



(51) International Patent Classification:

C12N 15/11 (2006.01) C07K 1/13 (2006.01)

C12N 15/63 (2006.01) G01N 33/52 (2006.01)

C12N 15/66 (2006.01) G01N 33/68 (2006.01)

C07K 1/00 (2006.01) G01N 33/58 (2006.01)

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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(21) International Application Number:

PCT/IL2020/050275

(22) International Filing Date:

09 March 2020 (09.03.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

267744 30 June 2019 (30.06.2019) IL

62/960,321 13 January 2020 (13.01.2020) US

(71) Applicant: NATIONAL INSTITUTE FOR BIOTECHNOLOGY IN THE NEGEV LTD. [IL/IL]; Bldg. 39, Ben Gurion University of the Negev, 84105 Beer-Sheva (IL).

(72) Inventors: ELIA, Natalie; 330 HaBrosh Street, 8534000 Sde Tzvi (IL). SEGAL, Inbar; 8 HaRav Neriya Street, 7575122 Rishon LeTsiyon (IL). ARBELY, Eyal; 20 Tabib Street, 7532020 Rishon LeTsiyon (IL). NACHMIAS, Dikla; 40 Hadas Street, 8496500 Omer (IL).

(74) Agent: KESTEN, Dov et al.; GEYRA KESTEN, 100 Hashmonaim Street, P.O. Box 52630, 6713317 Tel Aviv (IL).

(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,

(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))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: LABELLING MOLECULES AND METHODS OF USE THEREOF

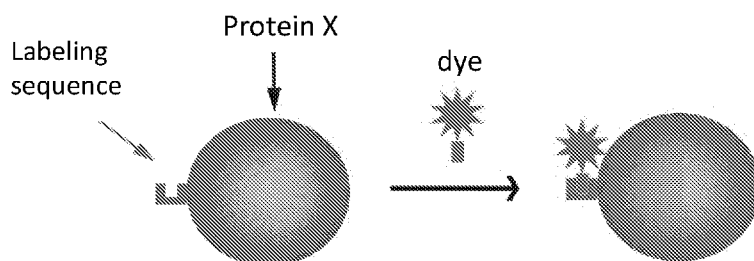


Figure 1A

(57) Abstract: The present invention is directed to a polynucleotide molecule including a first polynucleotide sequence encoding a labelling sequence that is operably linked to a second polynucleotide sequence encoding a protein of interest with a linker of 2-6 amino acids.



LABELLING MOLECULES AND METHODS OF USE THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[001] This application claims the benefit of priority of Israel Patent Application No. 267744 titled "LABELLING MOLECULES AND METHODS OF USE THEREOF", filed June 30, 2019, and of U.S. Provisional Patent Application No. 62/960,321 titled "LABELLING MOLECULES AND METHODS OF USE THEREOF", filed January 13, 2020, the contents of which are incorporated herein by reference in their entirety.

FIELD OF INVENTION

[002] The present invention is in the field of protein labelling.

BACKGROUND

[003] Tracking the dynamics of proteins and cellular components in live cells is key to understanding their functions. For this, fluorescent protein (e.g., GFP) or self-labelling protein (e.g., Halo-Tag) tags are routinely attached to proteins in cells. While these tags are vigorous and easy-to-implement, they are large and bulky (e.g., GFP, ~27 kDa; Halo-tag, 33 kDa), such that their attachment could affect the dynamics and function of the protein under study. Using genetic code expansion (GCE) and bioorthogonal chemistry, it is now possible to non-invasively attach fluorescent dyes (Fl-dyes) to specific protein residues, thereby allowing essentially "tag-free" labelling of proteins in live cells. In GCE-based labelling, a non-canonical amino acid (ncAA) carrying a functional group is incorporated into the protein sequence in response to an in-frame codon reassignment site via an orthogonal tRNA/tRNA-synthetase pair. Labelling is then carried out by a rapid and specific bioorthogonal reaction between the functional group and the Fl-dye.

[004] While the ncAA (and consequently the Fl-dye) can, in theory, be incorporated anywhere in the protein sequence, finding a suitable labelling site can be laborious and time-consuming in practice for several reasons. First, the efficiency of ncAA incorporation varies at different locations in the protein, with no guidelines for the preferred sequence context having been reported. Second, prior knowledge or functional assays are necessary to ensure that insertion of the ncAA does not substantially affect protein structure and function. Third, the functional group in the ncAA should be solvent-exposed to allow efficient bioorthogonal conjugation with the Fl-dye. All these requirements are protein-specific, such that any attempt at labelling via this approach begins with

a screen for suitable incorporation sites. Consequently, despite its great potential, GCE-based labelling is presently not widely used in mammalian live cell imaging applications.

[005] There is a need for a minimal peptide labelling sequence for use in efficient and direct labelling of proteins in live cells, such as without the need of undergoing screening steps currently associated with GCE-based labelling.

SUMMARY

[006] The present invention provides, in some embodiments, polynucleotide molecules encoding labelling sequences, such as for labelling proteins in a cell. Methods and kits for labelling proteins in live cells are also provided.

[007] According to a first aspect, there is provided a polynucleotide molecule comprising: (a) first polynucleotide sequence encoding a labelling sequence, comprising: i. a nucleotide sequence encoding a linker sequence; and ii. a codon reassignment site encoding a ncAA; wherein the codon reassignment site is contiguous to the linker sequence; and (b) a second polynucleotide sequence encoding a protein, wherein the first polynucleotide sequence is operably linked to the second polynucleotide, and wherein the linker comprises 2-6 amino acids selected from the group consisting of: Glycine (G) and Serine (S).

[008] According to another aspect, there is provided an expression vector comprising the polynucleotide molecule of the invention.

[009] According to another aspect, there is provided a cell comprising: (a) the polynucleotide molecule of the invention; (b) the expression vector comprising the polynucleotide of the invention; or any combination of (a) and (b).

[010] According to another aspect, there is provided a method for labelling a protein in a cell, comprising the step of incubating the herein disclosed cell in the presence of: i. a non-canonical amino acid (ncAA) carrying a functional group; ii. an orthogonal tRNA/tRNA-synthetase pair; and iii. a fluorescent (Fl)-dye, wherein the ncAA is incorporated by the orthogonal tRNA/tRNA-synthetase pair to a protein transcribed from the expression vector, and wherein the functional group of the ncAA reacts with the Fl-dye, thereby labelling a protein in the cell.

[011] According to another aspect, there is provided a method for labelling a protein in a cell, wherein the protein is being processed through the secretory pathway, the method comprising incubating the herein disclosed cell in the presence of: i. a non-canonical amino acid (ncAA) carrying a functional group; ii. an orthogonal tRNA/tRNA-synthetase pair; and iii. a fluorescent

(Fl)-dye, wherein the ncAA is incorporated by the orthogonal tRNA/tRNA-synthetase pair to a protein transcribed from the expression vector, and wherein the functional group of the ncAA reacts with said Fl-dye, thereby labelling a protein in the cell.

[012] According to another aspect, there is provided a kit comprising the herein disclosed expression vector of, and at least one of: i. a non-canonical amino acid (ncAA) carrying a functional group; ii. an orthogonal tRNA/tRNA-synthetase pair; iii. a fluorescent dye; and iv. instructions for use of the expression vector with any one of the non-canonical amino acid (ncAA) carrying a functional group; the orthogonal tRNA/tRNA-synthetase pair; and the fluorescent dye.

[013] In some embodiments, the linker sequence comprises a Glycine-Serine dipeptide.

[014] In some embodiments, the linker sequence comprises an amino acid sequence as set forth in SEQ ID NO: 1 (GGSG).

[015] In some embodiments, the first polynucleotide encoding a labelling sequence further comprises a nucleotide sequence encoding a tag.

[016] In some embodiments, the tag is selected from a group consisting of: HA, c-Myc, and FLAG.

[017] In some embodiments, the labelling sequence has a molecular weight of 0.5 – 2 kDa.

[018] In some embodiments, the labelling sequence is operable in a pH value lower than 7.

[019] In some embodiments, the second polynucleotide sequence encodes a protein localized in a specific cellular component.

[020] In some embodiments, the cellular component is selected from the group consisting of: endoplasmic reticulum (ER), plasma membrane (PM), peroxisome, lysosome, multivesicular bodies (MVBs), and exosome.

[021] In some embodiments, the second polynucleotide sequence encodes a microtubule protein.

[022] In some embodiments, the first polynucleotide sequence is located upstream of the second polynucleotide sequence.

[023] In some embodiments, the first polynucleotide is located downstream of the second polynucleotide sequence.

[024] In some embodiments, the first polynucleotide encoding a labelling sequence is located at a predetermined site of the second polynucleotide sequence encoding a protein.

[025] In some embodiments, the predetermined site is contiguous to the start codon of the second polynucleotide sequence encoding a protein.

[026] In some embodiments, the expression vector further comprises a tRNA/tRNA-synthetase pair comprising a polynucleotide encoding a tRNA and a polynucleotide encoding an amino-acyl tRNA synthetase.

[027] In some embodiments, the tRNA/tRNA-synthetase pair are an orthogonal pair to the cell endogenous tRNAs and aminoacyl-tRNA synthetase.

[028] In some embodiments, the protein processed through the secretory pathway is selected from the group consisting of: a secreted protein, an integral protein, and a transmembrane protein.

[029] In some embodiments, the protein undergoes endocytosis.

[030] In some embodiments, the incubating results in (i) incorporation of said ncAA to the polypeptide; (ii) reaction of the functional group of the ncAA with the Fl-dye; or both.

[031] Unless otherwise defined, all technical and/or scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of embodiments of the invention, exemplary methods and/or materials are described below. In case of conflict, the patent specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be necessarily limiting.

[032] Further embodiments and the full scope of applicability of the present invention will become apparent from the detailed description given hereinafter. However, it should be understood that the detailed description and specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE FIGURES

[033] **Figs. 1A-1B** are illustrations of non-limiting schemes depicting the use of a labelling sequence for fluorescence labelling of proteins via bioorthogonal chemistry. **(1A)** A schematic representation of the labelling approach showing a labelling sequence operably linked upstream of the protein (X) sequence. Binding of the Fl-dye to the labelling sequence results in labelling of the protein. **(1B)** A graphical illustration of an experimental design.

[034] **Figs. 2A-2K** are non-limiting illustrations, micrographs, and graphs showing the optimization of a minimal labelling sequence. (**Fig. 2A**) A schematic representation of labelling sequences of Probes 1-4 with α -tubulin protein. α -tubulin^{45TAG} was used as a reference. (**2B**) Western blot analysis of HEK293T cells transfected with expression vectors comprising the polynucleotide sequences depicted in **Fig. 2A**. The cells were incubated for 48 hours in the presence of ncAA BCN-Lysine. (**Figs. 2C-2K**) COS7 cells transfected with expression vectors comprising the polynucleotide sequences depicted in **Fig. 2A**. The cells were incubated for 48 hours in the presence of the ncAA BCN-Lysine, labelled with SiR-tet for 1 hour and imaged live. (**Figs. 2C-2G**) Images representing maximum intensities of 3D z-stacks of representative cells expressing: α -tubulin^{45TAG} (**2C**); probe 1 (**2D**); probe 2 (**2E**); probe 3 (**2F**); and probe 4 (**2G**). Scale-bar: 10 μ m. (**Fig. 2H**) SNR values measured in the COS7 cells expressing the reference α -tubulin^{45TAG}, Probe 3- α -tubulin and Probe 4- α -tubulin. Statistical significance was determined by ANOVA ***p<0.0001, * p<0.05 (n=25). (**Figs. 2I-2J**, upper micrographs) Zoom-in images of the regions marked in squares in **Figs. 2C** and **2G** for the reference α -tubulin^{45TAG} and Probe 4- α -tubulin, respectfully. Intensity values along the dotted line drawn in the bottom graphs, demonstrate the improvement in SNR. Scale-bar: 2 μ m. Results presented in **Figs. 2B-2J** were obtained in at least 3 independent experiments. (**Fig. 2K**) A schematic view of the optimized Probe 4.

[035] **Figs. 3A-3T** are fluorescent micrographs and vertical bar graphs depicting plasma membrane (PM), peroxisomes and lysosomes labelling of COS7 live cells. (**3A-3F**) Fluorescent micrographs of cells expressing Probe 4-GFP-CAAX (**3A-3C**) or GFP-CAAX mutated at position 150 to incorporate the ncAA BCN-Lysine (**3D-3F**) and labelled with SiR-Tet showing plasma membrane labelling. Intensity values were normalized to background levels. (**3A and 3D**) GFP visualization; (**3B and 3E**) SiR visualization; (**3C and 3F**) merging of (**3A-3B**) and (**3D-3E**), respectively. (**3G-3J**) Fluorescent micrographs and a graph showing peroxisome labelling in cells expressing Probe 4-GFP-SKL. (**3J**) Colocalization analysis between intensity values of the 488 (GFP; **3G**) and 640 (SiR; **3H**) channels, Pearson correlation value = 0.814. (**3I**) merging of (**3G-3H**). (**3K-3N**) Probe 4-Lamp1 labelled with SiR-Tet and (**3O-3R**) Lamp1-mVenus, respectively. (**3K, 3L, 3O, and 3P**) do not comprise an added treatment. (**3M, 3N, 3Q, and 3R**) comprise a treatment with the lysosome inhibitor chloroquine (3 h, 120 μ M). (**3L, 3N, 3P, and 3R**) show zoom-in images of a subset of the cells (correspond to the squares in the **3K, 3M, 3O, and 3Q**, respectively). Results presented in each panel were obtained in at least 3 independent experiments. Scale-bars: 10 μ m (**3K, 3M, 3O, and 3Q**); and 2 μ m (**3L, 3N, 3P, and 3R**). (**3S-3T**) are vertical bar graphs showing *in vitro* analysis of the florescence intensity of SiR-

Tet (**3S**) and TAMRA-Tet (**3T**) diluted in HEPES buffer in different pH, in the presence of the ncAA BCN-Lysine or absence of the ncAA BCN-Lysine. For each pH value the left bars represent the presence of the ncAA BCN-Lysine, and the right bars represent the absence of the ncAA BCN-Lysine.

[036] **Figs. 4A-4Q** are fluorescent micrographs and a graph showing MVBs, exosomes and ER exhibit specific cellular labelling. (**4A-4C**) COS7 cells expressing the MVB marker, CD63 conjugated to Probe 4 were fixed and stained with (**4A**) anti-HA and (**4B**) anti-CD63 antibodies. (**4C**) is a merge image of (**4A and 4B**). (**4D**) MVB labelling in cells expressing Probe 4-CD63 labelled with TAMRA-Tet. (**4E-4F**) Exosome labelling in cells expressing Probe 4-Exo70 labelled with TAMRA-Tet. (**4F**) is a zoom-in image of a subset of the cells, corresponding to the square in **4E**. (**4G-4H**) Exosome labelling in cells expressing Exo70-GFP. (**4H**) is a zoom-in image of a subset of the cells, corresponding to the square in **4G**. (**4I-4J**) ER labelling in cells expressing Probe 4- ER^{cb5}TM labelled with TAMRA-Tet. (**4J**) is a zoom-in image of a subset of the cells, corresponding to the square in **4I**. (**4K-4L**) ER labelling in cells expressing ER^{cb5}TM-GFP. (**4L**) is a zoom-in image of a subset of the cells, corresponding to the square in **4K**. (**4M-4P**) FRAP analysis of cells expressing Probe 4-ER^{cb5}TM labelled with TAMTA-Tet. (**4M-4O**) are snapshots of a subset of representative cells. (**4M**) Pre-bleaching; (**4N**) Bleaching (0 sec); and (**4O**) Recovery (48.8 sec). Sequential images of the bleached region of interest are shown in **4P**. (**4Q**) A plot representing the exponential fit of fluorescence intensity recovery versus time after photobleaching in the region of interest. Intensity values were corrected for unintentional bleaching and normalized to intensity levels measured pre-bleaching. Results presented in each panel were obtained in at least 3 independent experiments. Scale-bars: 10 μ m (**4M-4O**), 2 μ m (**4P**).

[037] **Figs. 5A-5R** are a micrograph, fluorescent micrographs, graphs showing labelling sequence variations. (**5A**) A micrograph of a western blot analysis of HEK293T cells transfected with an expression vector comprising a labelling sequence upstream of a GFP-SKL sequence. The labelling sequences are (from left to right): Probe 4 (HA-GGSG-ncAA), FLAG-GGSG-ncAA, Myc-GGSG-ncAA, and a GGSG-ncAA. (**Figs. 5B-5E**) Live cell images of COS7 cells transfected with an expression vector comprising a labelling sequence and a cellular component encoding sequence. (**Figs. 5B-5E**) Cells comprising FLAG epitope-GGSG-ncAA-GFP-SKL. (**5E**) A plot representing colocalization analysis between intensity values of the 488 (GFP; **5B**) and 640 (SiR; **5C**) channels, pearson correlation value= 0.858. (**5D**) merging of (**5B-5C**). (**5F-5G**) Cells comprising FLAG epitope-GGSG-ncAA-protein: (**5F**) an α -tubulin protein; and (**5G**)

an EXO70 protein. **(5H-5J)** Cells comprising Myc epitope-GGSG-ncAA-GFP-SKL. **(5H)** GFP visualization; **(5I)** SiR visualization; and **(5J)** merge of **(5H and 5I)**. **(5K)** Cells comprising Myc epitope-GGSG-ncAA-Exo70. **(5L)** Cells comprising Myc epitope-GGSG-ncAA- α -tubulin. **(5M-5N)** Cells comprising GGSG linker-ncAA-protein: **(5M)** α -tubulin protein **(5N)** an EXO70 protein. **(5O-5R)** Cells comprising epitope-GGSG-ncAA-GFP-SKL. **(5R)** Colocalization analysis between intensity values of the 488 (GFP; **5O**) and 640 (SiR; **5P**) channels, Pearson correlation value = 0.766. **(5Q)** merging of **(5O-5P)**. Results presented in each panel were obtained in at least 3 independent experiments. Scale-bar: 10 μ m.

[038] **Figs. 6A-6B** are an illustration and a fluorescent micrograph describing the labelling of an internal region of a protein sequence. **(6A)** A schematic representation of the ShakerB channel. The labelling sequence was inserted in the extracellular loop connecting S3 and S4. **(6B)** COS7 cells expressing HA epitope-GGSG-ncAA-ShakerB and labelled with tetrazine-Alexa647.

[039] **Figs. 7A-7B** are fluorescent micrographs showing the labelling of the 3' end of a protein. **(7A)** Exosome labelling in COS7 expressing Exo70 linked to a labelling sequence comprising ncAA-GGSG-HA. **(7B)** Exosome labelling in COS7 cells expressing Exo70 linked to a labelling sequence comprising GGSG-ncAA-HA.

[040] **Figs. 8A-8C** are fluorescent micrographs and a gel micrograph showing the labeling of a cellular protein using the GCE-tag. **(8A-8B)** are fluorescent micrographs of confocal slices showing maximum intensity projections taken from live cell images of COS7 cells, transfected with pBUD-Pyl-RS plasmid carrying the constructs: **(8A)** S. pep-GCE tag-EGFR; and **(8B)** S. pep-EGFR, wherein position 128 of EGFR was optimized for site specific labeling (TGA), and labeled with Cy3-Tet. Cy3-Tet is a cell impermeable dye. Intracellular labeling was observed in both **(8A and 8B)**, which results from EGFR endocytosis. Specific labeling was obtained for EGFR. **(8C)** In-gel fluorescence analysis of HEK293T cells transfected as in **(8A-8B)** and labeled with SiR-Tet. S. pep, signal peptide; GCE-tag, genetic code expansion tag; EGFR, epidermal growth factor receptor. Scale-bar: 10 μ m.

DETAILED DESCRIPTION

[041] The present invention, in some embodiments, provides a polynucleotide molecule comprising a labelling sequence for labelling a protein of interest in a cell, an expression vector comprising the polynucleotide molecule, and a cell expressing the vector. The present invention further concerns, in some embodiments, a method for labelling proteins in a live cell. A kit for labelling proteins in a cell is also provided.

[042] The invention is based on the surprising findings that cellular component dynamics in live cells can be tracked with a labelling sequence comprising as few as 5 residues.

The findings disclosed herein are based on the comprehensive analysis of protein labelling sequences of varying lengths, located at different positions of a protein and expressed in various cellular components.

Polynucleotide molecules and methods of use

[043] In some embodiments, the invention is directed to a polynucleotide molecule comprising: a first polynucleotide sequence encoding a labelling sequence, comprising: (a) a nucleotide sequence encoding a linker sequence comprising 2-6 amino acids selected from a group consisting of: Glycine (G) and Serine (S); and, a codon reassignment site encoding a ncAA is contiguous to the linker sequence; and; (b) a second polynucleotide sequence encoding a protein sequence (e.g., a protein of interest), wherein the first polynucleotide sequence is operably linked to the second polynucleotide.

[044] According to some embodiments, a method is provided for labelling a protein (e.g., a protein of interest) in a cell. According to some embodiments, the method comprises incubating the cell comprising the expression vector of the invention in the presence of: (a) a non-canonical amino acid (ncAA) carrying a functional group; (b) an orthogonal tRNA/tRNA-synthetase pair; and (c) a fluorescent dye, wherein the ncAA is incorporated by the orthogonal tRNA/tRNA-synthetase pair to a polypeptide transcribed from the expression vector, and the functional group of the ncAA reacts with the FI-dye, thereby labelling the protein in the cell.

[045] The term "nucleic acid" or "nucleotide" are used interchangeably and are well known in the art. A "nucleic acid" as used herein will generally refer to DNA, RNA or a derivative or an analog thereof, comprising a nucleobase. A nucleobase includes, for example, a naturally occurring purine or pyrimidine base found in DNA (e.g., an adenine "A," a guanine "G," a thymine "T" or a cytosine "C") or RNA (e.g., an A, a G, an uracil "U" or a C).

[046] The terms "nucleic acid molecule" and "nucleic acid sequence" include but not limited to single-stranded RNA (ssRNA), double-stranded RNA (dsRNA), single-stranded DNA (ssDNA), double-stranded DNA (dsDNA). According to some embodiments, a "nucleic acid sequence" refers to a sequence comprising at least 2 nucleotides.

[047] The terms "polynucleotide," "polynucleotide sequence," or "polynucleotide molecule" are used herein interchangeably. The aforementioned terms encompass nucleotide sequences and the like. A polynucleotide may be a polymer of RNA, DNA, or a hybrid thereof. In some

embodiments, the polynucleotide is single- or double-stranded. In some embodiments, the polynucleotide comprises single and double stranded regions. In some embodiments, the polynucleotide comprises an double stranded region and an over-hang region. In some embodiments, the polynucleotide comprises synthetic, non-natural, or altered nucleotide bases, or any combination thereof.

[048] According to some embodiments, the first polynucleotide molecule encodes a labelling sequence. According to some embodiments, the labelling sequence comprises a linker sequence and a codon reassignment site encoding a linker and a ncAA, respectively.

[049] As used herein, the term “linker” sequence encompasses a flexible linker or a rigid linker. According to some embodiments, the linker comprises at least 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acids, or any value and range therebetween. Each possibility represents a separate embodiment of the invention. According to some embodiments, the linker comprises at most 4, 5, 6, 7, 8, 9, 10, 12 amino acids, or any value and range therebetween. Each possibility represents a separate embodiment of the invention. According to some embodiments, the linker comprises 2-10, 2-8, 2-6, 2-5, 2-4 amino acids. Each possibility represents a separate embodiment of the invention. According to some embodiments, the linker sequence comprises a Glycine (G) amino acid and/or a Serine (S) amino acid.

[050] According to some embodiments, the linker comprises a Glycine-Serine dipeptide.

[051] In some embodiments, the linker consists of Glycine and Serine amino acid residues.

[052] According to some embodiments, the linker sequence comprises an amino acid sequence as set forth in SEQ ID NO: 1 (GGSG). According to some embodiments, the linker sequence consists of an amino acid sequence as set forth in SEQ ID NO: 1 (GGSG).

[053] As used herein, the phrase “Codon reassignment site” refers to an unassigned codon sequence or reassigned codon sequence which allows the site-specific installation of a ncAA into a protein of choice. In some embodiments, the codon reassignment site comprises a stop codon sequence. In some embodiments, the codon reassignment site is selected from: TAG sequence, TAA sequence, TGA sequence, or a 4-based codon sequence.

[054] According to some embodiments, the codon reassignment site is contiguous to the nucleotide sequence encoding a linker sequence. According to some embodiments, the codon reassignment site is located upstream of the nucleotide sequence encoding a linker sequence (e.g., at the 5' end). According to some embodiments, the codon reassignment site is located downstream of the nucleotide sequence encoding a linker (e.g., at the 3' end).

[055] As used herein the term “non-canonical amino acid” (ncAA) encompasses unnatural amino acids. According to some embodiments, the ncAA is selected from: 3-Iodo-L-tyrosine, N^ε-Benzyloxycarbonyllysine (ZLys), N^ε-Acetyllysine (AcLys), N^ε-Cyclopentylloxycarbonyl-L-lysine (Cyc), N^ε-(((1R,2R)-2-azidocyclopentyl)oxycarbonyl)-L-lysine (ACPK), o-Nitrobenzyl-Otyrosine, o-Nitrobenzyloxycarbonyl-N^ε-Llysine, N^ε-[(1-(6-Nitrobenzo[d][1,3]dioxol-5yl)ethoxy)carbonyl]-L-lysine, N^ε-[(2-(3-Methyl-3Hdiazirine-3-yl)ethoxy)carbonyl]-Llysine, (3-(3-Methyl-3Hdiazirine-3-yl)-propaminocarbonyl)N^ε-L-lysine (DiZPK), BCN (exo isomer), BCN (endo isomer), TCO, N^ε-(1-Methylcycloprop-2-enecarboxamido)lysine (CpK), N^ε-Acryloyl-L-lysine, pNO₂ZLys, TmdZLys, N^ε-Crotonyl-L-lysine (Kcr), 2-Chloro-L-phenylalanine, 2-Bromo-L-phenylalanine, 2-Iodo-L-phenylalanine, 2-Methyl-L-phenylalanine, 2-Methoxy-L-phenylalanine, 2-Nitro-L-phenylalanine, 2-Cyano-L-phenylalanine, and N^ε-(tert-Butoxycarbonyl)-L-lysine (BocLys).

[056] According to some embodiments, the ncAA is a BCN ncAA.

[057] According to some embodiments, the labelling sequence is operably linked to the protein (e.g., a protein of interest). According to some embodiments, the labelling sequence is operably linked upstream of the protein. According to some embodiments, the labelling sequence is operably linked to the protein between the 5' end and the 3' end of the protein (or the N' terminus and the C' terminus, respectively). According to some embodiments, the labelling sequence is contiguous and downstream of the start codon of a polynucleotide sequence encoding the protein. According to some embodiments, the labelling sequence is operably linked downstream of the protein coding sequence. In some embodiments, the labelling sequence is incorporated, embedded, located, or positioned within the protein coding sequence.

[058] As used herein, the operably linked is interchangeable to physically linked, contiguous, or any equivalent thereof.

[059] According to some embodiments, the labelling sequence is operably linked to the protein at a predetermined site. According to some embodiments, the labelling sequence is located at a predetermined site of the second polynucleotide sequence encoding a protein. According to some embodiments, the predetermined site may be α -helix, β -strand, or a structural loop.

[060] According to some embodiments, the predetermined site faces the lumen of a cellular component. According to some embodiments, the predetermined site faces the extracellular side

of the cell. According to some embodiments, the predetermined site faces the intracellular side of the cell.

[061] Determining the direction to which a labelling agent faces, is well within the capabilities of one of ordinary skill in the art. Non-limiting examples for methods for determining such direction include, but are not limited to, fluorescent assays, such as immunoassays targeting cell component specific biomarkers, e.g., cell membrane (e.g., actin fibre specific stain using phalloidin), or other cell organelles: ER (e.g., SERCA2, Calreticulin, COL1A1, Calnexin, HO-1, Heme oxygenase 2, RyR1, and UGGT1, among others), Golgi apparatus (e.g., B4GALT1, COG2, GALNT2, gamma Adaptin, GCC1, GLG1, GM130, Golgi Protein 58k, Golgin-97, GRASP65, MAN1A2, MAN2A1, RCAS1, and TGN46, and others).

[062] In some embodiments, the cell is selected from: an eukaryote cell, a prokaryote cell, and an archaean cell.

[063] According to some embodiments the cell is a eukaryote cell. According to some embodiments the cell is a prokaryote cell. In some embodiments, the cell is selected from: a mammalian cell, a bacterial cell, a yeast cell, a plant cell, and a synthetic cell.

[064] Synthetic cells are well known to a skilled person. A synthetic cell may be a minimal cell, artificial cell, cell-like system, semi-synthetic cell or cell-mimetic. According to some embodiments, the synthetic cell comprises sufficient information to allow the cell to carry out essential biological processes, such as, for example, transcription, translation, use of an energy source, transport of salts, nutrients and the like into and out of the organelle or cell, etc.

[065] As used herein, the terms "peptide", "polypeptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues. In another embodiment, the terms "peptide", "polypeptide" and "protein" as used herein encompass native peptides, peptidomimetics (typically including non-peptide bonds or other synthetic modifications), and peptide analogues peptoids and semipeptoids or any combination thereof. In another embodiment, a peptide, as described herein, has modifications rendering it more stable while in the body or more capable of penetrating into a cell. In one embodiment, the terms "peptide", "polypeptide" and "protein" apply to naturally occurring amino acid polymers. In another embodiment, the terms "peptide", "polypeptide" and "protein" apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid. According to some embodiments, the terms "peptide", "polypeptide" and "protein" apply to

amino acid polymers comprising one or more labelling moieties, compounds, agents, sequences, or any combination thereof.

[066] According to some embodiments, the protein is a cellular component protein. As used herein, the term "cellular component" refers to a cell organelle, or a cell compartment. In some embodiment, a cellular component protein refers to that the protein is present, functions, located, or can be identified particularly or predominantly in a cellular compartment. In some embodiment, a cellular component protein refers to that the protein is more likely to be present, function, located, or identified in a particular or specific cellular compartment over other cellular components, compartments, or areas of the cell. According to some embodiments, the cellular component is selected from: the plasma membrane (PM), peroxisome, lysosome, multivesicular bodies (MVBs), endoplasmic reticulum (ER), exosome, mitochondria, nucleus, Golgi apparatus, or centrosome.

[067] In some embodiments, the protein is processed through the secretory pathway. In some embodiments, the term "protein processed through the secretory pathway" refers to any protein comprising a signal peptide or any equivalent thereof which in turn results in the protein being targeted to the cellular secretory pathway. In some embodiments, the secretory pathway comprises: the cell membrane, the ER, the Golgi apparatus, a secretion vesicle, an exosome, an endosome, or any combination thereof. In some embodiments, the secretory pathway comprises other cellular components as disclosed hereinabove.

[068] In some embodiments, the method of the invention is optimized for labelling of a protein being processed through the secretory pathway.

[069] In some embodiments, for labelling a protein being processed through the secretory pathway, the labelling sequence is integrated and/or located contiguously to the C'-terminal end of the signal peptide of the protein. In some embodiments, for labelling a protein being processed through the secretory pathway, the labelling sequence is integrated into the protein in a position located C'-terminally of the signal peptide of the protein.

[070] According to some embodiments, the labelling sequence does not substantially affect the structure and/or function of the protein. According to some embodiments, the labelling sequence does not substantially affect the structure and/or function of the protein as compared to the same protein labeled with GFP. According to some embodiments, the labelling sequence does not substantially affect the structure and/or function of the protein as compared to a wild type protein, e.g., a non-labeled protein.

[071] According to some embodiments, the polynucleotide encoding the labelling sequence further comprises a nucleotide sequence encoding a tag sequence. According to some embodiments, the tag is at least: 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, or 20 amino acids, or any value and range therebetween. Each possibility represents a separate embodiment of the invention. According to some embodiments, the tag is at most: 10, 12, 14, 16, 18, 20, 25, 30, 50, or 75 amino acids, or any value and range therebetween. Each possibility represents a separate embodiment of the invention. According to some embodiments, the tag is 6-20, 6-18, 6-16, 6-15, 6-13, 6-12, 6-12, 7-11, 7-12, 7-13, 7-14, 8-16, 8-14, 8-13, 8-12, 8-10 amino acids. Each possibility represents a separate embodiment of the invention.

[072] According to some embodiments, the tag sequence is an epitope. According to some embodiments, the epitope is selected from a list including, but not limited to, 6-His tag (HHHHHH; SEQ ID NO: 2), Myc tag (EQKLISEEDL; SEQ ID NO: 3), Hemagglutinin or HA tag (YPYDVPDYA; SEQ ID NO: 4), S tag (KETAAAKFERQHMDL; SEQ ID NO: 5), FLAG or DYKDDDDK tag (SEQ ID NO: 6), E Tag (GAPVPPDPLEPR; SEQ ID NO: 7), AVI Tag (GLNDIFEAQKIEWHE; SEQ ID NO: 8), HSV tag (QPELAPEDPED; SEQ ID NO: 9), AU1 tag (DTYRYI; SEQ ID NO: 10), CBP tag (KRRWKKNFIAVSAANRFKKISSSGAL; SEQ ID NO: 11), VSV-g tag (YTDIEMNRLGK; SEQ ID NO: 12), Glu-Glu tag (EMYMPE; SEQ ID NO: 13), AU5 tag (TDFYLK; SEQ ID NO: 14), KT3 tag (KPPTPPPEPET; SEQ ID NO: 15). According to some embodiments, the tag sequence is a HA tag, FLAG tag, or a Myc tag.

[073] According to some embodiments, the tag sequence is located at the 5' end of the labelling sequence. According to some embodiments, the tag sequence is located at the 3' end of the labelling sequence.

[074] It will be clear to one of ordinary skill in the art that the polypeptide molecular weight is measured in Kilodalton (kDa) units. According to some embodiments, the labelling sequence molecular weight is at least: 0.3, 0.4, 0.44, 0.46, 0.48, 0.5, 0.52, 0.54, 0.56, 0.58, 0.6, 0.7, 0.8, 0.9, 1, or 2 kDa, or any value and range therebetween. Each possibility represents a separate embodiment of the invention. According to some embodiment, the labelling sequence molecular weight is at most: 0.6, 0.8, 1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 2.2, 2.5, 2.8, 3, 4, 5, or 10 kDa, or any value and range therebetween. Each possibility represents a separate embodiment of the invention. According to some embodiment, the labelling sequence is between 0.4-10, 0.4-8, 0.4-6, 0.4-4, 0.4-2.5, 0.4-2, 0.4-1.8, 0.5-2.5, 0.5-2, 0.5-1.8, 0.5-1.7, 0.5-1.65 kDa. Each possibility represents a separate embodiment of the invention.

[075] According to some embodiments the labelling sequence is operable in a pH value lower than 7. According to some embodiments, an operable labelling sequence refers to a labelling sequence that is able to bind a FI-Dye. According to some embodiments, the labelling sequence is operable in a pH value greater than 7. According to some embodiments, the labelling sequence is operable in a pH value of 7. According to some embodiments, the labelling sequence is operable in a pH value ranging from 4-7, 4-6, 4-5.5, or 4-5. Each possibility represents a separate embodiment of the invention.

[076] According to some embodiments, the protein is selected: α -tubulin, CAAX, SKL, Lysosomal-associated membrane protein 1 (Lamp1), EP^{cb5}TM (e.g., a mitochondrial marker), EXO70 (e.g., an exosome marker), CD63 (e.g., multivesicular bodies (MVBs) marker), and ShakerB (e.g., a potassium ion channel).

[077] As used herein, "CAAX" refers to any polypeptide comprising the CAAX box or motif, wherein Cysteine (C), A is any aliphatic amino acid, and the identity of X determines which enzyme acts on the protein. For example, when X = M, S, Q, A, or C, the enzyme farnesyltransferase recognizes the motif and adds a hydrophobic molecule to the C residue, e.g., a process termed prenylation, isoprenylation, or lipidation. As an alternative example, when X = L or E, the enzyme geranylgeranyltransferase I recognizes the motif and adds a hydrophobic molecule to the C residue.

[078] As used herein, "SKL", refers to any polypeptide comprising the peroxisomal targeting signal sequence of Serine (S)-Lysine (K)-Leucine (L).

[079] The term "expression" as used herein refers to the biosynthesis of a gene product, including the transcription and/or translation of said gene product. Thus, expression of a nucleic acid molecule may refer to transcription of the nucleic acid fragment (e.g., transcription resulting in mRNA or other functional RNA) and/or translation of RNA into a precursor or mature protein (polypeptide).

[080] Expressing of a gene within a cell is well known to one skilled in the art. It can be carried out by, among many methods, transfection, viral infection, or direct alteration of the cell's genome. In some embodiments, the gene is in an expression vector such as plasmid or viral vector.

[081] A vector nucleic acid sequence generally contains at least one origin of replication for propagation in a cell and optionally additional elements, such as a heterologous polynucleotide sequence, an expression control element (e.g., a promoter, enhancer), a selectable marker (e.g.,

antibiotic resistance), a poly-Adenine sequence, a labelling sequence, or any combination thereof.

[082] It will be appreciated that other than containing the necessary elements for the transcription and translation of the inserted coding sequence (encoding the polypeptide), the expression construct of the present invention can also include sequences engineered to optimize stability, production, purification, yield or activity of the expressed polypeptide.

[083] According to some embodiments, the vector is a DNA plasmid. In some embodiments, the vector is delivered via non-viral methods or via viral methods. According to some embodiments, the viral vector is a retroviral vector, a herpes viral vector, an adenoviral vector, an adeno-associated viral vector, or a poxviral vector.

[084] In some embodiments, the promoter is a constitutive promoter, for example RNA polymerase III (Pol III) promoter. In some embodiments, the promoter is an inducible promoter. According to some embodiments, the promoter is a viral promoter. In some embodiments, the promoter is selected from: elongation factor 1 α -subunit promoter (EF1 α), U6, cytomegalovirus (CMV), H1, or other promoters shown effective for expression in eukaryotic cells. In some embodiments, the gene is operably linked to the promoter. In one embodiment, the expression of a protein coding sequence is driven by a number or a plurality of promoters.

[085] The term “operably linked” is intended to mean that the nucleotide sequence of interest is linked to the element or elements in a manner that allows for expression of the nucleotide sequence (e.g., in an *in-vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

[086] In some embodiments, nucleic acid sequences are transcribed by RNA polymerase II (RNAP II and Pol II). RNAP II is an enzyme found in eukaryotic cells. It catalyzes the transcription of DNA to synthesize precursors of mRNA and most snRNA and microRNA.

[087] In some embodiments, mammalian expression vectors include, but are not limited to, pcDNA3, pcDNA3.1 (\pm), pGL3, pZeoSV2(\pm), pSecTag2, pDisplay, pEF/myc/cyto, pCMV/myc/cyto, pCR3.1, pSinRep5, DH26S, DHBB, pNMT1, pNMT41, pNMT81, which are available from Invitrogen, pCI which is available from Promega, pMbac, pPbac, pBK-RSV and pBK-CMV which are available from Strategene, pTRES which is available from Clontech, and their derivatives, or pBUD available from Addgene.

[088] Various methods can be used to introduce the expression vector of the present invention into cells. Such methods are generally described in Sambrook et al., *Molecular Cloning: A*

Laboratory Manual, Cold Springs Harbor Laboratory, New York (1989, 1992), in Ausubel et al., Current Protocols in Molecular Biology, John Wiley and Sons, Baltimore, Md. (1989), Chang et al., Somatic Gene Therapy, CRC Press, Ann Arbor, Mich. (1995), Vega et al., Gene Targeting, CRC Press, Ann Arbor Mich. (1995), Vectors: A Survey of Molecular Cloning Vectors and Their Uses, Butterworths, Boston Mass. (1988) and Gilboa et al. [Biotechniques 4 (6): 504-512, 1986] and include, for example, stable or transient transfection, lipofection, electroporation and infection with recombinant viral vectors. In addition, see U.S. Pat. Nos. 5,464,764 and 5,487,992 for positive-negative selection methods.

[089] In some embodiments, the vector further comprises a polynucleotide encoding an orthogonal tRNA. In some embodiments, the vector further comprises a polynucleotide encoding a tRNA-synthetase.

[090] The term “orthogonal pair” refers to a tRNA/tRNA-synthetase pair which does not occur naturally in the cell. As used herein the terms “tRNA synthase”, “amino-acyl tRNA synthetase” and “aaRS” refer to an enzyme capable of acylating a tRNA with an amino acid or amino acid analog. In some embodiments, the aaRS does not catalyze aminoacylation of an endogenous tRNA. In some embodiments, the aaRS aminoacylate the orthogonal tRNA with a ncAA.

[091] According to some embodiments the orthogonal tRNA/tRNA-synthetase pair is derived from Eukaryote, Bacteria, or Archaea. According to some embodiments the orthogonal tRNA/tRNA-synthetase pair is derived from an organism selected from: *E. coli*, *Methanocaldococcus jannaschii*, *Methanosarcina barkeri*, *Desulfitobacterium hafniense*, *Methanobacterium thermoautotrophicum*LeuRS/*Halobacterium* sp., *Methanosarcina mazei*, *Saccharomyces cerevisiae*, and *Bacillus stearothermophilus*. According to some embodiments, the orthogonal tRNA/tRNA-synthetase pair is a chimeric pair.

[092] According to some embodiments, the dye used is a fluorescent organic dye or tetrazine-dye. According to some embodiments, the tetrazine-dye is selected from: ATTO425, ATTO465, ATTO488, ATTO532, Cy3, carboxytetramethylrhodamine (TAMRA), ATTO550, ATTO565, ATTO590, ATTO594, ATTO620, Si-rhodamine (SiR), ATTO647N, Cy5, ATTO655, ATTO680, ATTO700, AZ503, AZ519, JF6466, HMSiR, and Cy5B.

[093] According to some embodiments, the cell is a mammalian cell. According to some embodiments, the cell is a living mammalian cell. In some embodiments the labeled cell is being labeled while being alive. In some embodiments, the labeled cell is not being labeled while being fixed or in the presence of a fixation solution or a fixative.

Kits

[094] According to some embodiments, a kit is provided comprising an expression vector comprising the polynucleotide molecule of the invention and at least one of a non-canonical amino acid (ncAA) carrying a functional group, an orthogonal tRNA/tRNA-synthetase pair, and a fluorescent dye.

[095] According to some embodiments, the kit comprises instructions for using a polynucleotide molecule with the ncAA, the orthogonal tRNA/tRNA-synthetase pair, and/or the fluorescent dye. Non-limiting examples of instructions include the ratio of each element, the sequence of contacting the cell with the orthogonal tRNA/tRNA-synthetase pair, and/or the fluorescent dye, and the like. In some embodiments, the instructions comprise the ratio between each of the elements, the order of steps according to which the cell is being contacted with the orthogonal tRNA/tRNA-synthetase pair, and/or the fluorescent dye, and the like.

[096] In some embodiments of the herein disclosed kits, the polynucleotide molecule and at least one of the ncAA, the orthogonal tRNA/tRNA-synthetase pair, and the fluorescent dye are packaged within a container. In some embodiments, the container is made of a material selected from: thin-walled film or plastic (transparent or opaque), paperboard-based, foil, rigid plastic, metal (e.g., aluminum), glass, etc.

[097] In some embodiments, the content of the kit is packaged, as described below, to allow for storage of the components until they are needed.

[098] In some embodiments, some or all components of the kit may be packaged in suitable packaging to maintain sterility.

[099] In some embodiments, the herein disclosed kit, polynucleotide molecule and at least one of the ncAA, the orthogonal tRNA/tRNA-synthetase pair, and the fluorescent dye are stored in separate containers within the main kit containment element e.g., box or analogous structure, may or may not be an airtight container, e.g., to further preserve the sterility of some or all of the components of the kit.

[0100] In some embodiments, the dosage amount of the polynucleotide molecule and at least one of the ncAA, the orthogonal tRNA/tRNA-synthetase pair, and the fluorescent dye provided in a kit as disclosed herein, may be sufficient for a single application or for multiple applications.

[0101] In those embodiments, the kit may have multiple dosage amounts of the polynucleotide molecule and at least one of the ncAA, the orthogonal tRNA/tRNA-synthetase pair, and the

fluorescent dye packaged in a single container, e.g., a single tube, bottle, vial, Eppendorf and the like.

[0102] In some embodiments, the kit may have multiple dosage amounts of the polynucleotide molecule and at least one of the ncAA, the orthogonal tRNA/tRNA-synthetase pair, and the fluorescent dye individually packaged such that certain kits may have more than one container of each.

[0103] In some embodiments, multiple dosage amounts of the polynucleotide molecule and at least one of the ncAA, the orthogonal tRNA/tRNA-synthetase pair, and the fluorescent dye may be packed in single separated containers.

[0104] In some embodiments, the kit contains instructions for preparing the composition used therein and for how to practice the methods of the invention.

[0105] In some embodiments, the kit further comprises a measuring utensil such as syringe, measuring spoon or a measuring cup.

[0106] In some embodiments, the instructions may be present in the kit as a package insert, in the labelling of the container of the kit or components thereof (i.e., associated with the packaging or sub-packaging) etc. In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, e.g. CD-ROM, diskette, etc. In other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g. via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions is recorded on a suitable substrate.

[0107] According to some embodiments, "predetermined" is defined to a location on the protein sequence that does not substantially affect the structure and/or function of the protein.

[0108] As used herein, the term "Genetic code expansion" (GCE) refers to reprogramming a codon to encode for a ncAA.

[0109] As used herein, the term "about" when combined with a value refers to plus and minus 10% of the reference value. For example, a length of about 1000 nanometers (nm) refers to a length of 1000 nm \pm 100 nm.

[0110] It is noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example,

reference to "a polynucleotide" includes a plurality of such polynucleotides and reference to "the polypeptide" includes reference to one or more polypeptides and equivalents thereof known to those skilled in the art, and so forth. It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely", "only" and the like in connection with the recitation of claim elements, or use of a "negative" limitation.

[0111] In those instances where a convention analogous to "at least one of A, B, and C, etc." is used, in general such a construction is intended in the sense one having skill in the art would understand the convention (e.g., "a system having at least one of A, B, and C" would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). It will be further understood by those within the art that virtually any disjunctive word and/or phrase presenting two or more alternative terms, whether in the description, claims, or drawings, should be understood to contemplate the possibilities of including one of the terms, either of the terms, or both terms. For example, the phrase "A or B" will be understood to include the possibilities of "A" or "B" or "A and B."

[0112] It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable sub-combination. All combinations of the embodiments pertaining to the invention are specifically embraced by the present invention and are disclosed herein just as if each and every combination was individually and explicitly disclosed. In addition, all sub-combinations of the various embodiments and elements thereof are also specifically embraced by the present invention and are disclosed herein just as if each and every such sub-combination was individually and explicitly disclosed herein.

[0113] Additional objects, advantages, and novel features of the present invention will become apparent to one ordinarily skilled in the art upon examination of the following examples, which are not intended to be limiting. Additionally, each of the various embodiments and aspects of the present invention as delineated hereinabove and as claimed in the claims section below finds experimental support in the following examples.

[0114] Various embodiments and aspects of the present invention as delineated hereinabove and as claimed in the claims section below find experimental support in the following examples.

EXAMPLES

[0115] Generally, the nomenclature used herein, and the laboratory procedures utilized in the present invention include molecular, biochemical, microbiological and recombinant DNA techniques. Such techniques are thoroughly explained in the literature. See, for example, "Molecular Cloning: A laboratory Manual" Sambrook et al., (1989); "Current Protocols in Molecular Biology" Volumes I-III Ausubel, R. M., ed. (1994); Ausubel et al., "Current Protocols in Molecular Biology", John Wiley and Sons, Baltimore, Maryland (1989); Perbal, "A Practical Guide to Molecular Cloning", John Wiley & Sons, New York (1988); Watson et al., "Recombinant DNA", Scientific American Books, New York; Birren et al. (eds) "Genome Analysis: A Laboratory Manual Series", Vols. 1-4, Cold Spring Harbor Laboratory Press, New York (1998); methodologies as set forth in U.S. Pat. Nos. 4,666,828; 4,683,202; 4,801,531; 5,192,659 and 5,272,057; "Cell Biology: A Laboratory Handbook", Volumes I-III Cellis, J. E., ed. (1994); "Culture of Animal Cells - A Manual of Basic Technique" by Freshney, Wiley-Liss, N. Y. (1994), Third Edition; "Current Protocols in Immunology" Volumes I-III Coligan J. E., ed. (1994); Stites et al. (eds), "Basic and Clinical Immunology" (8th Edition), Appleton & Lange, Norwalk, CT (1994); Mishell and Shiigi (eds), "Strategies for Protein Purification and Characterization - A Laboratory Course Manual" CSHL Press (1996); all of which are incorporated by reference. Other general references are provided throughout this document.

Materials and Methods

Cell culture

[0116] COS7 and HEK293T cells were grown in Dulbecco's Modified Eagle Medium (DMEM; Life Technologies, Carlsbad, CA) supplemented with 10% fetal bovine serum (FBS), 2 mM glutamine, 10,000 U/ml penicillin and 10 mg/ml streptomycin.

Plasmids and constructs

[0117] The polynucleotide sequences were operably linked to an elongation factor 1 α -subunit promoter (EF1 α) and sub-cloned into a single expression vector pBUD-BCNK-RS using *NotI/KpnI* restriction sites. The vector further comprised *pylT* encoding for tRNA_{CUA}^{Py1}, and Pyrrolysyl-tRNA synthetase. Sequences encoding cellular component markers were then inserted in frame using the *KpnI/XhoI* restriction sites at pBUD-BCNK-RS. All constructs were sequenced before use.

Incorporation of the ncAA to proteins in cells

[0118] Twenty-four hours prior to transfection, cells were plated at 20% confluency using the following dishes: live cell imaging, 4-well chamber slide (Ibidi, Martinsried, Germany); Western blot, 12-well plate (NUNC, Rochester, NY); immunostaining, #1.0 coverslips (Menzel, Braunschweig, Germany). Cells were transfected with pBUD-BCNK-RS plasmids carrying different tags and cellular component markers using Lipofectamine 2000 (Life Technologies, Carlsbad, CA) according to the manufacturer's protocol, and incubated for 48 hours in the presence of the ncAA BCN-Lysine (0.5 mM; Synaffix, Oss, Netherlands) in growth media supplemented with 100 μ M ascorbic acid (Sigma Aldrich, Israel).

Bioorthogonal labelling

[0119] 48 hours post transfection, cells were washed with fresh medium (3 \times quick wash followed by 3 \times 30 min wash) at 37°C, incubated with SiR-Tet (1–2 μ M, 1 h; Spirochrome, Stein am Rhein, Switzerland) or TAMRA-Tet (2 μ M, 1 h; Jena BioScience, Germany) and washed again with fresh medium (3 \times quick wash and 3 \times 30 min wash) at 37°C.

Western blot

[0120] Cells were harvested 48 hours post transfection using RIPA lysis buffer (150 mM NaCl, 1% NP-40, 0.5% deoxycholate, 0.1% SDS, 50 mM Tris [pH 8.0]) supplemented with complete protease inhibitor for 30 minutes at 4°C. Total protein concentrations were measured with BCA Protein Assay Kit (Pierce Biotechnology) and equal total protein amounts were loaded. Membranes were stained with the following primary antibodies: rabbit anti-HA (1:4000; Applied Biological Materials, Richmond, Canada), mouse anti-GAPDH (1:1000; Applied Biological Materials), mouse anti-GFP (1:1000, Applied Biological Materials) and with rabbit or mouse-peroxidase secondary antibodies (1:10,000; Jackson ImmunoResearch, West Grove, PA).

Immunofluorescence

[0121] Cells were fixed 48 hours post transfection with 4% paraformaldehyde (PFA) and co-stained with rabbit anti-HA (1:500) and Mouse anti-CD63 (1:200; Abcam, Cambridge, MA) primary antibodies and with Alexa Fluor 488 anti-rabbit and Alexa Fluor 594 anti-mouse (1:500, Life Technologies) secondary antibodies. Cells were mounted with Fluoromount-G (SouthernBiotech, Birmingham, AL).

Live cell imaging and image processing

[0122] Cells were imaged on a fully incubated confocal spinning-disk microscope at 37°C (Marianas; Intelligent Imaging, Denver, CO) using a 63 \times oil objective (numerical aperture 1.4),

and recorded on an electron-multiplying charge-coupled device camera (pixel size, 0.079 μm ; Evolve; Photometrics, Tucson, AZ). FRAP experiments were performed on cells expressing Probe 4-ER^{cb5}TM and labelled with TAMRA-Tet. After 4 baseline time-points, bleaching was carried out using a 405 laser. Recovery after bleaching was recorded for 1 minute with 1.8 second intervals. Image processing, FRAP analysis fitting, and unintentional bleaching corrections were performed using SlideBook, version 6 (Intelligent Imaging, Denver, CO).

In-vitro analysis of pH sensitivity of tetrazine-conjugated FI-dyes

[0123] SiR-Tet (1 μM) or TAMRA-Tet (2 μM) dyes were diluted in HEPES buffers of different acidity (pH ranges from 5 to 9) in the presence or absence of the ncAA BCN-Lysine (0.5 mM). As a control, FI-dyes were also diluted in 0.1 M HCl and 0.1 M NaOH. Samples were then loaded in triplicates onto a 384-well plate (Greiner), and incubated for 30 minutes at 37°C. Fluorescence intensity was recorded using infinite M1000 plate reader (Tecan, Männedorf, Switzerland) using 652/674 nm or 545/575 nm excitation / emission wavelengths, respectively.

EXAMPLE 1

Labelling sequence structure

[0124] Potential labelling sequences were cloned into an expression vector. The labelling were located upstream to a protein coding sequence and comprised a sequence encoding a linker operably linked to a codon reassignment site comprising a TAG stop codon (UAG mRNA) and/or a sequence encoding a tag polypeptide. During ribosomal translation the ncAA BCN-Lysine is incorporated at the codon reassignment site in response to an in-frame UAG codon using a specific, orthogonal pair of tRNA/ tRNA-Synth. Finally, a tetrazine-conjugated FI-dye is covalently attached to BCN-Lysine via a bioorthogonal reaction. As a result, the protein is directly labelled with FI-dye via a small polypeptide tag (**Figs. 1A-1B**).

[0125] Labelling potency was initially evaluated by conjugating potential labelling sequences to α -tubulin and evaluating microtubule (MT) labelling obtained in live mammalian cells in the presence of tetrazine-silicon-rhodamine (SiR-Tet). MT labelling obtained upon site-specific incorporation of BCN-Lysine at α -tubulin position 45 (α -tubulin^{45TAG}) was used as a reference.

[0126] The labelling sequences were inserted contiguous and downstream of the start codon of a polynucleotide sequence encoding the protein. Introducing a labelling sequence comprising of only a TAG stop codon sequence to the α -tubulin resulted in no MT labelling in the presence of SiR-Tet, indicating that the 5' end labelling sequences need to be longer than a single residue.

[0127] Additional sequences carrying labelling sequences at the 5' end of the sequence encoding a protein were examined. The labelling sequences comprised different lengths and were designed based on the commonly used 9 amino-acid (AA)-long hemagglutinin (HA) epitope as backbone. The TAG stop codon was either introduced into the HA sequence (replacing the last codon in the epitope) (Probe 1), immediately after the HA sequence (Probe 2), after a short commonly used flexible linker comprising glycine-serine (GS) (Probe 3) or after the short commonly used flexible linker glycine-glycine-serine-glycine (GGSG) (Probe 4) (**Fig. 2A**). Low but noticeable levels of α -tubulin were observed using any of these sequences, indicating that α -tubulin comprising a labelling sequence at the 5' end was expressed in cells (**Fig. 2B**). In live cell labelling experiments with SiR-Tet, little to no MT decoration was obtained using α -tubulin tagged with Probes 1 or 2, while clear and specific labelling was observed using Probes 3 and 4 (**Figs. 2C and 2E**). Signal-to-noise ratios (SNRs) were significantly higher in cells expressing α -tubulin labelled with Probe 4 (as compared to Probe 3) and even higher than those obtained in cells expressing α -tubulin^{45TAG} (**Fig. 2D**). Therefore, it can be concluded that a labelling sequence comprising glycine and serine residues (specifically the GGSG linker) is the most suitable for α -tubulin labelling in live mammalian cells (**Fig. 2F**).

EXAMPLE 2

Diverse cellular components labelling

[0128] To test the applicability of Probe 4 for labelling diverse cellular components, COS7 cells were transfected with expression vectors comprising Probe 4 fused to the 5' end of the sequence of a cellular component protein marker. Specifically, GFP-CAAX was used as a plasma membrane (PM) marker, GFP-SKL as a peroxisomal marker, Lamp1 as a lysosomal marker, CD63 as a marker of multivesicular bodies (MVBs), ER^{cb5}TM as a endoplasmic reticulum (ER) marker, Exo70 as an exosomal marker and Mito^{cb5}TM(mito) or mito-DsRed as mitochondrial markers.

[0129] Clear PM labelling was obtained in both the GFP and SiR channels upon expressing GFP-CAAX. Moreover, SNRs measured for PM labelling were similar using either GFP-CAAX (**Fig. 3Ai**) or the optimized ncAA incorporation site in GFP (**Fig. 3Aii**). COS7 cells expressing the peroxisome marker GFP-SKL fused to Probe 4 and labelled with SiR-Tet, exhibited co-localization of GFP and SiR in small puncta throughout the cell (**Fig. 3B**). Apart from the non-specific nuclear labelling obtained in the SiR channel, essentially all SiR-Tet-labelled puncta

were positive for GFP (Pearson correlation = 0.814). Together, these data indicate that both the PM and peroxisome were labelled and properly targeted in cells.

[0130] Effective lysosome labelling was obtained using Probe 4 labelling sequence with Lamp1 in the presence of SiR-Tet (**Figs. 3C-3D**). Overall, labelling was comparable to that obtained with Lamp1-mVenus, with the expected increase in lysosome numbers upon chloroquine treatment being observed using either tag (i.e., labelling sequence or mVenus). Of note, while lysosome labelling using Lamp1-mVenus highlighted both filled and hollow structures, only filled structures were observed in cells labelled with Probe 4-Lamp1. This difference indicates the position of the tag; the mVenus tag faced the cytosol, while the labelling sequence faced the lysosomal lumen. Labelling lysosomal lumen proteins using conventional fluorescent proteins is challenging due to the pH sensitivity of the latter. SiR fluorescence, on the other hand, is much less sensitive to pH as can be seen in **Fig. 3E**. **Figure 3E**, describes an *in vitro* analysis of SiR-Tet and TAMRA-Tet diluted in HEPES buffer in different pH, in the presence or absence of the ncAA BCN-Lysine. Therefore, SiR fluorescence is suitable for intra-lysosome labelling.

[0131] Initial attempts at labelling MVBs, ER, mitochondria and exosomes using probe 4 and SiR-Tet resulted in no specific staining. Yet, in immunostaining experiments using anti-CD63 antibodies, probe 4-CD63 co-localized with endogenous CD63 (**Fig. 4A**). The inventors thus reasoned that probe 4-CD63 was properly expressed and targeted in cells yet failed to bind the FI-dye via the bioorthogonal reaction. Consistent with this notion, substituting SiR-Tet FI-dye with TAMRA-Tet resulted in specific labelling of MVBs, ER and exosomes using probe 4 with CD63, ER^{cb5}TM and Exo70, respectively (**Figs. 4B-4C, and 4E**). Mitochondrial labelling was not obtained using either SiR- Tet or TAMRA-Tet together with Probe 4-Mito^{cb5}TM or mito-DsRed probes. The successful labelling of probe 4 with the cellular component markers of MVBs, ER and exosomes were similar to those obtained using conventional FI-protein markers (**Figs. 4D-4F**).

EXAMPLE 3

Tracking labelled cellular structures in live cells

[0132] FI-dye labelling of the Probe 4 constructs allowed live cell recordings and tracking of any of the labelled cellular structures (i.e., MT, PM, peroxisomes, ER, MVB and exosomes). Fluorescence recovery after photobleaching (FRAP) experiments were successfully performed using Probe 4 -ER^{cb5}TM ER marker, with recovery time being comparable to those measured in the ER using VSVG-GFP (**Figs. 4G-4H**). It is worth mentioning that the overall size of the Probe

4-ER^{cb5}TM ER marker is considerably smaller than VSVG-GFP (Probe 4-tag-ER^{cb5}TM, ~8.5 kDa; VSVG-GFP, ~84.5 kDa). Taken together, these results indicate that the newly developed labelling sequence for cellular component reported here can be employed to track cellular component dynamics in live cells with much smaller probes.

EXAMPLE 4

Labelling sequence versatility

[0133] The 14 residue-long Probe 4 includes the complete HA tag sequence and can therefore be exploited for other applications, such as immunoblot, immunofluorescence and immunoprecipitation (**Figs. 2B, 4A, and 5A**). To expand the of the labelling sequence, the inventors tested whether the HA epitope can be replaced by a Myc epitope (10 residues) or a FLAG (8 residues) epitope (**Fig. 5A-5F**). GFP-SKL carrying HA, FLAG, or Myc epitope with the GGSG linker, and a TAG stop codon were successfully expressed in cells, with SiR-Tet fluorescence co-localizing with GFP (Pearson correlation values: FLAG, 0.858, Myc, 0.73) (**Figs. 5A-5B, and 5D**). Efficient labelling with the FLAG and Myc tags was also obtained for Exo70 (**Figs. 5C, and 5E**). MT labelling, however, appeared to be compromised using FLAG or Myc tags, with very few cells exhibiting MT staining using the Myc epitope (**Figs. 5C, and 5F**). These results demonstrate the versatility of the system in terms of the epitope used in the labelling sequence and indicate that the labelling sequence can be employed both for fluorescent labelling and for other classical applications involving these epitopes.

EXAMPLE 5

Labelling sequence length optimization

[0134] Next examined was whether a labelling sequence comprising the linker and the ncAA is sufficient for site-specific bioorthogonal labelling of different cellular component. The inventors tested cells comprising a labelling sequence comprising GGSG linker and ncAA operably linked to α -tubulin protein (**Fig. 5Gi**) or a labelling sequence comprising GGSG linker and ncAA operably linked to an EXO70 protein (**Figure 5Gii**). The plot represents colocalization analysis between intensity values of the 488 (GFP) and 640 (SiR) channels, pearson correlation value=0.766. Results presented in each panel were obtained in at least 3 independent experiments. Scale-bar: 10 μ m.

[0135] While MT labelling was almost completely abolished, specific labelling was obtained for peroxisomes and exosomes, albeit at reduced levels (**Figs. 5G-5H**) (Pearson correlation =

0.766). These results indicate that the labelling sequence can be reduced to as few as five residues.

EXAMPLE 6

Labelling of an internal region of a protein sequence

[0136] Some proteins cannot be tagged at their 3' end (corresponding to the polynucleotide 5' end). The inventors therefore tested whether Probe 4 can be incorporated in internal regions of a protein sequence. For that, the voltage dependent potassium channel, ShakerB, was used. Both the 5' end and 3' end of ShakerB face the inner side of the plasma membrane and were shown to be important for the activity of the channel. The channel crosses the membrane six times with the transmembrane domains connected by unfolded loops (**Fig. 6A**). Inserting the probe in the extracellular loop connecting transmembrane domains 3 and 4 (at position 345) resulted in efficient labelling in the presence of tetrazine-Alexa647 (**Fig. 6B**). Labelling efficiency was higher than by specifically mutating residue 345 to BCN-Lysine. These results indicate that the labelling sequence can be incorporated in different locations relative to the protein sequence.

EXAMPLE 7

Labelling the 3' end of a protein sequence

[0137] The inventors examined the labelling sequence efficiency when incorporated to the 3' end of a polynucleotide. COS7 cells were introduced with sequence comprising a labelling sequence operably linked at the 3' end of a sequence encoding an Exo70 protein.

[0138] The labelling sequence comprised a codon reassignment site, a GGSG linker, and a HA tag. The labelling sequence was arranged in either a 5' to 3' order of ncAA-GGSG-HA (**Fig. 7A**) or a 5' to 3' order of GGSG- ncAA-HA (**Fig. 7B**).

[0139] Images taken from transfected COS7 cells labelled with Tet-TAMRA exhibited successful labelling of exosomes in live cells with either labelling sequence (**Figs. 7A-7B**).

EXAMPLE 8

Labelling of a cellular protein using the GCE-tag

[0140] The inventors examined whether the successful labelling is exclusively attributed to the GCE-tag being located in the N-terminal end. The GCE-tag was inserted between the signal peptide and the protein coding sequence. Successful and specific labelling was obtained for GCE-

tagged extracellular protein EGFR, as indicated by in-gel fluorescence and live-cell imaging (**Fig. 8**). Thus, besides its use in organelle labelling, the GCE-tag can be used for labelling extracellular proteins (and possibly intracellular proteins) in live cells.

[0141] Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims.

CLAIMS

What is claimed is:

1. A polynucleotide molecule comprising:
 - a. a first polynucleotide sequence encoding a labelling sequence, comprising:
 - i. a nucleotide sequence encoding a linker sequence; and,
 - ii. a codon reassignment site encoding a ncAA;wherein the codon reassignment site is contiguous to the linker sequence; and
 - b. a second polynucleotide sequence encoding a protein,wherein the first polynucleotide sequence is operably linked to the second polynucleotide, and wherein said linker comprises 2-6 amino acids selected from the group consisting of: Glycine (G) and Serine (S).
2. The polynucleotide molecule of claim 1, wherein said linker sequence comprises a Glycine-Serine dipeptide.
3. The polynucleotide molecule of claim 1, wherein said linker sequence comprises an amino acid sequence as set forth in SEQ ID NO: 1 (GGSG).
4. The polynucleotide molecule of any one of claims 1 to 3, wherein said first polynucleotide encoding a labelling sequence further comprises a nucleotide sequence encoding a tag.
5. The polynucleotide molecule of claim 4, wherein said tag is selected from a group consisting of: HA, c-Myc, and FLAG.
6. The polynucleotide molecule of any one of claims 1 to 5, wherein said labelling sequence has a molecular weight of 0.5 – 2 kDa.
7. The polynucleotide molecule of any one of claims 1 to 6, wherein said labelling sequence is operable in a pH value lower than 7.
8. The polynucleotide molecule of any one of claims 1 to 7, wherein said second polynucleotide sequence encodes a protein localized in a specific cellular component.

9. The polynucleotide molecule of claim 8, wherein the cellular component is selected from the group consisting of: endoplasmic reticulum (ER), plasma membrane (PM), peroxisome, lysosome, multivesicular bodies (MVBs), and exosome.

10. The polynucleotide molecule of any one of claims 1 to 9, wherein said second polynucleotide sequence encodes a microtubule protein.

11. The polynucleotide molecule of any one of claims 1 to 10, wherein said first polynucleotide sequence is located upstream of said second polynucleotide sequence.

12. The polynucleotide molecule of any one of claims 1 to 10, wherein said first polynucleotide is located downstream of said second polynucleotide sequence.

13. The polynucleotide molecule of any one of claims 1 to 10, wherein said first polynucleotide encoding a labelling sequence is located at a predetermined site of said second polynucleotide sequence encoding a protein.

14. The polynucleotide molecule of claim 13, wherein said predetermined site is contiguous to the start codon of the second polynucleotide sequence encoding a protein.

15. An expression vector comprising the polynucleotide molecule of any one of claims 1 to 14.

16. The expression vector of claim 15, further comprising a tRNA/tRNA-synthetase pair comprising a polynucleotide encoding a tRNA and a polynucleotide encoding an amino-acyl tRNA synthetase.

17. A cell comprising:

- a. the polynucleotide molecule of any one of claims 1 to 14;
- b. the expression vector of claim 15 or 16; or
- c. any combination of (a) and (b).

18. The cell of claim 17, wherein said tRNA/tRNA-synthetase pair are an orthogonal pair to the cell endogenous tRNAs and aminoacyl-tRNA synthetase.

19. A method for labelling a protein in a cell, comprising the step of incubating the cell of claim 17 or 18 in the presence of:

- i. a non-canonical amino acid (ncAA) carrying a functional group;
- ii. an orthogonal tRNA/tRNA-synthetase pair; and
- iii. a fluorescent (Fl)-dye,

wherein said ncAA is incorporated by said orthogonal tRNA/tRNA-synthetase pair to a protein transcribed from said expression vector, and wherein said functional group of said ncAA reacts with said Fl-dye,

thereby labelling a protein in the cell.

20. A method for labelling a protein in a cell, wherein said protein is being processed through the secretory pathway, the method comprising incubating the cell of claim 17 or 18 in the presence of:

- i. a non-canonical amino acid (ncAA) carrying a functional group;
- ii. an orthogonal tRNA/tRNA-synthetase pair; and
- iii. a fluorescent (Fl)-dye,

wherein said ncAA is incorporated by said orthogonal tRNA/tRNA-synthetase pair to a protein transcribed from said expression vector, and wherein said functional group of said ncAA reacts with said Fl-dye,

thereby labelling a protein in the cell.

21. The method of claim 20, wherein said protein processed through the secretory pathway is selected from the group consisting of: a secreted protein, an integral protein, and a transmembrane protein.

22. The method of claim 20 or 21, wherein said protein undergoes endocytosis.

23. The method of any one of claims 19 to 22, wherein said incubating results in (i) incorporation of said ncAA to said polypeptide; (ii) reaction of the functional group of said ncAA with said Fl-dye; or both.

24. A kit comprising the expression vector of claim 15, and at least one of:

- i. a non-canonical amino acid (ncAA) carrying a functional group;

- ii. an orthogonal tRNA/tRNA-synthetase pair;
- iii. a fluorescent dye; and
- iv. instructions for use of the expression vector with any one of said non-canonical amino acid (ncAA) carrying a functional group; said orthogonal tRNA/tRNA-synthetase pair; and said fluorescent dye.

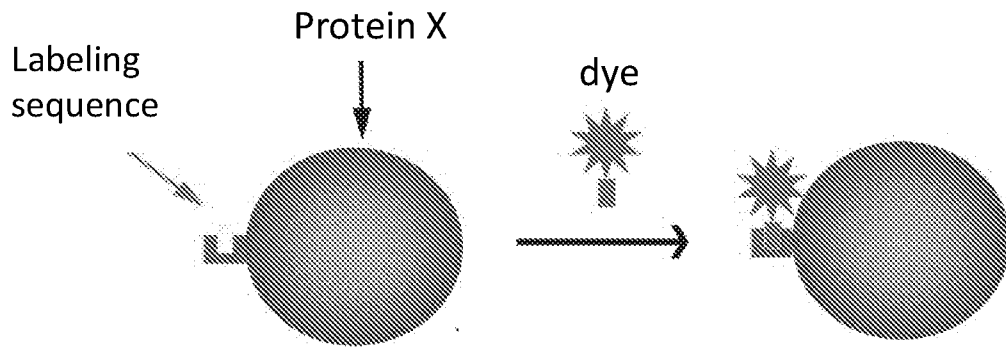


Figure 1A

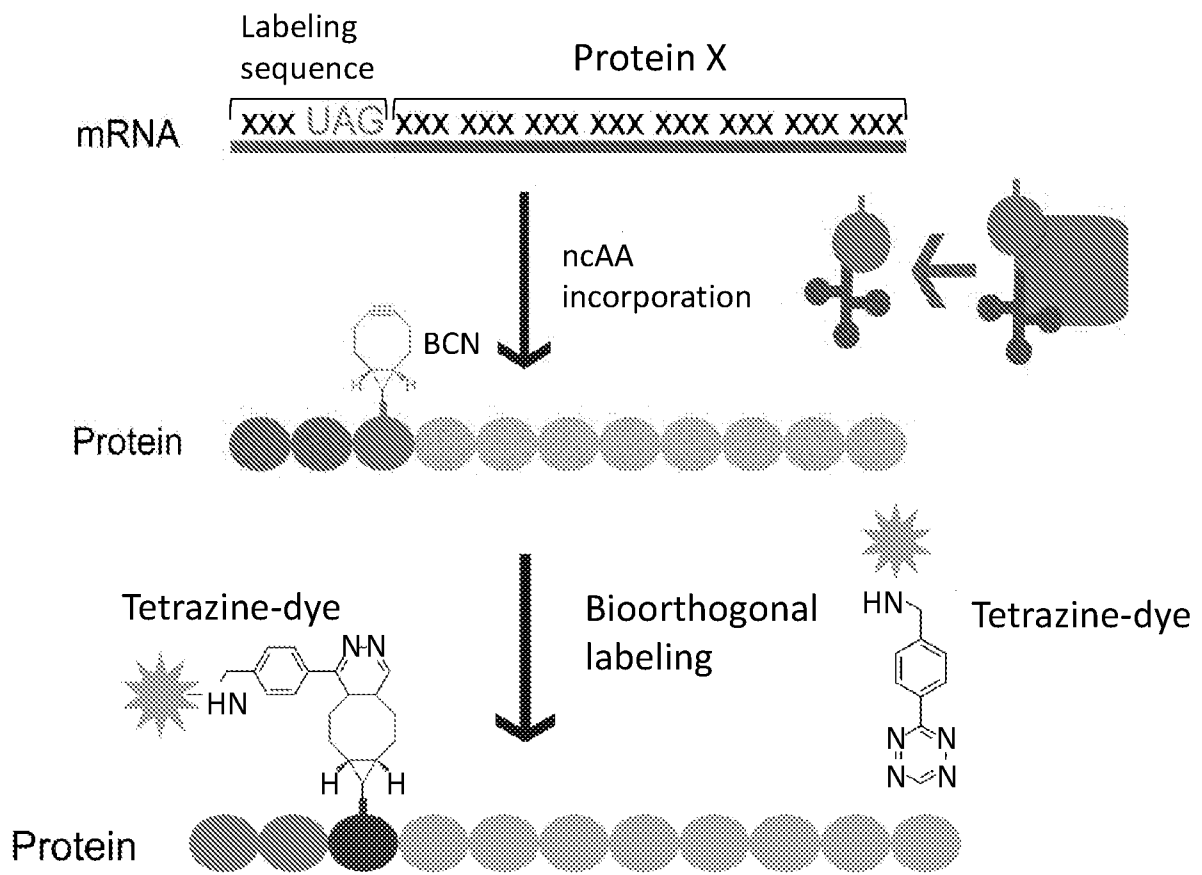


Figure 1B

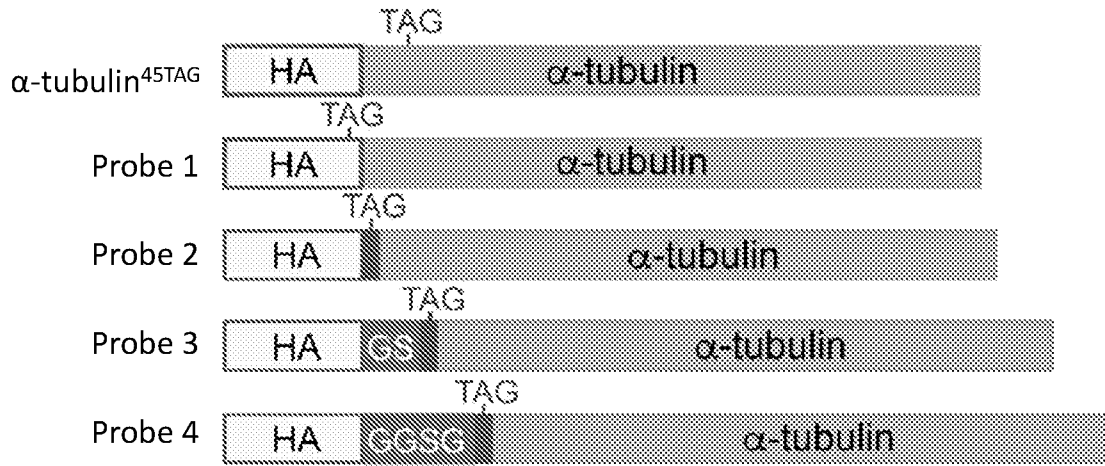


Figure 2A

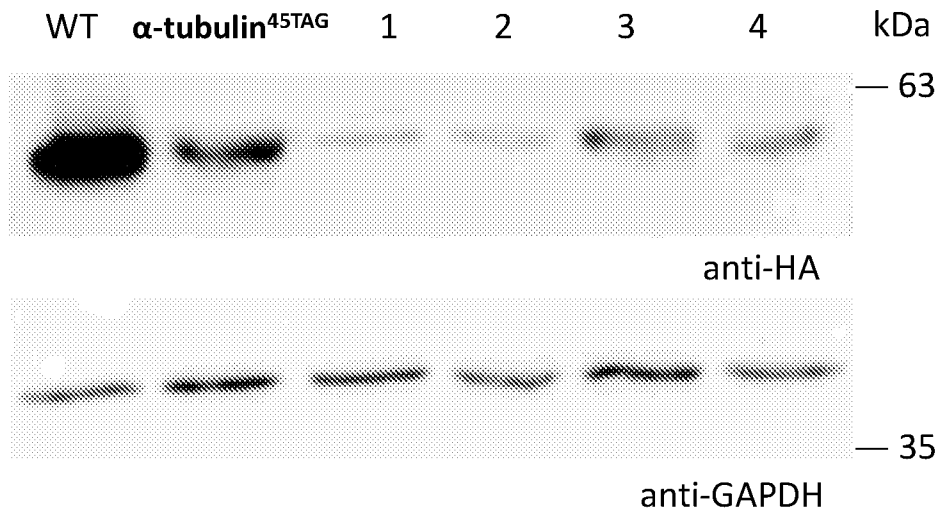


Figure 2B

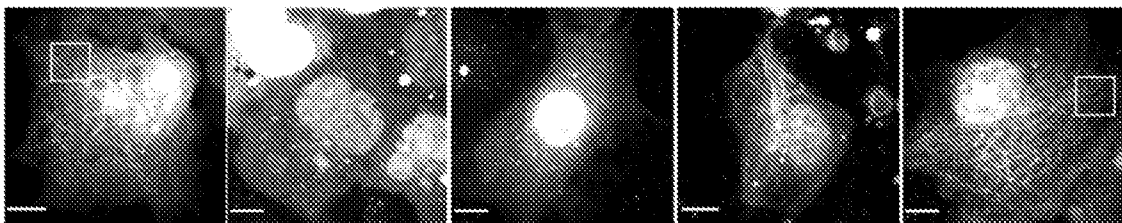


Figure 2C

Figure 2D

Figure 2E

Figure 2F

Figure 2G

3/10

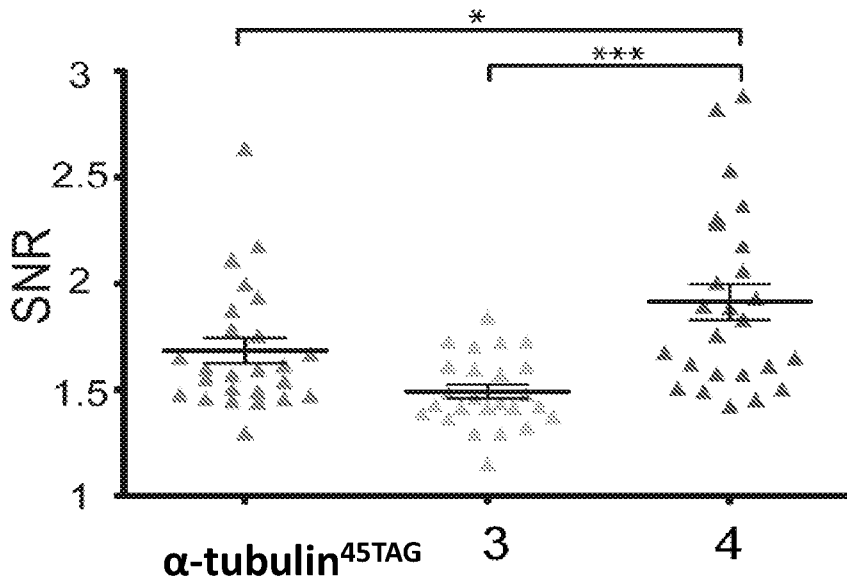


Figure 2H

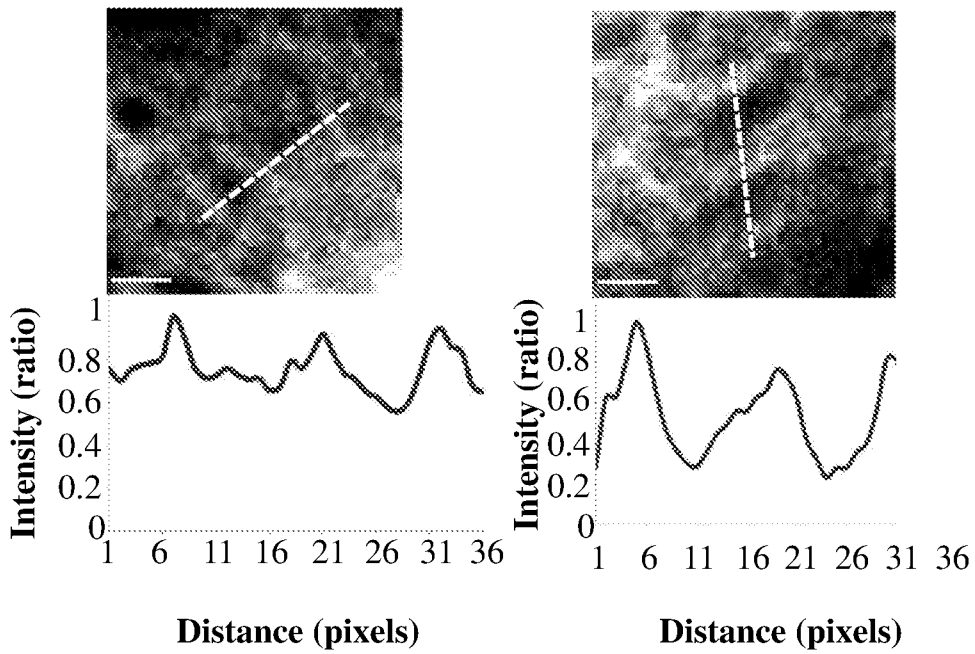


Figure 2I

Figure 2J

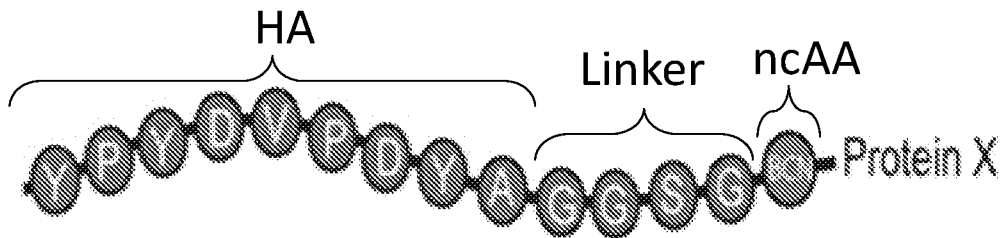


Figure 2K

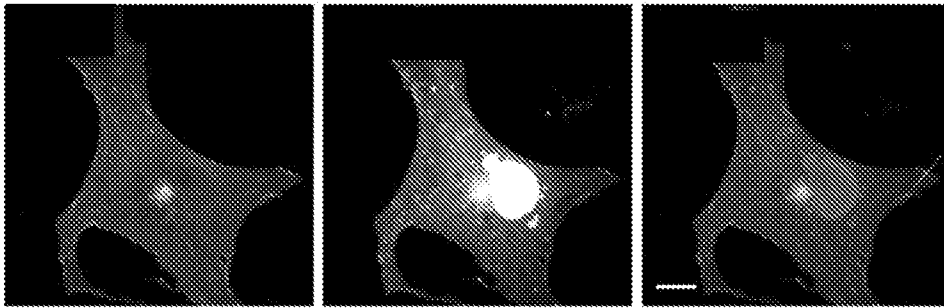


Figure 3A

Figure 3B

Figure 3C

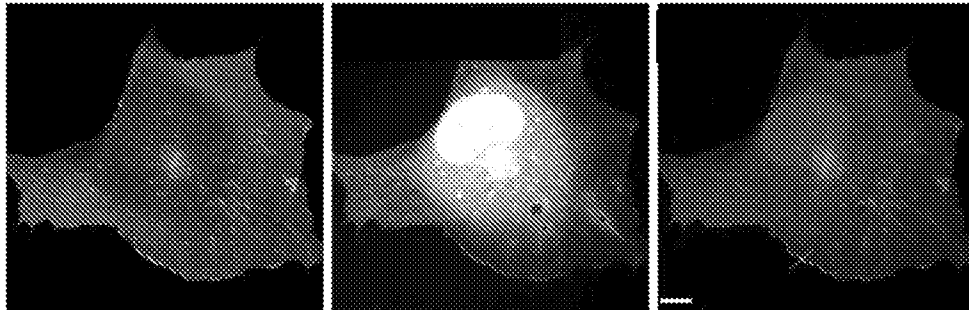


Figure 3D

Figure 3E

Figure 3F

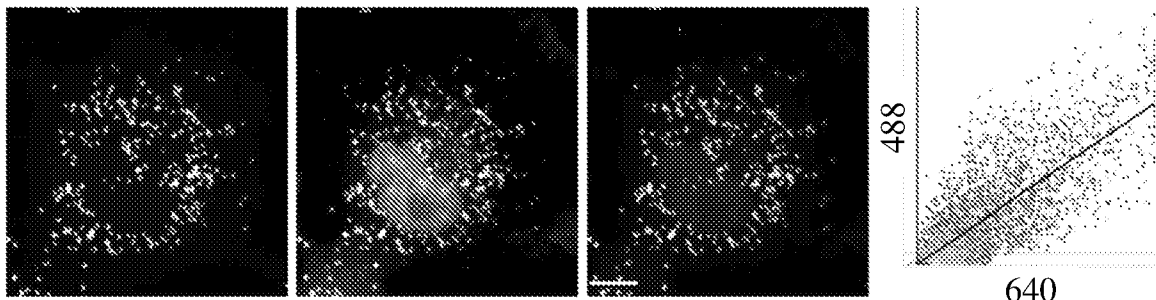


Figure 3G

Figure 3H

Figure 3I

Figure 3J

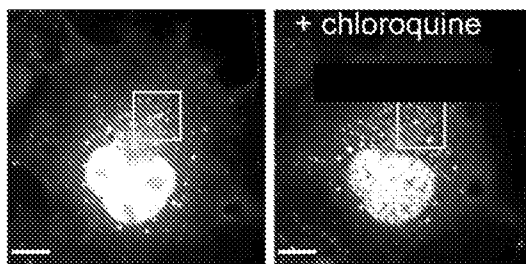


Figure 3K

Figure 3M

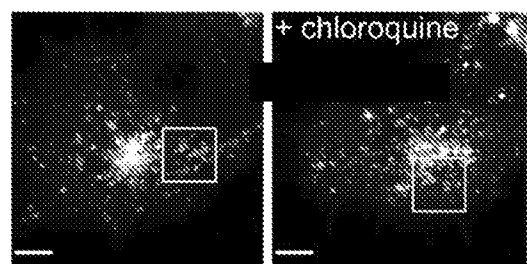


Figure 3O

Figure 3Q

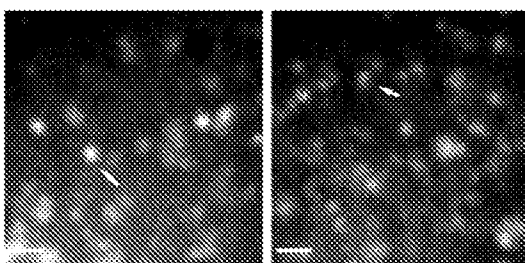


Figure 3L

Figure 3N

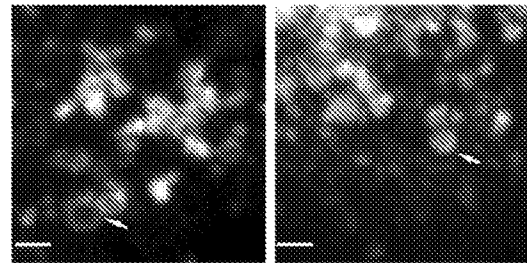


Figure 3P

Figure 3R

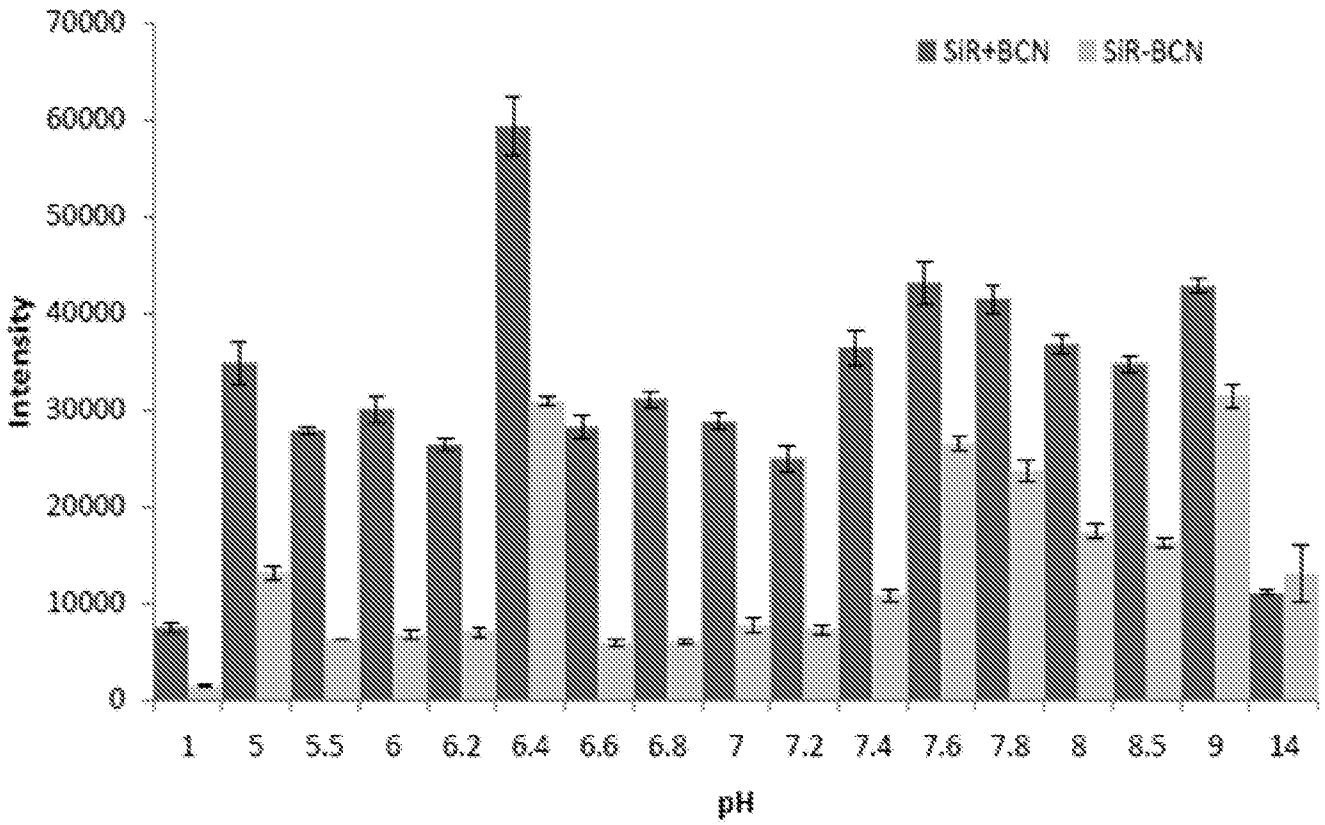


Figure 3S

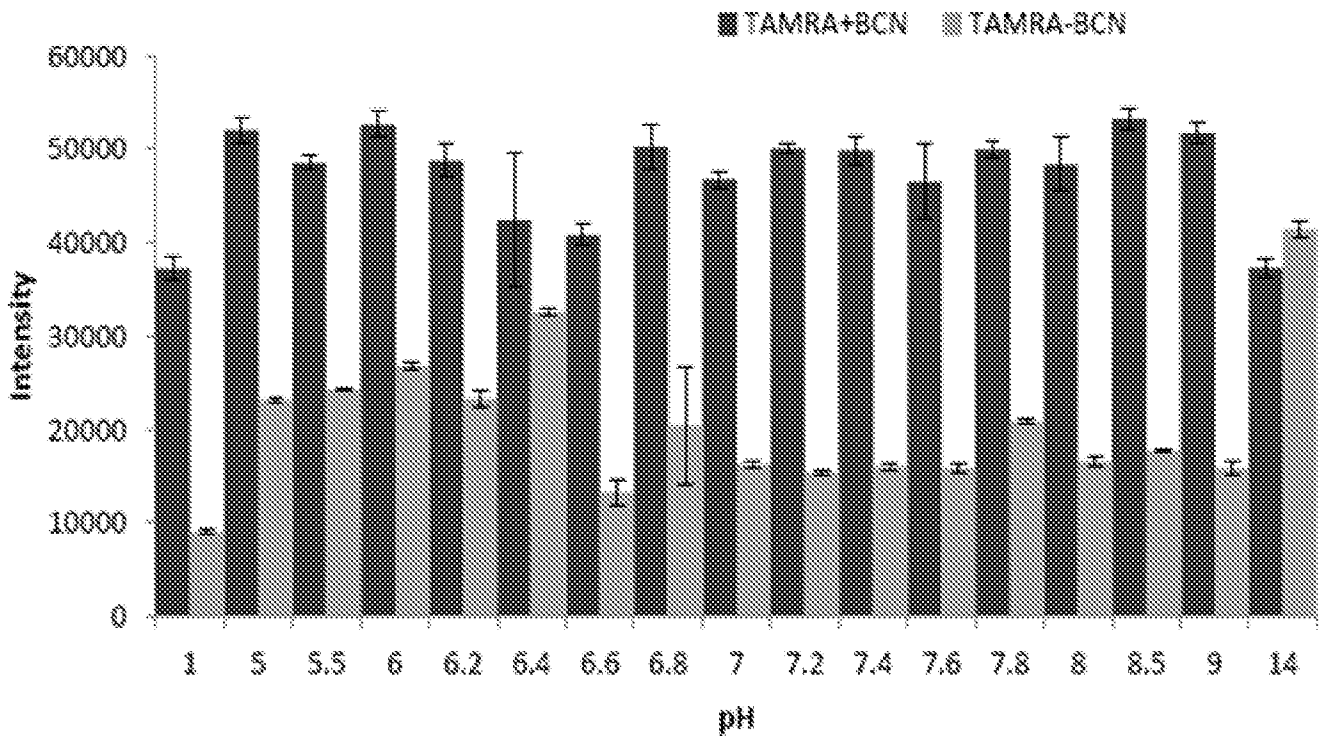


Figure 3T

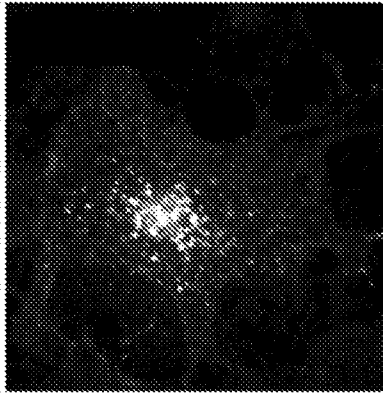


Figure 4A

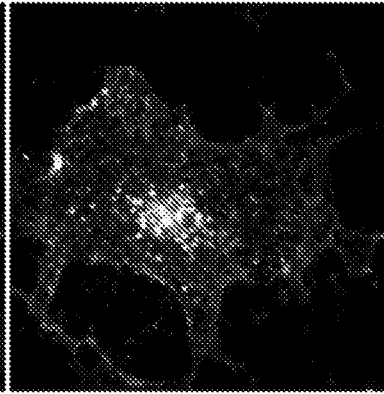


Figure 4B

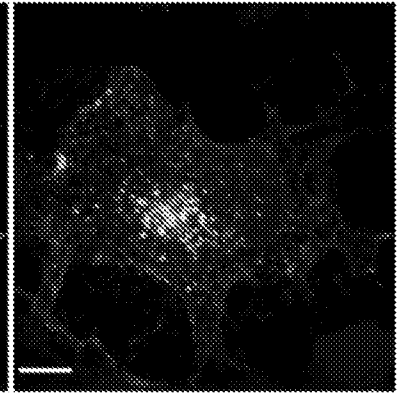


Figure 4C

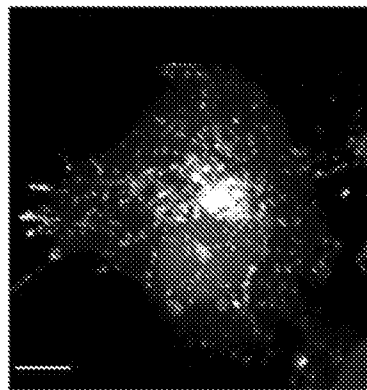


Figure 4D

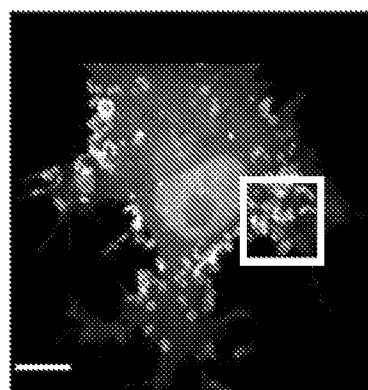


Figure 4E

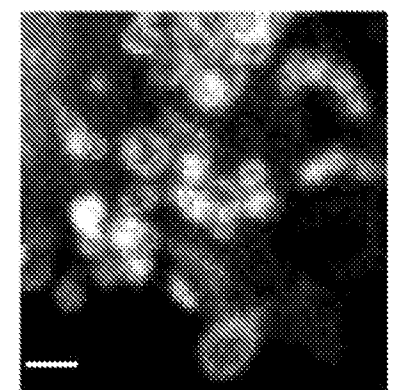


Figure 4F

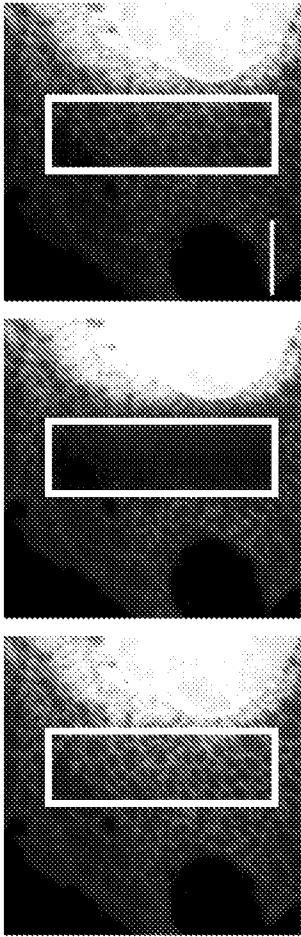


Figure 4O

Figure 4N

Figure 4M

Pre-bleach 0 3.8 11.3 18.8 26.3 33.8 41.3 48.8

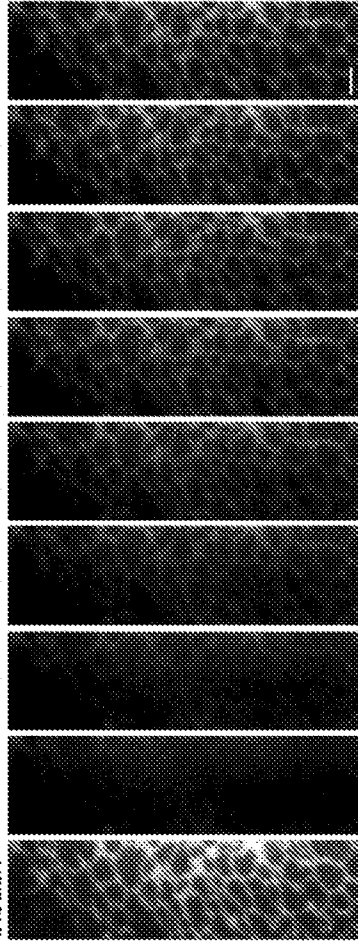


Figure 4P

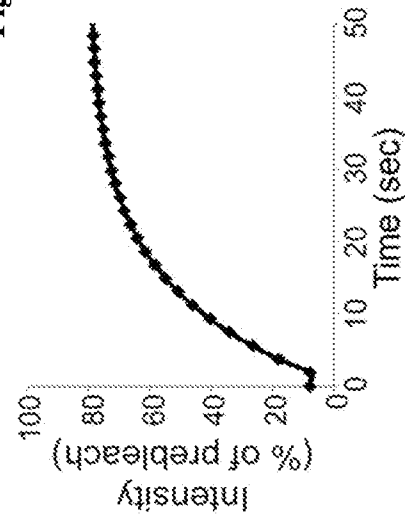


Figure 4Q

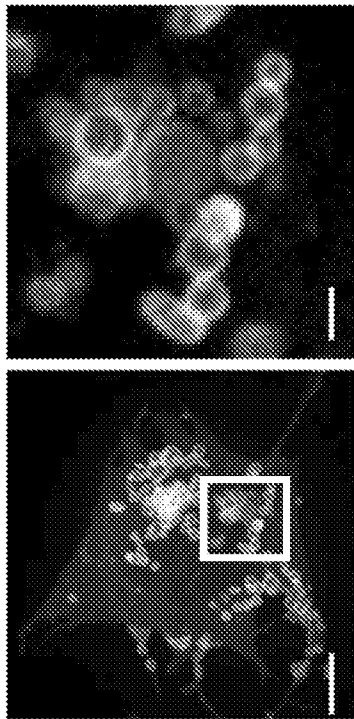


Figure 4H

Figure 4G

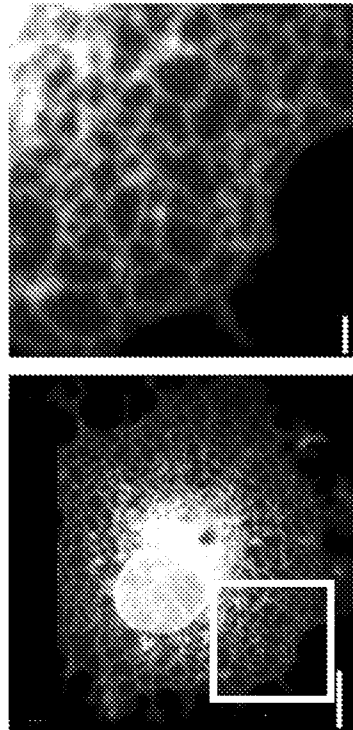


Figure 4J

Figure 4I

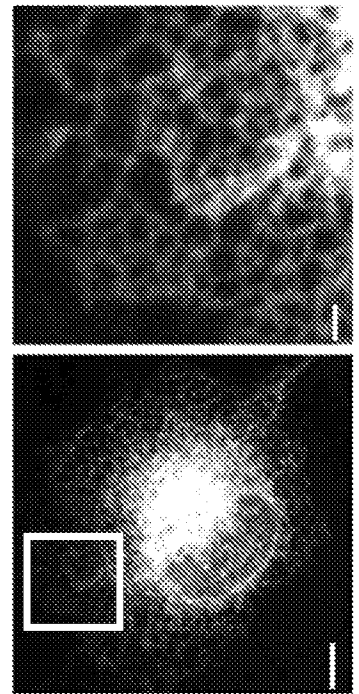


Figure 4L

Figure 4K

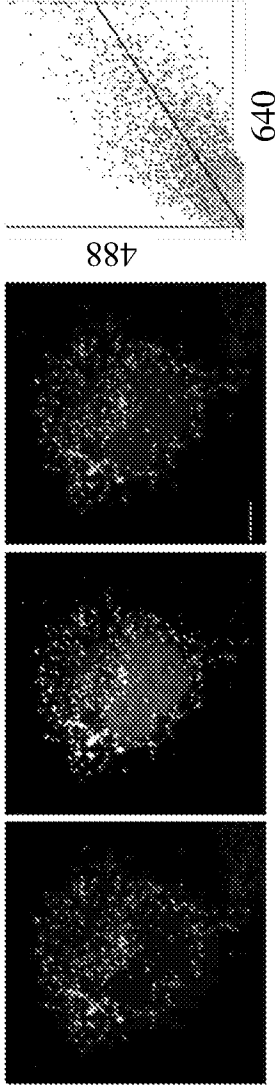
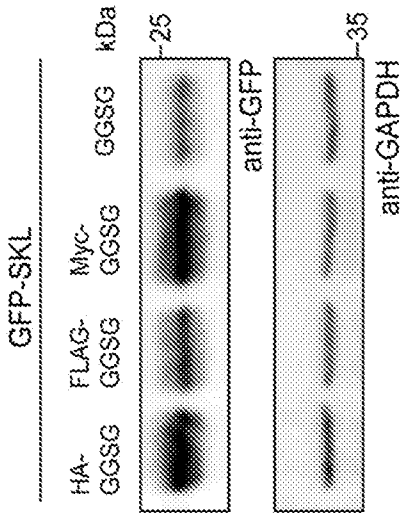


Figure 5A

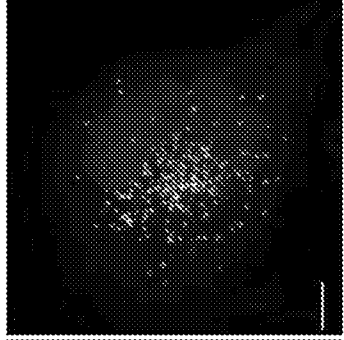
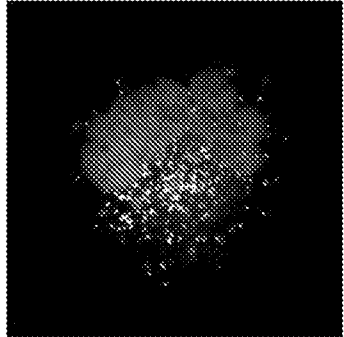
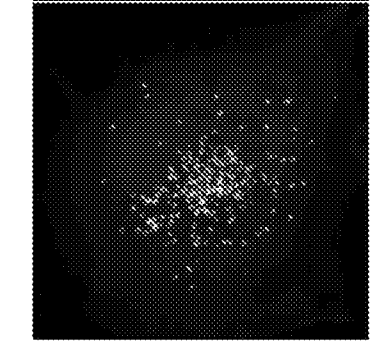
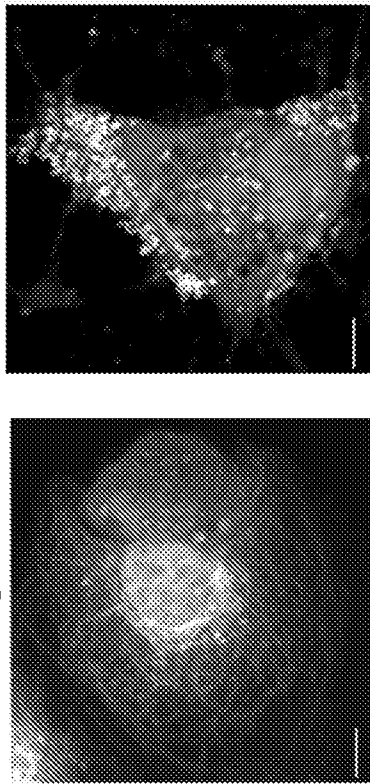


Figure 5F

Figure 5G

Figure 5H

Figure 5I

Figure 5J

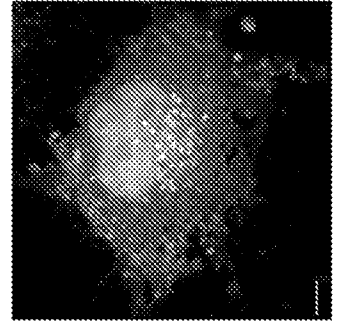
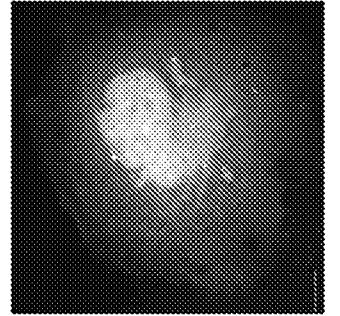
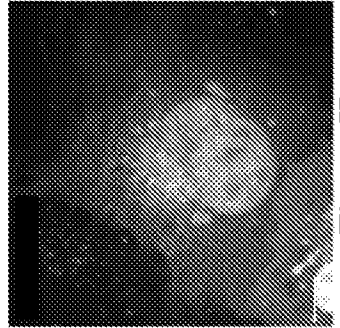
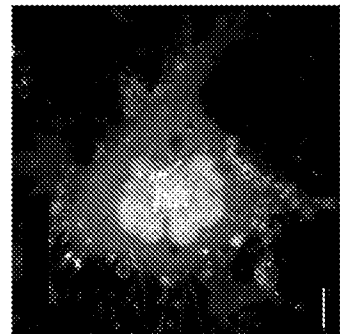


Figure 5K

Figure 5L

Figure 5M

Figure 5N

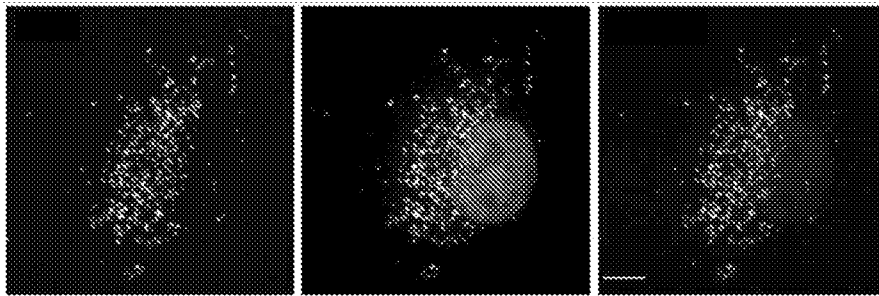


Figure 5O

Figure 5P

Figure 5Q

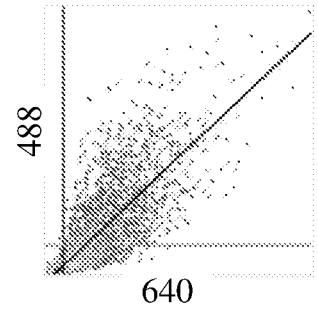


Figure 5R

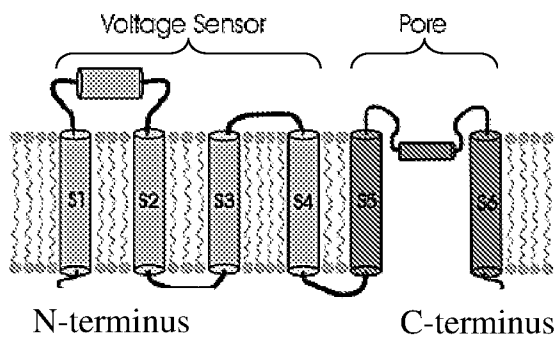


Figure 6A

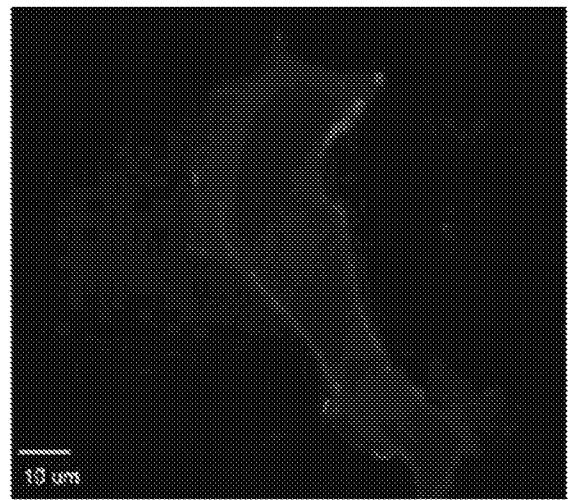


Figure 6B

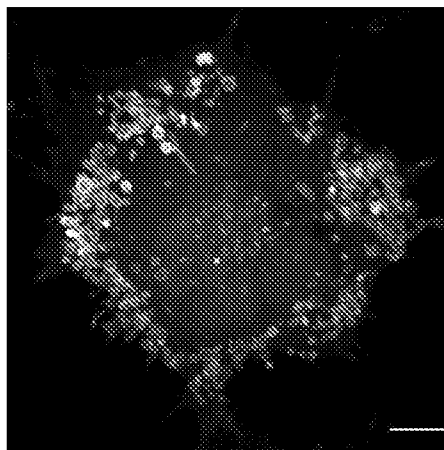


Figure 7A

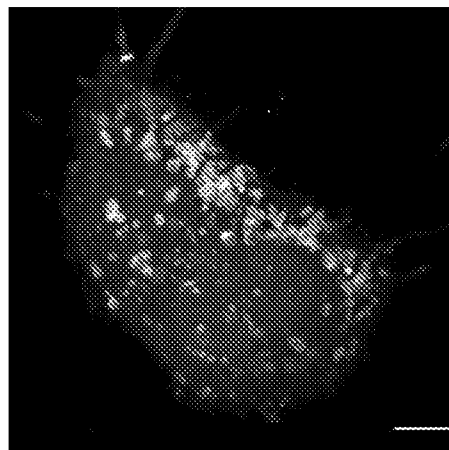


Figure 7B

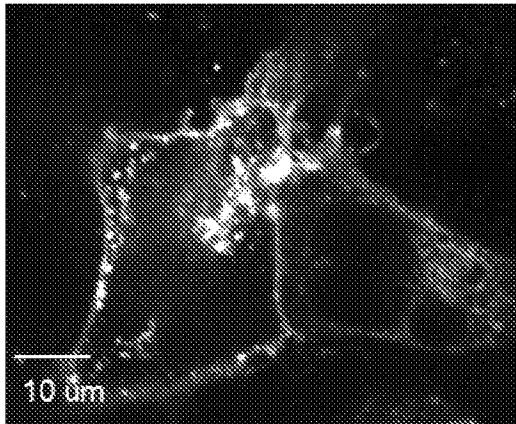


Figure 8A

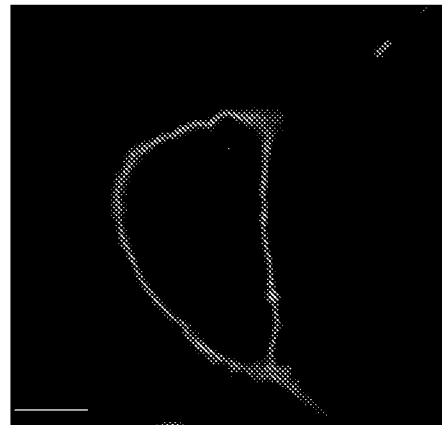


Figure 8B

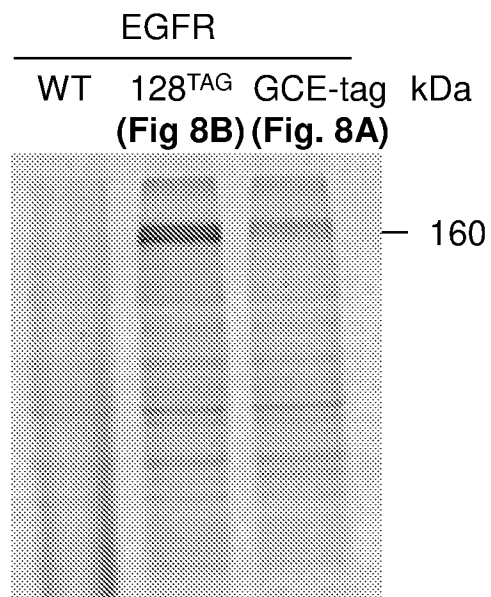


Figure 8C

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IL2020/050275

A. CLASSIFICATION OF SUBJECT MATTER See extra sheet.		
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) See extra sheet.		
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) See extra sheet.		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ALOUSH, Noa, et al. Live cell imaging of bioorthogonally labelled proteins generated with a single pyrrolysine tRNA gene. Scientific reports, 2018, 8.1: 14527. https://www.nature.com/articles/s41598-018-32824-1 . [Online] [Retrieved on 2019-09-15]. Retrieved from the Internet <URL: https://www.nature.com/articles/s41598-018-32824-1 > <DOI: 10.1038/s41598-018-32824-1 > 28 Sep 2018 (2018/09/28) abstract, figures 1A, 2A, page 2 fourth paragraph, page 9, "Discussion" third paragraph, figure 3A, plasmids h and i; page 3 under "Reducing the number of encoded PyIT genes increases the number of specifically labelled live cells",	1-19,23,24
X	WO 2016176689 A1 UNIV OREGON STATE [US]; MEHL RYAN A [US]; BLIZZARD ROBERT [US] 03 Nov 2016 (2016/11/03) page 15, second and third paragraph, page 8 last paragraph-page 9 first paragraph, example 3 on page 44, figure 10, page 17 last paragraph-page 18 first paragraph	1-3,6-19,23,24
X	WO 2013171485 A1 MEDICAL RES COUNCIL [GB]; CHIN JASON [GB]; LANG KATHRIN [GB] 21 Nov 2013 (2013/11/21) page 3 lines 22-28, page 3 second paragraph, page 4 lines 24-28, page 36, last paragraph, examples 2,7, figures 5b, 5c, claim 3	1-24
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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 "D" document cited by the applicant in the international application "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 10 May 2020	Date of mailing of the international search report 11 May 2020	
Name and mailing address of the ISA: Israel Patent Office Technology Park, Bldg.5, Malcha, Jerusalem, 9695101, Israel Email address: pctoffice@justice.gov.il	Authorized officer RON-COHEN Yael Telephone No. 972-73-3927142	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IL2020/050275

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>SEGAL, Inbar, et al. A straightforward approach for bioorthogonal labeling of proteins and organelles in live mammalian cells, using a short peptide tag. BMC Biology, 2020, 18.1: 1-16. [online] [retrieved on 2020-05-05]. Retrieved from the Internet <URL: https://link.springer.com/article/10.1186/s12915-019-0708-7><doi:10.1186/s12915-019-0708-7> 14 Jan 2020 (2020/01/14) The whole document.</p>	1-24

A. CLASSIFICATION OF SUBJECT MATTER:

IPC (20200101) C12N 15/11, C12N 15/63, C12N 15/66, C07K 1/00, C07K 1/13, G01N 33/52, G01N 33/68, G01N 33/58
CPC (20160501) C12N 15/11, C12Y 601/01, C07C 2602/14, C07K 2319/41, C07K 2319/42, C07K 2319/43, C12N 15/63, C12N 15/66, C07K 1/006, C07K 1/13, G01N 33/52, G01N 33/6803, G01N 33/582

B. FIELDS SEARCHED:

* Minimum documentation searched (classification system followed by classification symbols)

IPC (20200101) C12N 15/11, C12N 15/63, C12N 15/66, C07K 1/00, C07K 1/13, G01N 33/52, G01N 33/68, G01N 33/58
CPC (20160501) C12N 15/11, C12Y 601/01, C07C 2602/14, C07K 2319/41, C07K 2319/42, C07K 2319/43, C12N 15/63, C12N 15/66, C07K 1/006, C07K 1/13, G01N 33/52, G01N 33/6803, G01N 33/582

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

Databases consulted: Esp@cenet, Google Patents, BIOSIS, Google Scholar, DWPI, Derwent Innovation

Search terms used: non-canonical amino acid, ncAA, 3-Iodo-L-tyrosine, NE-Benzyloxycarbonyllysine, Zlys, NE Acetyllysine, AcLys, NE-Cyclopentylloxycarbon-Nitrobenzyloxycarbonyl-NE -L-lysine, NE-1-6 Nitrobenzo-d-1,3-dioxol-5yl-ethoxy-carbonyl-L-lysine, NE-2-3-Methyl-3Hdiazirin-3-yl-ethoxy-carbonyl-L-lysine, 3-3-Methyl-3Hdiazirine-3yl-propaminocarbonyl, NE-L-lysine, DiZPK, BCN, TCO, NE-1-Methylcycloprop-2-enecarboxamido-lysine, CpK or NE-Acryloyl-L-lysine, pNO2ZLys, TmdZLys or NE-Crotonyl-L-lysine, Kcr, 2-Chloro-L-phenylalanine, 2-Bromo-L-phenylalanine, 2-Iodo-L-phenylalanine, 2-Methyl-L-phenylalanine 2-Methoxy -L -phenylalanine, 2-Nitro-L-phenylalanine, 2-Cyano -L-phenylalanine, NE-tert-Butoxycarbonyl-L -lysine-R-BocLys, linker, GGSG, Gly-Gly-Ser, label, polynucleotide, Genetic code expansion, codon reassignment, cellular component, endoplasmic reticulum, plasma membrane, peroxisome, lysosome multi-vesicular bodies, MVBs, exosome, secretory pathway, endocytosis, applicant

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/IL2020/050275

Patent document cited search report	Publication date	Patent family member(s)	Publication Date
WO 2016176689 A1	03 Nov 2016	WO 2016176689 A1	03 Nov 2016
		EP 3288925 A1	07 Mar 2018
		EP 3288925 A4	20 Feb 2019
		US 2019077776 A1	14 Mar 2019
WO 2013171485 A1	21 Nov 2013	WO 2013171485 A1	21 Nov 2013
		BR 112014028588 A2	27 Jun 2017
		CA 2873793 A1	21 Nov 2013
		CN 104718285 A	17 Jun 2015
		CN 108410827 A	17 Aug 2018
		EP 2852664 A1	01 Apr 2015
		EP 2852664 B1	06 Sep 2017
		GB 201208875 D0	04 Jul 2012
		GB 201210303 D0	25 Jul 2012
		HK 1208704 A1	11 Mar 2016
		IN 9922DEN2014 A	14 Aug 2015
		JP 2015523965 A	20 Aug 2015
		JP 6343604 B2	13 Jun 2018
		US 2015148525 A1	28 May 2015