

Published in final edited form as:

*Nat Chem.* ; 4(4): 298–304. doi:10.1038/nchem.1250.

## Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction

Kathrin Lang<sup>1</sup>, Lloyd Davis<sup>1</sup>, Jessica Torres-Kolbus<sup>2</sup>, Chungjung Chou<sup>2</sup>, Alexander Deiters<sup>2,\*</sup>, and Jason W. Chin<sup>1,\*</sup>

<sup>1</sup>Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK

<sup>2</sup>Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695-8204, USA

### Abstract

The site-specific incorporation of bioorthogonal groups via genetic code expansion provides a powerful general strategy for site-specifically labelling proteins with any probe. However, the slow reactivity of the bioorthogonal functional groups that can be encoded genetically limits the utility of this strategy. We demonstrate the genetic encoding of a norbornene amino acid using the pyrrolysyl tRNA synthetase/tRNA<sub>CUA</sub> pair in *Escherichia coli* and mammalian cells. We developed a series of tetrazine-based probes that exhibit 'turn-on' fluorescence on their rapid reaction with norbornenes. We demonstrate that the labelling of an encoded norbornene is specific with respect to the entire soluble *E. coli* proteome and thousands of times faster than established encodable bioorthogonal reactions. We show explicitly the advantages of this approach over state-of-the-art bioorthogonal reactions for protein labelling *in vitro* and on mammalian cells, and demonstrate the rapid bioorthogonal site-specific labelling of a protein on the mammalian cell surface.

There is a pressing need for general methods to site-specifically label proteins, in diverse contexts, with user-defined probes. Current protein-labelling methods involve the use of fluorescent protein fusions<sup>1–4</sup>, self-labelling proteins (for example, SNAPtag, HALOtag and CLIPtag)<sup>5–8</sup>, ligases (such as biotin ligase, lipolic acid ligase, sortase and phosphopantetheinyl transferase)<sup>9–15</sup> and self-labelling tags (for example, tetracysteine and tetraserine)<sup>16,17</sup>. Although some of these approaches allow rapid labelling and have had substantial impact on biological studies, they require the use of protein fusions and/or the introduction of additional sequences into the protein of interest. This may disturb the structure and function of the protein and can make it challenging to place probes at any position in a protein. Moreover, the range of probes that can be incorporated by some of these techniques is limited<sup>3,4,18</sup>.

Ideal methods for protein labelling would (i) allow probes to be placed easily at any position in any protein expressed in prokaryotic and eukaryotic cells and organisms, (ii) be rapid and quantitative, (iii) be specific for a user-defined site in a protein, (iv) require no toxic reagents

\*alex\_deiters@ncsu.edu; chin@mrc-lmb.cam.ac.uk.

**Author contributions** K.L., L.D., J.T.K., C.C., A.D. & J.W.C. designed the research and analysed the data. K.L., L.D., J.T.K. and C.C. performed the experiments. K.L. and J.W.C. co-wrote the paper with input from the co-workers.

**Additional information** The authors declare no competing financial interests. Supplementary information and chemical compound information accompany this paper at [www.nature.com/naturechemistry](http://www.nature.com/naturechemistry). Reprints and permission information is available online at <http://www.nature.com/reprints>. Correspondence and requests for materials should be addressed to A.D. and J.W.C.

and generate no toxic by-products, (v) show 'turn-on' fluorescence with minimal off-site or background labelling and (vi) allow for labelling with diverse probes. In principle, the genetically encoded, site-specific incorporation of unnatural amino acids that bear bioorthogonal functional groups would allow the labelling of specific proteins at defined sites with essentially any probe.

Bioorthogonal groups, including azides, alkynes, ketones, anilines, alkenes, tetrazoles and 1,2-aminothiols, have been genetically encoded using amber suppressor aminoacyl tRNA synthetase/tRNA<sub>CUA</sub> pairs<sup>19–29</sup>. For established reactions that have been demonstrated on proteins, the rate constants for the corresponding model reactions<sup>30</sup> are in the range 10<sup>–2</sup> M<sup>–1</sup> s<sup>–1</sup> to 10<sup>–4</sup> M<sup>–1</sup> s<sup>–1</sup> (although for emerging approaches higher rates have been reported)<sup>29,31,32</sup>. Clearly, the rates of established reactions are sufficient to allow useful labelling of metabolically incorporated azido- and keto-bearing glycan analogues presented at high density on the cell surface, and the labelling of amino acid analogues incorporated nonspecifically throughout the proteome<sup>33–35</sup>. However, the sluggishness of established bioorthogonal reactions often makes it challenging to label proteins quantitatively at defined sites *in vitro* and may account for the fact that there are currently no examples of labelling proteins expressed on the mammalian cell surface using genetically encoded unnatural amino acids.

Recent advances in bioorthogonal chemistry demonstrate that strained alkenes, including norbornenes and *trans*-cyclooctenes, react rapidly and specifically with tetrazines in inverse electron-demand Diels–Alder cycloaddition reactions to form stable adducts. The rate constants of these reactions are orders of magnitude faster than those for other established bioorthogonal reactions<sup>36–38</sup>. Genetically encoding a component of these reactions is therefore an attractive strategy for realizing rapid and site-specific protein labelling (Fig. 1a).

## Results and discussion

### Synthesis and genetic encoding of a norbornene-containing amino acid

The pyrrolysyl-tRNA synthetase/tRNA<sub>CUA</sub> pair (PylRS/tRNA<sub>CUA</sub>) from *Methanosarcina* species (*M. barkeri* (*Mb*) and *M. mazei* (*Mm*)), which naturally incorporates pyrrolysine (**1**, Fig. 1b), is orthogonal to endogenous tRNAs and aminoacyl-tRNA synthetases in *Escherichia coli* and eukaryotic cells<sup>39–42</sup>. Using this PylRS/tRNA<sub>CUA</sub> pair, and its synthetically evolved derivatives, we and others directed the efficient incorporation of unnatural amino acids, including post-translationally modified amino acids, chemical handles and photocaged amino acids, at specific sites in desired proteins in *E. coli*, yeast and mammalian cells<sup>27,28,39,40,43–46</sup>. Moreover, recently we demonstrated the incorporation of unnatural amino acids, using this pair, in a whole animal<sup>42</sup>. We envisioned that this synthetase/tRNA pair might be used to site-specifically and quantitatively incorporate a norbornene-containing amino acid into proteins produced in diverse organisms, and that the norbornene-containing protein could be labelled rapidly and selectively with tetrazine-based probes.

We designed the norbornene-containing amino acid *Nε*-5-norbornene-2-yloxy-carbonyl-L-lysine (**2**, Fig. 1b) and synthesized it in three steps and 77% overall yield (Supplementary Information and Supplementary Scheme S1). To investigate whether **2** is a substrate for the *Mb*PylRS/tRNA<sub>CUA</sub> pair we transformed *E. coli* with *pBKPyIS* (which encodes *Mb*PylRS) and *psfGFP150TAGPyIT-His<sub>6</sub>* (which encodes *Mb*tRNA<sub>CUA</sub> and a C-terminally hexahistidine (His<sub>6</sub>)-tagged superfolder green fluorescent protein (sfGFP) gene with an amber codon at position 150). In the presence of **2** (1 mM), full-length sfGFP was isolated in good yield (Fig. 2, 4 mg l<sup>–1</sup> of culture), comparable to the yields obtained for other well-

incorporated unnatural amino acids<sup>28,32,45</sup>. GFP expression was clearly amino acid-dependent. Similarly, myoglobin (Myo) bearing an amber codon at position 4 and T4 lysozyme (T4L) bearing an amber codon at position 83 produced good yields of protein in the presence, but not absence, of **2** (Fig. 2 and Supplementary Fig. S1). The incorporation of **2** was further confirmed by electrospray ionization mass spectrometry (ESI-MS) of purified proteins (Fig. 2 and Supplementary Fig. S1).

### Synthesis of biocompatible tetrazines

To create unsymmetrical tetrazines that contain a unique reactive group for functionalization with biophysical probes (Fig. 1c, Supplementary Scheme S2 and Supplementary Information) we reacted equimolar quantities of 5-amino-2-cyanopyridine and 2-cyanopyridine (or 2-cyanopyrimidine) with an excess of aqueous hydrazine to obtain *s*-dihydropyridazines **S5a** and **S6a** (ref. 36). Treatment of these dihydropyridazines with a mixed anhydride formed *in situ* from isobutylchloroformate and *N*-*tert*-butyloxycarbonylglycine afforded compounds **S5b** and **S6b**, respectively, which were oxidized readily to their corresponding tetrazines **5** and **6** with sodium nitrate in acetic acid (direct oxidation of dihydropyridazines **S5a** and **S6a** produced tetrazines that bear amino groups with markedly reduced reactivity, as expected based on increased  $\pi$ -conjugation of the amino group lone pair in these tetrazines). Acidic deprotection of the *tert*-butyloxycarbonyl groups afforded tetrazines **S5c** and **S6c**. The primary amino group in these tetrazine derivatives provides a handle for further functionalization with biophysical probes.

We envisioned that analogues of **5** and **6** that bear a carboxy group in place of the amine would be more electron deficient and potentially more reactive in inverse electron-demand cycloaddition reactions with norbornenes. To create tetrazines **7** and **8**, we reacted *N*-*tert*-butyloxycarbonylethylenediamine with 6-cyanopyridine-3-carboxylic acid under standard amide-coupling conditions. The resulting nitrile **S7a** was reacted with acetonitrile or 2-cyanopyrimidine in aqueous hydrazine to give dihydropyridazines **S7b** and **S8b**, respectively, which after oxidation with sodium nitrate afforded tetrazines **7** and **8**, respectively. Deprotection of **8** under acidic conditions gave tetrazine **S8c**. The primary amino group in this tetrazine derivative provides a handle for further functionalization with biophysical probes. All the tetrazines synthesized were stable in MeOH/H<sub>2</sub>O and dimethylsulfoxide (DMSO)/H<sub>2</sub>O at room temperature for several days, as judged by liquid chromatography–mass spectrometry (data not shown).

### Kinetic analysis of the rapid tetrazine Diels–Alder cycloaddition

Tetrazines **5–8** react readily with 5-norbornene-2-ol to form the corresponding dihydropyridazines **S15** and its isomeric forms **S16** in protic solvents in >96% conversion (Supplementary Fig. S2 and Supplementary Information). The rate constants for these reactions were determined under pseudo first-order conditions by following the exponential decay over time in the ultraviolet absorbance of the tetrazine at 320 or 300 nm (Supplementary Fig. S3). The reactions were faster in more polar solvent systems, that is in solvent mixtures of higher water content, as expected<sup>36,47</sup>. Tetrazine **8** displays the highest activity towards 5-norbornene-2-ol with second-order rate constants of approximately 9 M<sup>-1</sup> s<sup>-1</sup> in H<sub>2</sub>O/MeOH (95:5) at 21 °C, whereas **5** reacts with a rate constant of approximately 1 M<sup>-1</sup> s<sup>-1</sup> under the same conditions (Supplementary Table S1 and Supplementary Information). This confirms that the tetrazine–norbornene reaction is orders of magnitude faster than established bioorthogonal reactions<sup>30</sup>.

### Tetrazine-based fluorophores: 'turn-on' fluorogenic probes

To create fluorescent probes based on **5**, **6** and **8**, the primary amino groups of **S5c**, **S6c** and **S8c** were conjugated to succinimidylesters or isothiocyanates of fluorescein, tetramethylrhodamine (TAMRA) and boron dipyrromethene (BODIPY) dyes (Supplementary Information, Supplementary Figs S4 and S5, Supplementary Table S2).

The fluorescence of the visible light emitting tetrazine conjugate **9** and BODIPY tetrazine conjugate **13** was reduced substantially with respect to the fluorescence of the succinimidyl or isothiocyanate derivatives of the parental fluorophores. This is in agreement with recent work that shows fluorophores can be quenched by energy transfer to a proximal tetrazine chromophore that absorbs between 510 and 530 nm (ref. 48). However, even though **5**, **6** and **8** have very similar absorption spectra, the reduction in fluorescence of the dye conjugates was dependent on the specific combination of tetrazine and fluorophore. For example, **9** (**5**-TAMRA-X) showed a much greater reduction in fluorescence with respect to the parent TAMRA-X than did **10** (**6**-TAMRA-X) and **12** (**8**-TAMRA-X). Fluorescein (emission maximum at 518 nm) was quenched minimally by conjugation to **8**. The fluorescence of **9**, **11** and **13** was turned on upon cycloaddition with 5-norbornene-2-ol, leading to a 5–10 fold gain in fluorescence intensity (Fig. 3a, Supplementary Fig. S5).

### Rapid labelling of norbornene-containing proteins with tetrazine-based probes

To demonstrate that the tetrazine–dye probes react efficiently and specifically with recombinant proteins that bear site-specifically incorporated **2**, purified sfGFP-**2**, Myo-**2** and T4L-**2** were incubated overnight with fluorophore **9** (10 equiv.) at room temperature. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE)-based fluorescence imaging and ESI–MS analysis (Fig. 3a and Supplementary Fig. S7) confirmed the quantitative labelling of the proteins that contained **2**, whereas no nonspecific labelling was detected with the control proteins that contained *Ne-tert*-butyloxycarbonyl-L-lysine (**3**) in place of **2** at the same site. In additional experiments we showed, using mass spectrometry, the specific and quantitative labelling of proteins that contained **2** with tetrazine derivatives **5**, **6** and **8**, as well as with tetrazine fluorophores **12**, **13** and **14** (Supplementary Figs S6 and S7). Previous labelling experiments of proteins containing unnatural amino acids with specific fluorophores required washing steps to remove free dye non-covalently associated with the protein. We found that we could image the specific labelling of proteins that contained **2** without washing the sample or the gel; this improvement may, at least in part, result from the 'turn-on' fluorescence of the tetrazine fluorophores.

To further probe the specificity of the reaction between the genetically encoded norbornene and the tetrazine-based fluorophores, we performed the labelling reaction on the proteome of *E. coli* expressing either sfGFP-**2**-His<sub>6</sub> or Myo-**2**-His<sub>6</sub> (Fig. 3c and Supplementary Fig. S8). We controlled the level of recombinant protein expression so that it was equal to or less than that of many endogenous proteins by modulating the concentration of **2** added to cells. This ensures that any specific labelling of the target protein versus native proteins was not an artefact of the abundance of the target protein. Cells were harvested 3.5 hours after the induction of protein expression, washed with PBS and incubated with fluorophore probes (**12** or **13**) at room temperature. After washing the cell pellets, the cells were lysed and the reaction mixtures were analysed by SDS–PAGE to assess the protein levels. Fluorescence scanning of SDS–PAGE gels revealed that the tetrazine–norbornene cycloaddition is highly specific for **2** with respect to other soluble *E. coli* proteins. The low background fluorescence suggests that any nonspecific labelling of cellular proteins is minimal. Cells that incorporate amino acid **3**, or no amino acid, in place of **2** at the amber codon show a comparable background labelling profile (compare lanes 11 and 12 to lanes 13–15 and lanes

16 and 17 to lanes 18–20 in Fig. 3c). This suggests that background fluorescence arising from the reaction of tetrazine probes with **2** incorporated at the C-terminus of endogenous proteins in response to amber codons is a minimal contributor to the observed background in these cellular labelling experiments.

To show that the high rate constants measured on small molecules translate into rapid protein labelling, we labelled myoglobin bearing **2** at position 4 with **12** (10 equiv.). In-gel fluorescence imaging of the labelling reaction as a function of time (Fig. 3d) demonstrates that the reaction is complete in approximately 30 minutes. Rapid labelling of proteins that incorporate **2** was also observed with probes **9** and **12** (Supplementary Fig. S9). In contrast, the labelling of an alkyne-containing amino acid at the same site in myoglobin required 50 equiv. of azide fluorophore and 18 hours to reach completion in a copper-catalysed [3 + 2] click reaction<sup>28</sup>. This demonstrates that the labelling method we report here has a clear advantage for the labelling of recombinant proteins.

### Site-specific protein labelling on the mammalian cell surface

Although it has been possible to label abundant molecules at multiple chemical handles on cell surfaces via metabolic incorporation of bioorthogonal functional groups<sup>33–35</sup>, there are no reports of labelling single, genetically defined sites on proteins on the mammalian cell surface using any of the unnatural amino acids that currently can be genetically encoded.

We demonstrated that **2** can be encoded genetically with high efficiency into proteins in mammalian cells using the *MmPyIRS*/tRNA<sub>CUA</sub> pair by western blot, fluorescence imaging and mass spectrometry<sup>46</sup> (Fig. 4a,b and Supplementary Fig. S10). To show the site-specific labelling of a mammalian protein, we introduced an amber codon into an epidermal growth factor receptor (EGFR)-GFP fusion gene at position 128 in the extracellular portion of the receptor in a vector that contained *MmPyIRS*, creating *pMmPyIRS-EGFR(128TAG)-GFP-HA*. We transfected HEK293 cells with *pMmPyIRS-EGFR(128TAG)-GFP-HA* and *p4CMVE-U6-PyIT* that encodes four copies of the *MmPyItRNA*<sub>CUA</sub>. In the presence of **2** or **3**, cells produced full-length EGFR-GFP that can be visualized at the cell membrane by fluorescence microscopy. To demonstrate the specific labelling of EGFR-GFP that contained **2** with tetrazine fluorophores we treated cells with **9** (200 nM), washed the cells and imaged the red fluorescence arising from TAMRA labelling as well as the green fluorescence arising from expression of full-length EGFR-GFP, in which the C-terminal GFP is intracellular. Clear labelling of cells that bear EGFR-**2**-GFP was observed within two hours and TAMRA fluorescence clearly co-localized with cell-surface EGFR-GFP fluorescence. No labelling was observed for cells in the same sample that did not express EGFR-GFP, and cells bearing EGFR-**3**-GFP were not labelled with **9**. These observations confirm that **2** at position 128 of EGFR was labelled specifically with the tetrazine-TAMRA conjugate **9** (Fig. 4c and Supplementary Figs S11–S14).

Next we aimed to compare the site-specific tetrazine labelling of **2** on the surface of mammalian cells with the labelling of a site-specifically incorporated azide, using a cyclooctyne, via a reaction previously employed to successfully label azides installed into cell-surface glycans and throughout the proteome<sup>33,34</sup>. We first demonstrated that an azide-containing amino acid, *N*<sub>ε</sub>-(2-azidoethoxycarbonyl-L-lysine) (**4**) (Fig. 1b), can be incorporated efficiently into proteins in mammalian cells using the *PyIRS*/tRNA<sub>CUA</sub> pair (Supplementary Fig. S15). We then incorporated **4** into EGFR-GFP at position 128 with an efficiency comparable to that for the incorporation of **2**, as judged by GFP fluorescence. However, when we attempted to label **4** with a cyclooctyne-based fluorophore (**S17**, TAMRA-DIBO-alkyne, commercially available from Invitrogen, Supplementary Fig. S4) under conditions identical to those used to label **2** with tetrazine fluorophores, we did not



observe specific labelling (Supplementary Fig. S16). Similarly, when we attempted to label **4** under conditions provided by the supplier we did not observe specific labelling of cell-surface EGFR (Supplementary Fig. S17). These results suggest that the norbornene-tetrazine reaction provides a clear advantage over established bioconjugation reactions with unnatural amino acids for protein labelling on the mammalian cell surface.

## Conclusions and outlook

In conclusion, we report the efficient synthesis and site-specific, genetically encoded incorporation of the norbornene-containing amino acid **2** into proteins in *E. coli* and mammalian cells. We describe the development of a series of tetrazine-based probes that exhibit 'turn-on' fluorescence on their rapid reaction with norbornenes. We demonstrate that proteins that bear **2** can be specifically labelled *in vitro*, in complex mixtures and on the surface of mammalian cells, and explicitly demonstrate the advantages of this approach for site-specific protein labelling. We are currently exploring extensions of the approaches described here for imaging site-specifically labelled proteins in cells and whole organisms to provide new biological insights. In addition, we are pursuing the discovery and genetic encoding of new, rapid bioorthogonal chemistries in proteins.

## Methods

Protocols for the chemical syntheses of norbornene lysine **2** and various tetrazine probes, as well as protocols on mass spectrometry, determination of kinetic rate constants and cloning for mammalian cells are given in the Supplementary Information.

## Protein expression and purification

To express sfGFP with an incorporated unnatural amino acid, we transformed *E. coli* DH10B cells with *pBKPyIS* (which encodes *MbPyIRS*) and *psfGFP150TAGPyIT-His<sub>6</sub>* (which encodes *MbtRNA<sub>CUA</sub>* and a C-terminally hexahistidine tagged sfGFP gene with an amber codon at position 150). Cells were recovered in 1 ml of suboptimal broth media (supplemented with 0.2% glucose) for one hour at 37 °C before incubation (16 hours, 37 °C, 230 revolutions per minute (r.p.m.)) in 100 ml of Luria broth (LB) that contained kanamycin (50 µg ml<sup>-1</sup>) and tetracycline (25 µg ml<sup>-1</sup>). Of this overnight culture, 20 ml was used to inoculate one litre of LB supplemented with kanamycin (25 µg ml<sup>-1</sup>) and tetracycline (12 µg ml<sup>-1</sup>), and then incubated at 37 °C. At  $D_{600}$  0.4 to 0.5, a solution of **2** in H<sub>2</sub>O was added to give a final concentration of 2 mM. After 30 minutes, protein expression was induced by the addition of arabinose to a final concentration of 0.2%. After three hours of induction, cells were harvested by centrifugation and frozen at -80 °C until required. Cells were thawed on ice and suspended in 30 ml of lysis buffer (Tris-HCl (10 mM), imidazole (20 mM), NaCl (200 mM), pH 8, phenylmethanesulfonylfluoride (1 mM), lysozyme (1 mg ml<sup>-1</sup>), deoxyribonuclease A (100 µg ml<sup>-1</sup>), Roche protease inhibitor).

Proteins were extracted by sonication at 4 °C. The extract was clarified by centrifugation (20 minutes, 21,000g, 4 °C), Ni<sup>2+</sup>-NTA beads (Qiagen) (600 µl) were added to it and the mixture was incubated with agitation for one hour at 4 °C. Beads were collected by centrifugation (ten minutes, 1,000g). The beads were resuspended three times in 30 ml of wash buffer (Tris-HCl (10 mM), imidazole (20 mM), NaCl (200 mM), pH 8) and spun down at 1,000g. Subsequently, the beads were resuspended in 10 ml of wash buffer and transferred to a column. The protein was eluted with 3 ml of wash buffer supplemented with 200 mM imidazole and further purified by size-exclusion chromatography employing a HiLoad 16/60 Superdex 75 Prep Grade column (GE Life Sciences) at a flow rate of 1 ml min<sup>-1</sup> (buffer: Tris-HCl (20 mM), NaCl (100 mM), pH 7.4). Fractions that contained the protein were pooled and concentrated with an Amicon Ultra-15 3 kDa MWCO centrifugal filter device

(Millipore). Purified proteins were analysed by 4–12% SDS–PAGE and their mass confirmed by mass spectrometry (see Supplementary Information). Sperm-whale myoglobin and T4 lysozyme with incorporated **2** were prepared in the same way, except that cells were transformed with p*Myo4TAGPyIT*-His<sub>6</sub> (which encodes *Mb*tRNA<sub>CUA</sub> and a C-terminally hexahistidine-tagged sperm-whale myoglobin gene with an amber codon at position 4) and pBKPyIS or p*T4L83TAGPyIT*-His<sub>6</sub> (which encodes *Mb*tRNA<sub>CUA</sub> and a C-terminally hexahistidine-tagged T4 lysozyme gene with an amber codon at position 83) and pBKPyIS. Yields of purified proteins were up to 4 mg l<sup>-1</sup>.

### Protein labelling via tetrazine–norbornene cycloaddition

For *in vitro* labelling of purified proteins with tetrazine–dye conjugates, purified recombinant proteins with site-specifically incorporated **2**, sfGFP-**2**, Myo-**2**, T4L-**2** (all ~10 μM in Tris-HCl (20 mM), NaCl (100 mM), pH 7.4) were incubated with 10 equiv. of the tetrazine–dye conjugate **9** (2 mM in DMSO). The solution was incubated at room temperature and aliquots were taken after 12 hours and analysed by SDS–PAGE and (after desalting with a C4-ZIPTIP) by ESI-MS. The SDS–PAGE gels were either stained with Coomassie or scanned with a Typhoon imager to visualize in-gel fluorescence.

For *in vitro* labelling of purified proteins with tetrazine–dye conjugates as a function of time, 2 nmol of purified Myo-**2** (10 μM in Tris-HCl (20 mM), NaCl (100 mM), pH 7.4) was incubated with 20 nmol of tetrazine–dye conjugate **12** (10 μl of a 2 mM solution in DMSO). At different time points (0, 30 s, 1 min, 2 min, 3 min, 5 min, 10 min, 30 min, 1 h, 2 h) 8 μl aliquots were taken from the solution, quenched with a 200-fold excess of 5-norbornene-2-ol and plunged into liquid nitrogen. Samples were mixed with NuPAGE lithium dodecyl sulfate (LDS) sample buffer supplemented with 5% β-mercaptoethanol, heated for ten minutes to 90 °C and analysed by 4–12% SDS–PAGE. The amounts of labelled proteins were quantified by scanning the fluorescent bands with a Typhoon Trio phosphoimager (GE Life Sciences). Bands were quantified with the ImageQuant TL software (GE Life Sciences) using rubber-band background subtraction. In-gel fluorescence showed that labelling was complete within 30 minutes using 10 equiv. tetrazine–fluorophore **12** (Fig. 3d). In a similar experiment, sfGFP-**2** was incubated with tetrazine–fluorophore **12** or **9** and samples analysed at different time points (Supplementary Fig. S9).

To label the whole *E. coli* proteome with tetrazine–dye conjugates, *E. coli* DH10B cells that contained either p*sfGFP150TAGPyIT*-His<sub>6</sub> and pBKPyIS or p*Myo4TAGPyIT*-His<sub>6</sub> and pBKPyIS were inoculated into LB that contained kanamycin (50 μg ml<sup>-1</sup>) and tetracycline (25 μg ml<sup>-1</sup>). The cells were incubated with shaking overnight at 37 °C at 250 r.p.m. Of this overnight culture, 2 ml were used to inoculate 100 ml of LB supplemented with kanamycin (25 μg ml<sup>-1</sup>) and tetracycline (12 μg ml<sup>-1</sup>) and incubated at 37 °C. At OD<sub>600</sub> 0.5, 3 ml culture aliquots were removed and supplemented with different concentrations (1 mM, 2 mM and 5 mM) of **2** or 1 mM of **3**. After 30 minutes of incubation with shaking at 37 °C, protein expression was induced by the addition of 30 μl of 20% arabinose. After 3.5 hours of expression, cells were collected by centrifugation (16,000g, five minutes) of 1 ml of cell suspension. The cells were resuspended in PBS buffer, spun down again and the supernatant discarded. This process was repeated twice more. Finally, the washed cell pellet was suspended in 100 μl PBS and incubated with 3 μl of tetrazine–dye conjugate **12** or **13** (2 mM in DMSO) at room temperature overnight. The cells were collected again by centrifugation and washed twice with 1 ml PBS by suspending and centrifugation. Finally, the cells were resuspended in 100 μl of NuPAGE LDS sample buffer supplemented with 5% β-mercaptoethanol, heated at 90 °C for ten minutes and centrifuged at 16,000g for ten minutes. The crude cell lysate was analysed by 4–12% SDS–PAGE to assess protein levels. Gels were either Coomassie stained or scanned with a Typhoon imager to make fluorescent

bands visible. Western blots were performed with antibodies against the hexahistidine tag (Cell Signaling Technology, His tag 27E8 mouse mAb #2366).

### Incorporation of **2** in mammalian cells

HEK293 cells were seeded onto a Corning 96-well plate and grown to approximately 90% confluence in 10% fetal bovine serum Dulbecco's modified eagle medium (FBS/DMEM) with penicillin/streptomycin. Cells were transfected with two plasmids, *pMmPyIRS-mCherry-TAG-EGFP-HA* and *p4CMVE-U6-PyIT*, which contain four copies of the wild-type pyrrolysyl tRNA (see Supplementary Information). Transfection was carried out using the Lipofectamine 2000 Transfection Reagent from Invitrogen, according to the manufacturer's protocol. The growth media in which the cells were transfected was 10% FBS/DMEM, and it contained either **2** (1 mM), **3** (1 mM) or no additional amino acid, as indicated. Cells were imaged on a Zeiss 710 laser-scanning microscope to assay eGFP and mCherry expression after 16–24 hours. Cells were then lysed using 1X Reporter Lysis Buffer (Promega) supplemented with CompleteMini protease inhibitor cocktail (Roche). After lysis the cell debris was pelleted and the supernatant that contained soluble proteins removed and added to 4X NuPage LDS sample buffer (Invitrogen). Samples were loaded and run out by SDS-PAGE. Western blotting was carried out to detect full-length reporter protein using rabbit anti-HA (Sigma) antibody, detected with an anti-rabbit HRP conjugate (Cell Signaling). As a transfection control, Western blotting was also carried out to detect the synthetase using a mouse anti-FLAG antibody (AbFrontier) detected by an HRP-conjugated anti-mouse secondary antibody (Cell Signaling). MS/MS analysis was carried out to show the incorporation of **2** (see Supplementary Information).

### Labelling in mammalian cells

Cells were seeded and grown on 35 mm microdishes (Ibidi) coated with poly-L-lysine (Sigma). At ~90% confluence, cells were transfected using Lipofectamine 2000 (Invitrogen) with two plasmids, *p4CMVE-U6-PyIT* and *pMmPyIRS-EGFR(128TAG)-GFP-HA*. The transfection was carried out in DMEM with 0.1% FBS and containing 1 mM of **2**, **3** or **4**, as indicated. After transfection, cells were grown for 16 hours and then incubated in amino acid-free DMEM with 0.1% FBS for 2–5 hours. Then, the hEGFR-eGFP fusion was labelled with 200 nM of tetrazine-dye conjugate **9** (tet1-TAMRA-X) for 2–16 hours, as indicated, washed for ten minutes in DMEM with 0.1% FBS and imaged on a Zeiss LSM 780 or Zeiss LSM 710 laser scanning microscope with a Plan Apochromat 63× oil-immersion objective and using a 1× or 2× scan zoom, averaging 16. An Argon laser (488 nm) was used to excite eGFP, which was detected between 493 nm and 554 nm. TAMRA was excited using a diode-pumped solid state 561 nm laser and detected at 566–685 nm. Cells transfected in the presence of amino acid **4** were grown for 16 to 24 hours after transfection. According to the suppliers protocols, cells were washed in Dulbecco's phosphate buffered saline (DPBS) with 1% FBS, incubated with DiBO-TAMRA dye (Invitrogen) in DPBS with 1% FBS for 16 hours, washed four times in DPBS 1% FBS and imaged in DPBS 1% FBS.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

We thank the Medical Research Council (U105181009, UD99999908), the European Research Council and the National Institutes for Health (GM079114) for funding. J.T.K. was supported by the National Science Foundation Graduate Research Fellowship under Grant No. NSF 0750733. We thank R. Mehl, who pointed out developments in inverse electron-demand Diels-Alder reactions while on sabbatical in the Chin lab.

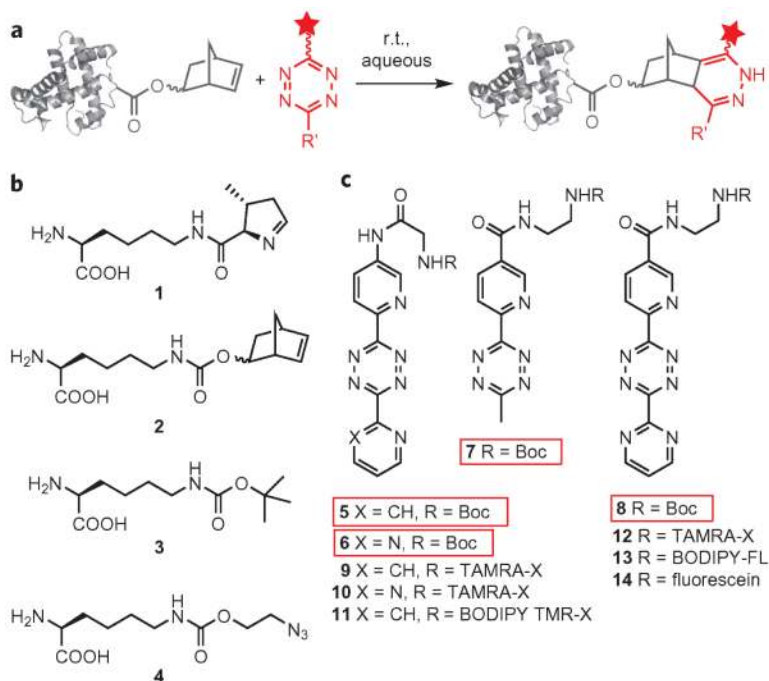


## References

1. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. *Science*. 1994; 263:802–805. [PubMed: 8303295]
2. Heim R, Prasher DC, Tsien RY. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl Acad. Sci. USA*. 1994; 91:12501–12504. [PubMed: 7809066]
3. Giepmans BN, Adams SR, Ellisman MH, Tsien RY. The fluorescent toolbox for assessing protein location and function. *Science*. 2006; 312:217–224. [PubMed: 16614209]
4. Shaner NC, Steinbach PA, Tsien RY. A guide to choosing fluorescent proteins. *Nature Methods*. 2005; 2:905–909. [PubMed: 16299475]
5. Los GV, et al. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol*. 2008; 3:373–382. [PubMed: 18533659]
6. Keppler A, et al. A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nature Biotechnol*. 2003; 21:86–89. [PubMed: 12469133]
7. Kosaka N, et al. *In vivo* stable tumor-specific painting in various colors using dehalogenase-based protein-tag fluorescent ligands. *Bioconjug. Chem*. 2009; 20:1367–1374. [PubMed: 19514716]
8. Gautier A, et al. An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol*. 2008; 15:128–136. [PubMed: 18291317]
9. George N, Pick H, Vogel H, Johnsson N, Johnsson K. Specific labeling of cell surface proteins with chemically diverse compounds. *J. Am. Chem. Soc*. 2004; 126:8896–8897. [PubMed: 15264811]
10. Zhou Z, Koglin A, Wang Y, McMahon AP, Walsh CT. An eight residue fragment of an acyl carrier protein suffices for post-translational introduction of fluorescent pantetheinyl arms in protein modification *in vitro* and *in vivo*. *J. Am. Chem. Soc*. 2008; 130:9925–9930. [PubMed: 18593165]
11. Yin J, et al. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc. Natl Acad. Sci. USA*. 2005; 102:15815–15820. [PubMed: 16236721]
12. Fernandez-Suarez M, et al. Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nature Biotechnol*. 2007; 25:1483–1487. [PubMed: 18059260]
13. Uttamapinant C, et al. A fluorophore ligase for site-specific protein labeling inside living cells. *Proc. Natl Acad. Sci. USA*. 2010; 107:10914–10919. [PubMed: 20534555]
14. Popp MW, Antos JM, Grotenbreg GM, Spooner E, Ploegh HL. Sortagging: a versatile method for protein labeling. *Nature Chem. Biol*. 2007; 3:707–78. [PubMed: 17891153]
15. Antos JM, et al. Site-specific *N*- and *C*-terminal labeling of a single polypeptide using sortases of different specificity. *J. Am. Chem. Soc*. 2009; 131:10800–10801. [PubMed: 19610623]
16. Griffin BA, Adams SR, Tsien RY. Specific covalent labeling of recombinant protein molecules inside live cells. *Science*. 1998; 281:269–272. [PubMed: 9657724]
17. Halo TL, Appelbaum J, Hobert EM, Balkin DM, Schepartz A. Selective recognition of protein tetraserine motifs with a cell-permeable, pro-fluorescent bis-boronic acid. *J. Am. Chem. Soc*. 2009; 131:438–439. [PubMed: 19105691]
18. Hinner MJ, Johnsson K. How to obtain labeled proteins and what to do with them. *Curr. Opin. Biotechnol*. 2010; 21:766–776. [PubMed: 21030243]
19. Chin JW, et al. Addition of *p*-azido-l-phenylalanine to the genetic code of *Escherichia coli*. *J. Am. Chem. Soc*. 2002; 124:9026–9027. [PubMed: 12148987]
20. Zhang Z, Wang L, Brock A, Schultz PG. The selective incorporation of alkenes into proteins in *Escherichia coli*. *Angew. Chem. Int. Ed*. 2002; 41:2840–2842.
21. Chin JW, et al. An expanded eukaryotic genetic code. *Science*. 2003; 301:964–967. [PubMed: 12920298]
22. Deiters A, et al. Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*. *J. Am. Chem. Soc*. 2003; 125:11782–11783. [PubMed: 14505376]
23. Deiters A, Cropp TA, Summerer D, Mukherji M, Schultz PG. Site-specific PEGylation of proteins containing unnatural amino acids. *Bioorg. Med. Chem. Lett*. 2004; 14:5743–5745. [PubMed: 15501033]

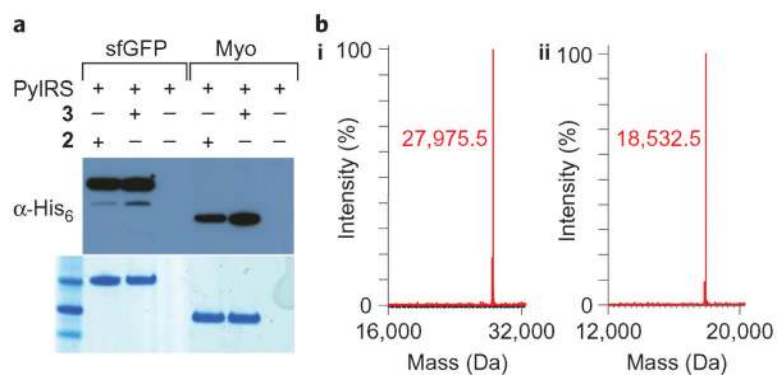
24. Mehl RA, et al. Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* 2003; 125:935–939. [PubMed: 12537491]
25. Wang L, Zhang Z, Brock A, Schultz PG. Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc. Natl Acad. Sci. USA.* 2003; 100:56–61. [PubMed: 12518054]
26. Carrico ZM, Romanini DW, Mehl RA, Francis MB. Oxidative coupling of peptides to a virus capsid containing unnatural amino acids. *Chem. Commun.* 2008:1205–1207.
27. Fekner T, Li X, Lee MM, Chan MK. A pyrrolysine analogue for protein click chemistry. *Angew. Chem. Int. Ed.* 2009; 48:1633–1635.
28. Nguyen DP, et al. Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNA(CUA) pair and click chemistry. *J. Am. Chem. Soc.* 2009; 131:8720–8721. [PubMed: 19514718]
29. Wang Y, Song W, Hu WJ, Lin Q. Fast alkene functionalization in vivo by photoclick chemistry: HOMO lifting of nitrile imine dipoles. *Angew. Chem. Int. Ed.* 2009; 48:5330–5333.
30. Agard NJ, Baskin JM, Prescher JA, Lo A, Bertozzi CR. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* 2006; 1:644–648. [PubMed: 17175580]
31. Wang J, et al. A biosynthetic route to photoclick chemistry on proteins. *J. Am. Chem. Soc.* 2010; 132:14812–14818. [PubMed: 20919707]
32. Nguyen DP, Elliott T, Holt M, Muir TW, Chin JW. Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. *J. Am. Chem. Soc.* 2011; 133:11418–11421. [PubMed: 21736333]
33. Laughlin ST, Bertozzi CR. Imaging the glycome. *Proc. Natl Acad. Sci. USA.* 2009; 106:12–17. [PubMed: 19104067]
34. Prescher JA, Bertozzi CR. Chemical technologies for probing glycans. *Cell.* 2006; 126:851–854. [PubMed: 16959565]
35. Johnson JA, Lu YY, Van Deventer JA, Tirrell DA. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr. Opin. Biotechnol.* 2010; 14:774–780.
36. Blackman ML, Royzen M, Fox JM. Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels–Alder reactivity. *J. Am. Chem. Soc.* 2008; 130:13518–13519. [PubMed: 18798613]
37. Devaraj NK, Weissleder R, Hilderbrand SA. Tetrazine-based cycloadditions: application to pretargeted live cell imaging. *Bioconjug. Chem.* 2008; 19:2297–2299. [PubMed: 19053305]
38. Devaraj NK, Weissleder R. Biomedical applications of tetrazine cycloadditions. *Acc. Chem. Res.* 2011; 44:816–827. [PubMed: 21627112]
39. Mukai T, et al. Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* 2008; 371:818–822. [PubMed: 18471995]
40. Neumann H, Peak-Chew SY, Chin JW. Genetically encoding *N*(epsilon)-acetyllysine in recombinant proteins. *Nature Chem. Biol.* 2008; 4:232–234. [PubMed: 18278036]
41. Hancock SM, Uprety R, Deiters A, Chin JW. Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *J. Am. Chem. Soc.* 2010; 132:14819–14824. [PubMed: 20925334]
42. Greiss S, Chin JW. Expanding the genetic code of an animal. *J. Am. Chem. Soc.* 2011; 133:14196–14199. [PubMed: 21819153]
43. Polcarpo CR, et al. Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* 2006; 580:6695–6700. [PubMed: 17126325]
44. Li X, Fekner T, Ottesen JJ, Chan MK. A pyrrolysine analogue for site-specific protein ubiquitination. *Angew. Chem. Int. Ed.* 2009; 48:9184–9187.
45. Nguyen DP, Garcia Alai MM, Kapadnis PB, Neumann H, Chin JW. Genetically encoding *N*(epsilon)-methyl-l-lysine in recombinant histones. *J. Am. Chem. Soc.* 2009; 131:14194–14195. [PubMed: 19772323]
46. Gautier A, et al. Genetically encoded photocontrol of protein localization in mammalian cells. *J. Am. Chem. Soc.* 2010; 132:4086–4088. [PubMed: 20218600]

47. Wijinen JW, Zavarise S, Engberts JBFN, Cahrton MI. J. Substituent effects on an inverse electron demand hetero Diels–Alder reaction in aqueous solution and organic solvents: cycloaddition of substituted styrenes to di(2-pyridyl)-1,2,4,5-tetrazine. *J. Org. Chem.* 1996; 61:2001–2005.
48. Devaraj NK, Hilderbrand S, Upadhyay R, Mazitschek R, Weissleder R. Bioorthogonal turn-on probes for imaging small molecules inside living cells. *Angew. Chem. Int. Ed.* 2010; 49:2869–2872.



**Figure 1. Scheme to label proteins via an inverse electron-demand Diels–Alder cycloaddition, and structural formulae of relevant compounds**

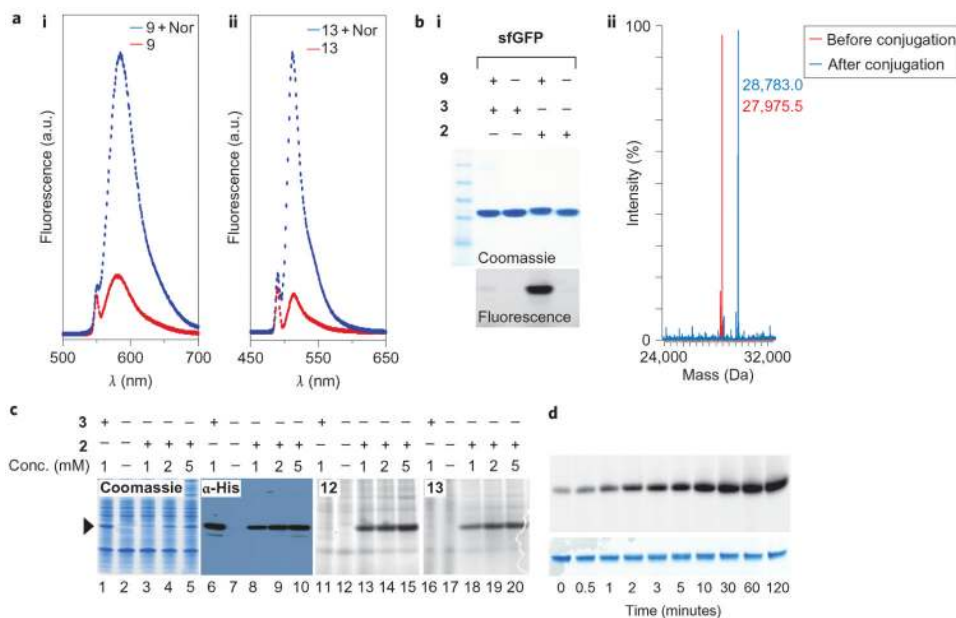
**a.** Genetically encoded norbornenes react rapidly with tetrazines, bearing probes (red star), in aqueous solution at ambient temperature and pressure to site-specifically label proteins. **b.** Amino acid structures of pyrrolysine (**1**), *N* $\epsilon$ -5-norbornene-2-ylloxycarbonyl-*L*-lysine (**2**), *N* $\epsilon$ -tert-butyloxycarbonyl-*L*-lysine (**3**) and *N* $\epsilon$ -2-azidoethyloxycarbonyl-*L*-lysine (**4**). **c.** Structures (**5–14**) of tetrazines and tetrazine–fluorophores used in this study. TAMRA-X, BODIPY TMR-X and BODIPY-FL are common names for fluorophores: their structural formulae are shown in Supplementary Fig. S4. Red boxes denote parent tetrazines. r.t. = room temperature.



**Figure 2. Efficient, genetically-encoded incorporation of 2 using the PylRS/tRNA<sub>CUA</sub> pair in *E. coli***

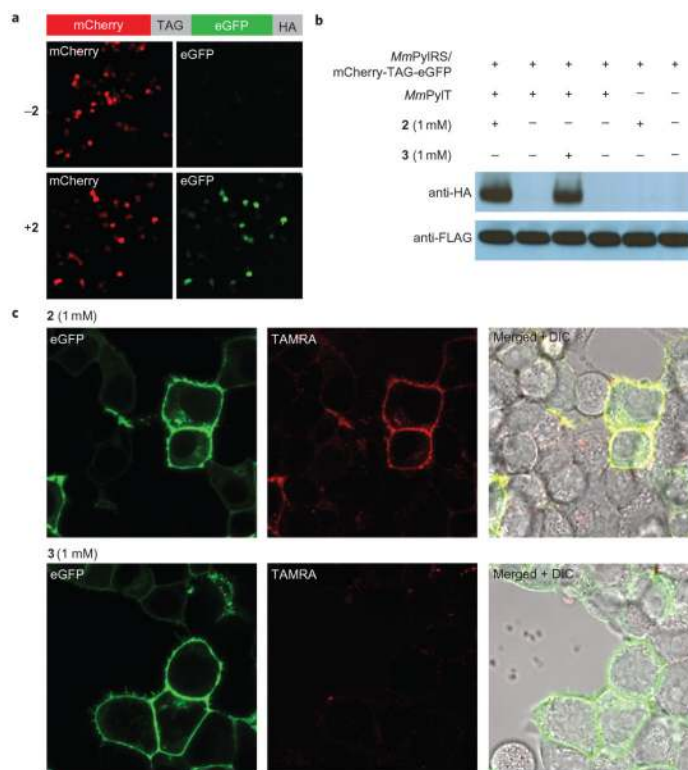
**a**, Amino acid dependent expression of sfGFP that bears an amber codon at position 150 and myoglobin that bears an amber codon at position 4. **b**, Mass spectrometry characterization of amino acid incorporation. **i**, sfGFP-2-His<sub>6</sub>, found: 27,975.5±1.5 Da, calculated: 27,977.5 Da. **ii**, Myo-2-His<sub>6</sub>, found: 18,532.5±1.5 Da, calculated: 18,532.2 Da.





### Figure 3. Characterization of tetrazine-norbornene reactions

**a**, 'Turn-on' fluorescence of tetrazine fluorophores **9** (**i**) and **13** (**ii**) on reaction with 5-norbornene-2-ol (Nor). **b**, Specific and quantitative labelling of sfGFP that bears **2**, demonstrated by SDS-PAGE (Coomassie staining and in-gel fluorescence) (**i**) and mass spectrometry (**ii**) before bioconjugation (red spectrum, found  $27,975.5 \pm 1.5$  Da, expected  $27,977.5$  Da) and after bioconjugation (blue spectrum, found  $28,783.0 \pm 1.5$  Da, expected  $28,784.4$  Da). **c**, Specificity of labelling **2** in sfGFP versus the *E. coli* proteome. Lanes 1–5: Coomassie-stained gel showing proteins from *E. coli* producing sfGFP in the presence of the indicated concentration of unnatural amino acids **2** or **3**. Lanes 6–10: The expressed protein was detected in lysates using an anti-His<sub>6</sub> antibody. Lanes 11–20: fluorescence images of protein labelled with the indicated fluorophore **12** or **13**. **d**, Labelling of myoglobin that bears **2** at position 4 with fluorophore **12**: fluorescence imaging (top) and Coomassie-stained loading control (bottom).



**Figure 4. Site-specific incorporation of 2 into proteins in mammalian cells and the specific labelling of EGFR-GFP on the cell surface with 9**

**a**, Cells that contain the PylRS/tRNA<sub>CUA</sub> pair and the mCherry(TAG)eGFP-HA reporter produced GFP only in the presence of **2**. **b**, Western blots confirm that the expression of full length mCherry(TAG)eGFP-HA is dependent on the presence of **2**. **c**, Specific and rapid labelling of a cell surface protein in live mammalian cells. EGFR-GFP that bears **2** or **3** at position 128 is visible as green fluorescence at the membrane of transfected cells (left panels). Treatment of cells with **9** (200 nM) leads to selective labelling of EGFR that contains **2** (middle panels). Right panels show merged green and red fluorescence images, DIC=differential interference contrast. Cells were imaged four hours after the addition of **9**.