3016

Authorized officer

Giry, Murielle

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/GB2019/051842

| C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                                                                                                                                                                               |                       |
|------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Category*                                            | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                            | Relevant to claim No. |
| A                                                    | US 2016/299141 A1 (KRAUS W.L. [US] ET AL.)<br>13 October 2016 (2016-10-13)<br>abstract<br>example 1<br>paragraph [0348]<br>claims 24,61                                                                                                                                                                                                       | 1-21                  |
| A                                                    | -----<br>MARTELLO R. ET AL.: "Proteome-wide<br>identification of the endogenous<br>ADP-ribosylome of mammalian cells and<br>tissue",<br>NATURE COMM.,<br>vol. 7, 12917,<br>30 September 2016 (2016-09-30), pages<br>1-13, XP055524651,<br>cited in the application<br>abstract<br>page 2<br>figure 1<br>page 10, column 2 - page 12, column 1 | 1-21                  |
| A                                                    | -----<br>WO 2016/201257 A2 (THE JOHNS HOPKINS<br>UNIVERSITY [US])<br>15 December 2016 (2016-12-15)<br>the whole document                                                                                                                                                                                                                      | 1-21                  |
| A                                                    | -----<br>GAGNÉ J.-P. ET AL.: "Hydrofluoric<br>acid-based derivatization strategy to<br>profile PARP-1 ADP-ribosylation by<br>LC-MS/MS",<br>J. PROTEOME RES.,<br>vol. 17, no. 7, 29 May 2018 (2018-05-29),<br>pages 2542-2551, XP055611022,<br>the whole document                                                                              | 1-21                  |
| X,P                                                  | -----<br>KALESH K. ET AL.: "An integrated chemical<br>proteomics approach for quantitative<br>profiling of intracellular<br>ADP-ribosylation",<br>SCIENTIFIC REP.,<br>vol. 9, no. 1, 6655,<br>30 April 2019 (2019-04-30), pages 1-12,<br>XP055611018,<br>the whole document<br>-----                                                          | 1-21                  |

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2019/051842

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s) | Publication<br>date |
|-------------------------------------------|---------------------|----------------------------|---------------------|
| US 2016299141                             | A1                  | 13-10-2016                 | NONE                |
| -----                                     |                     |                            |                     |
| WO 2016201257                             | A2                  | 15-12-2016                 | NONE                |
| -----                                     |                     |                            |                     |