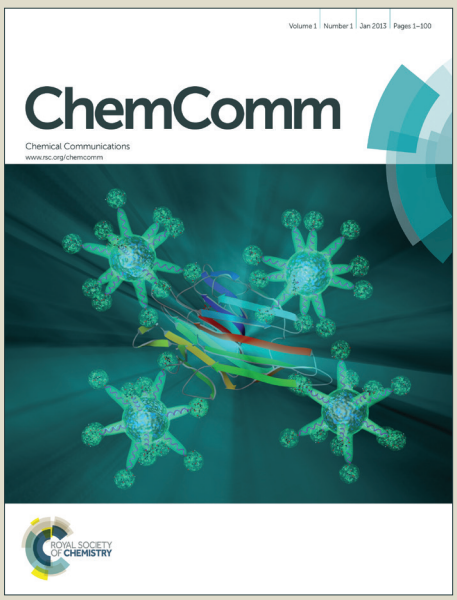
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Chemical Labeling of Intracellular Proteins via Affinity Conjugation and Strain-promoted Cycloadditions in Live Cells

Xi Chen,[a,b] Fu Li,[a,b] and Yao-Wen Wu*[a,b]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX

DOI: 10.1039/b000000x

A versatile chemical labeling approach was developed, where intracellular proteins were first incorporated with a bioorthogonal group via affinity conjugation, and subsequently labeled via strain-promoted cyclodaddition reactions in live cells.

Studying protein function *in vitro* or in live cells and organisms is essential in biological researches. Genetic tags, such as fluorescent proteins (FPs), are widely employed to detect proteins in cells. Yet, FPs show many limitations compared to organic dyes, including lack of environmental sensitivity, limited flexibility of modification and spectral range, and less amenability for temporal control.[1-3] In contrast, chemical probes are able to achieve properties that are not possible when using FPs, such as single-molecule imaging, fluorophore-assisted light inactivation, real-time detection of protein synthesis, and multicolour pulse-chase labeling.[1] In addition, many organic dyes are superior to FPs in terms of brightness, photo-stability, far red emission, environmental sensitivity, and the potential for modification of their spectra and biochemical properties. Moreover, chemical label is able to go beyond fluorescent modalities.[1, 4]

Chemical tagging approaches have been used to specifically label proteins with organic dyes or other probes.[5, 6] In these approaches, a protein of interest (POI) is fused with a protein domain or a polypeptide sequence, which is specifically modified with chemical probes. To date, various chemical tags have been developed, such as tetracysteine motif[7], His-tag[8], eDHFR[9], SNAP/CLIP,[10,11] SrtA,[12] HaloTag,[13] LAP tag,[14] PYP tag[15] and Spy tag.[16]

The eDHFR tag was introduced by Cornish's lab.[9] The antibiotics trimethoprim (TMP) has a high affinity (K_I = 1 nM) for *Escherichia coli* dihydrofolate reductase (eDHFR), but a much lower affinity (K_I = 4-8 μM) for mammalian DHFR.[17] Besides, eDHFR is small (18 kD, two thirds of GFP), monomeric and marginally perturbs the function of proteins that it fuses with.

We recently developed an affinity conjugation approach using engineered eDHFR tag and rationally designed a green emissive covalent TMP-AcBOPDIPY probe for rapid and fluorogenic labeling of proteins in live cells.[18] While this is an elegant design, the probe still suffers from limitation in the flexibility of varying the probe (Figure 1a). As a result, new probes have to be rationally designed, synthesized and tested individually to introduce different biochemical or biophysical properties. This could be laborious and may not be always successful.

Compared to this straightforward chemical labeling approach, a "tagging-then-labeling" approach could be more versatile. Proteins can be first incorporated with a bioorthogonal tag and be subsequently labeled via bioorthogonal reactions. In this study, a POI fused with an engineered eDHFR tag was introduced with an azide group via affinity conjugation (the "tagging" step). Then, the protein was labeled by a fluorescent probe *via* strain-promoted alkyne-azide cycloadditions (SPAAC),[19] facilitating biological investigations (Figure 1b,c).

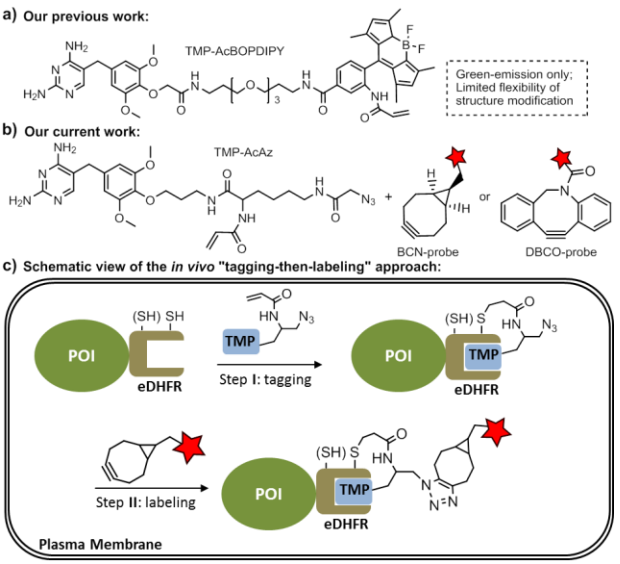
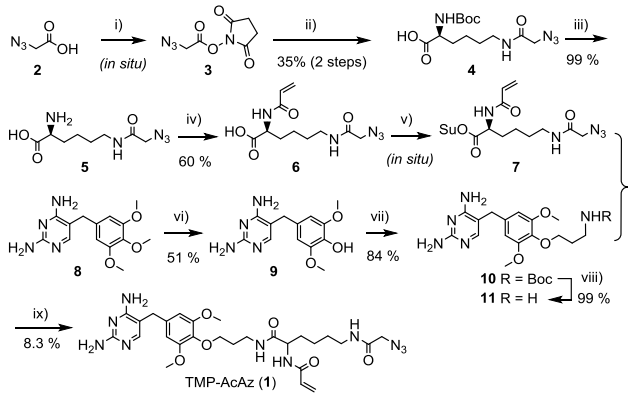


Figure 1. a, b) Comparison of our current work with the previous work. c) Schematic view the *in vivo* "tagging-then-labeling" concept: A POI fused with an appropriate eDHFR mutant was first treated with TMP-AcAz (**1**) ligand to introduce an azido group (Step I: tagging); subsequently, BCN-probe or DBCO-probe was applied, allowing the labeling of the POI in live cells (Step II: labeling) which facilitates biological investigations, e.g., intracellular FRET study.

We first designed and synthesized TMP-AcAz (**1**), which features TMP ligand for specific binding to eDHFR, an acrylamide group for covalent conjugation to cysteine on eDHFR and an azido group for bioorthogonal labeling. Due to proximity induced reactivity, the mildly reactive acrylamide group can only selectively attach to the cysteine residue introduced in the vicinity of the ligand binding site on eDHFR without interacting with other free thiols in live cells.[18]

We employed a convergent synthetic route to prepare TMP-



Scheme 1. Synthesis of TMP-AcAz (**1**): i) EDC, NHS, DMF, 2 h; ii) Boc-Lys-OH, DIEA, DMF, RT, overnight; iii) TFA/DCM (1:2, v:v), RT, 20 min; iv) acryloyl chloride, Na_2CO_3, H_2O; v) EDC, NHS, DMF, 3 h; vi) 48 % HBr, 100 °C, 20 min; vii) *N*-Boc 3-bromopropylamine, NaI, Cs_2CO_3, DMF, RT, 2h; viii) TFA/DCM, 20min; ix) DIEA, PBS, 2 h.

AcAz (**1**). 2-Azidoacetic acid (**2**) was *in situ* activated as an amine-reactive *N*-hydroxysuccinimidyl (NHS) ester (**3**) and then was coupled with *N*-Boc-Lys-OH to give the azido-lysine intermediate (**4**) which was deprotected to give *ε*-azidoacetyl-*L*-lysine (**5**). After acylation by acryloyl chloride, *N*-acryloyl-*ε*-azidoacetyl-*L*-lysine (**6**) was obtained as the first key building block which was then *in situ* activated as an NHS ester (**7**). In parallel, TMP (**8**) was regioselectively demethylated to give dimethoprim (**9**) and then coupled with *N*-Boc-3-bromopropylamine to give TMP-Pr-NHBoc (**10**). After deprotection, TMP-Pr-NH_2 (**11**) was obtained as the second key building block. Finally, TMP-Pr-NH_2 (**11**) was coupled with **7** at neutral condition in PBS buffer, giving the final product TMP-AcAz (**1**). Herein, the coupling between **7** and **11** is critical and needs to be performed in neutral PBS buffer. On the other hand, we prepared BCN- probes for strain-promoted cycloaddition reaction with azide (Supporting Information).

In this "tagging-then-labeling" approach, the first step is the incorporation of a bioorthogonal azido group to the engineered eDHFR tag in live cells. The tagging reaction rate is critical because a slow tagging rate will not only result in a prolonged reaction time, but also an insufficient incorporation of azide, which would strongly disfavour the subsequent labeling step. Therefore, we first sought to identify the eDHFR mutant with the highest reaction rate with TMP-AcAz. We prepared eight eDHFR mutants with single or double cysteines near the ligand binding pocket, including eDHFR_A19C, N23C, L28C, G51C, R52C, N23C/G51C, G51C/R52C and N23C/L28C (Figure 2b, Figure S1).[18, 22] We reasoned that the introduction of two cysteines may increase the probability of conjugation, resulting in faster reaction rate.

Since gel-electrophoresis and HPLC are difficult to monitor the reaction progress due to slight change of molecular weight (M.W.) and physical properties after conjugation, we used ESI-MS to monitor the reaction progress. Reaction was quenched by two volumes of quenching buffer (20% glycerol + 0.3% TFA). After reaction an increase of Δ *m/z* ~599 is observed during labeling which matches well with the M.W. of TMP-AcAz, 599.7 (Figure S1). We found that eDHFR_L28C displayed ca. 60-70 % conversion after 10 min (Figure S1a, S2) and complete conversion after 30 min (Figure S1b). In contrast, all other single mutants displayed only minor conversion after 10 min (Figure S1c-e). No detectable conjugation was observed for eDHFR_A19C after 10 min (Figure S1f). The reaction showed minor conversion after 3 h (Figure S1g). In the control, no reaction was detected in wild type (WT) eDHFR even after 48 h of reaction (Figure S1h). The reactivity of single cysteine mutants can be drawn in the following order: L28C > N23C ≈ G51C > R52C >> A19C. However, double cysteine mutants did not show significantly higher reaction rate than the single mutants (Figure S1i, j, k). Their reactivity can be drawn as: N23C/L28C > G51C/R52C > N23C/G51C. We also found that the addition of eDHFR cofactor, nicotinamide adenine dinucleotide phosphate (NADPH, 50 μM), led to further enhancement of the reaction rate (Figure 2Sl).

Next, we studied the reaction kinetics of eDHFR_N23C/L28C and eDHFR_L28C that showed the fastest reaction rates among the mutants. In order to monitor the conjugation progress, we employed a two-step method. eDHFR mutants (50 μM) were treated with TMP-AcAz (100 μM) in the presence of dithioerythritol (DTE, 100 μM) for different time intervals. The reaction mixture was treated with three volumes of the quenching/labeling buffer containing 8 M urea, 25 mM TMP, 2.67 mM *N*-(2-aminoethyl)maleimide and 267 μM Dibenzocyclooctyne (DBCO)-TAMRA at room temperature (RT)

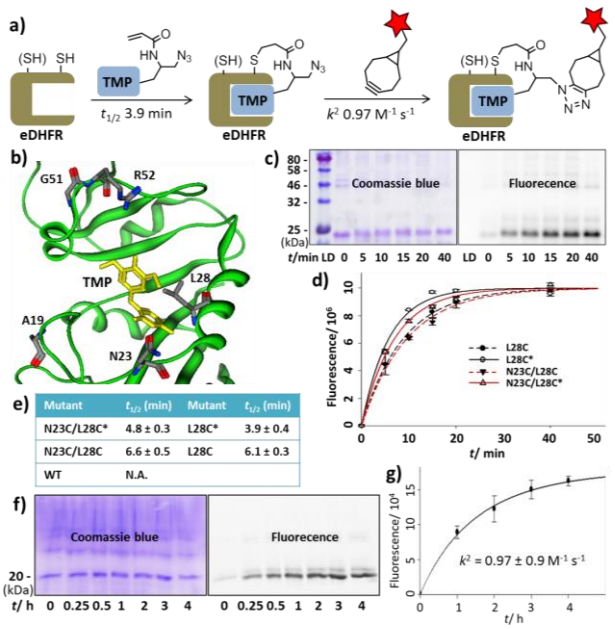


Figure 2. *In vitro* evaluation of the tagging and click labeling kinetics. a) Schematic view of the introduction of an azido group and then labeling by BCN-TAMRA. b) The crystal structure of eDHFR in complex with TMP where A19, N23, L28, G51 and R52 are highlighted. This crystal structure is obtained by aligning TMP into the crystal structure of eDHFR-DFA complex (PDB: 1RF7) and then energy minimized using MOE. c) Representative (eDHFR_L28C) coomassie blue stained and in-gel fluorescence images. This reveals the time-dependent conjugation of eDHFR_L28C with TMP-AcAz, which can be quantified by fluorescence intensity. d) Time-course of tagging reactions of eDHFR_L28C and N23C/L28C with (*) or without adding 50 μM NADPH. The lines show single exponential fitting. e) Tagging reaction half-life (* reaction in the presence of 50 μM NADPH). f) Representative coomassie blue stained and in-gel fluorescence image for time-dependent labeling by BCN-TAMRA. g) The labeling progress is fitted to the equation describing the 2nd-order reaction, giving a second order reaction constant of $k^2 = 0.97 \pm 0.09$ M^{-1} s^{-1}.

overnight to allow complete labeling (Figure S3). Afterwards, the conjugation reaction could be assayed by denaturing SDS-PAGE via in-gel fluorescence (Figure S4). This two-step approach enabled us to follow the conjugation progress quantitatively (Figure 2c, d). eDHFR_L28C and eDHFR_N23C/L28C showed a reaction half-life of 3.9 min and 4.8 min (Figure 2e), respectively, in the presentence of 50 µM NADPH, which represents the general concentration of NADPH in live cells.[20, 21] The tagging reaction of eDHFR_L28C by TMP-AcAz is only 2-fold slower than the labeling by TMP-AcBOPDIPY but 2-fold faster than the labeling by A-TMP probes.[18, 22] To our knowledge, the tagging reaction is faster than many other protein tagging reactions.[4, 22, 23]

Next, we studied the reaction kinetics of the second "labeling" step. We prepared a BCN-TAMRA probe which features a BCN (bicyclononyne) group for SPAAC reaction with azide, a PEG linker to increase aqueous solubility and a red emissive tetramethylrhodamine (TAMRA) fluorophore (Scheme S2).

eDHFR_L28C-Az was prepared by reacting of TMP-AcAz with eDHFR_L28C quantitatively. Afterwards, eDHFR_L28C-Az (50 µM) was treated with an excess of BCN-TAMRA (200 µM) in PBS buffer at 37 °C. At different time intervals, a reaction aliquot was taken and quenched by 100 mM 2-azidoacetic acid and immediately frozen in liquid nitrogen. Finally, the reaction mixture was subjected to SDS-PAGE and analysed by fluorescent scan and coomassie blue staining. A time-dependent incorporation of fluorophore into the protein was observed (Figure 2f). The time course of fluorescence incorporation was fitted to a second order reaction model, giving a kinetic constant of 0.97 ± 0.9 M^{-1} s^{-1}. This reaction is faster than previously reported reaction of SPAAC reaction with BCN (0.1-0.3 M^{-1} s^{-1}),[24] and is comparable with the fastest SPAAC reaction using biarylazacyclooctynone (BARAC).[25] This fast labeling rate will be favourable in the intracellular protein labeling in terms of reduction of reaction time and decrease of the amount of BCN-TAMRA probe.

After *in vitro* biochemical evaluation, we tested labeling of intracellular proteins in live cells using these reagents. HeLa cells were transiently expressed EGFP-eDHFR_N23C/L28C-NLS, containing the nuclear localization sequence (NLS), $(DPKKKRKV)_3$, which features three repetitive SV40 large T antigen nuclear localization signal (PKKKRKV) connected by aspartate (D).[26] Cells were treated with 50 µM TMP-AcAz, followed by labelling with 50 µM BCN-TAMRA. As can be seen from Figure 3a, specific labeling of the nucleus was observed, which co-localizes well with the EGFP channel. In contrast, no labeling was observed in cells expressing EGFP-eDHFR_N23C/L28C-NLS that were treated with BCN-TAMRA without tagging by TMP-AcAz (Figure 3a). Additionally, no specific labeling was detected in cells without transfection but treated with TMP-AcAz and BCN-TAMRA (Figure S5). In order to exclude any artifacts arised from EGFP, the experiments were performed in cells expressing eDHFR_N23C/L28C-NLS without fusion with EGFP. The same results were obtained (Figure 3b, S6a). To test the labeling at the plasma membrane, HeLa cells were transiently expressed EGFP-eDHFR_N23C/L28C-CAAX with K-Ras C-terminal sequence (KMSKDGKKKKKKSKTKCVIM), which target proteins to the plasma membrane. Specific labeling of intracellular proteins at the plasma membrane was achieved via the same procedure using TMP-AcAz/BCN-TAMRA (Figure S7). Inspired by these results, we advanced to label a target protein, centromere protein A (CENPA). CENPA (ca. 16 kDa) is a histone H3-like variant which serves as an epigenetic mark for centromere location and is required for recruitment and assembly of other centromere components and kinetochore proteins during mitosis. Pulse-chase labeling using chemical labeling approach made it possible to study CENPA function during mitotic progression.[27] We expressed EGFP-eDHFR_L28C-CENPA in live Hela cells and labeled it with 15 µM BCN-TAMRA. (Figure 3b, S6b).

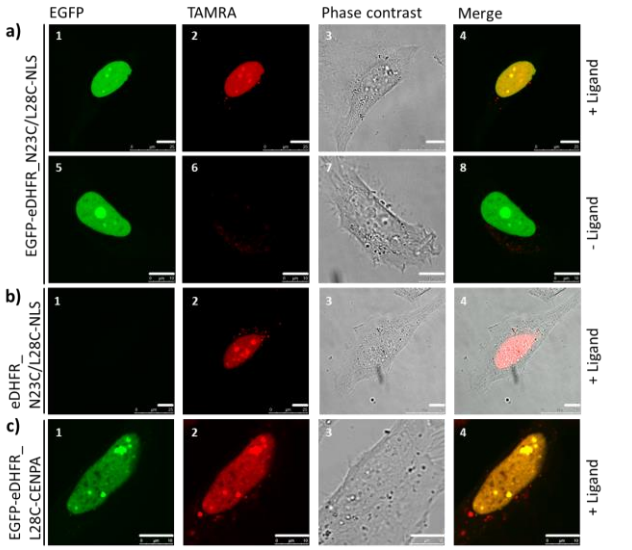


Figure 3. Confocal fluorescence microscopy images. Hela cells transiently expressing EGFP-eDHFR_N23C/L28C-NLS (a), eDHFR_N23C/L28C-NLS (b) or EGFP-eDHFR_L28C-CENPA (c) were treated by BCN-TAMRA. Specific labeling of proteins at nucleus was only observed for cells tagged by TMP-AcAz ligand. Scale bar: 10 µm.

In order to demonstrate the versatility of the labeling approach, we showed that the labeling can also be achieved in live cells using DBCO-TAMRA probe (Figure S8). ATTO647N is a superior fluorescent dye characterized by excellent photostability and brightness well suited for single molecule studies.[28] Hence, we prepared BCN-ATTO647N probe (Scheme S3) and used it to

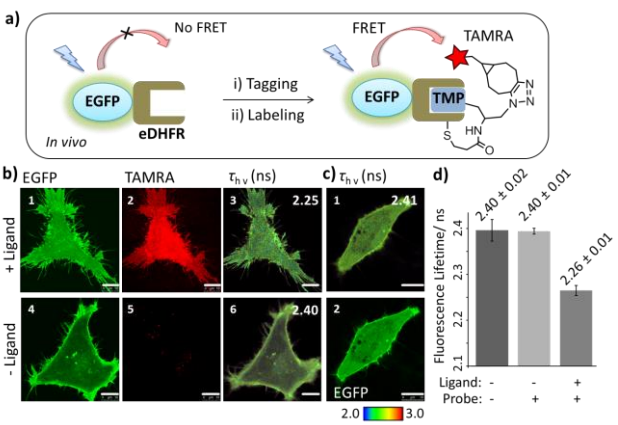


Figure 4. Intracellular FRET measurement. a) Schematic view of the intracellular FRET measurement. b) Confocal fluorescence microscopy and FLIM images of EGFP-eDHFR_L28C-CAAX with (top) or without (bottom) adding TMP-AcAz ligand before labeling by BCN-TAMRA. c) FLIM (#1) and confocal (#2) images of EGFP-eDHFR_L28C-CAAX. d) The fluorescence lifetime of EGFP. Scale bar: 10 µm.

label EGFP-eDHFR_N23C/L28C-NLS in live cells (Figure S9). These results showed that the introduced azido group is generally ready for SPAAC reactions in live cells.

After successful demonstration of the "tagging-then-labeling" approach in live cells, we proceeded to prove its utility in visualizing intracellular processes (Figure 4a). Fluorescence resonance energy transfer (FRET) is highly useful in cellular biology to study protein-protein interactions in cells. Fluorescence-lifetime imaging microscopy (FLIM) is an imaging technique based on the measurement of the lifetime of the donor. Energy transfer from the donor molecule to the acceptor molecule will lead to a decrease in the fluorescence lifetime of the former, which can be recorded by FLIM. Since FLIM measurements are insensitive to the concentration of fluorophores and can thus filter out artifacts resulted from changes in the concentration and emission intensity, FLIM is a very useful technique for FRET measurement to monitor interaction of biomolecules in cells.

Here, we measured the FRET between EGFP and the TAMRA fluorophore. In order to gain higher signal at the plasma membrane, the FLIM measurements were performed at the bottom of the cell. Hela cells expressing EGFP-eDHFR_L28C-CAAX were treated with BCN-TAMRA without TMP-AcAz, no specific labeling at plasma membrane was observed (Figure 4b, 4-5). We recorded the average fluorescence lifetime of 2.40 ± 0.01 ns for EGFP (Figure 4d) with a typical FLIM image given in Figure 4b (#6). The result is consistent with τ_{av} = 2.40 ± 0.2 ns (Figure 4d) of cells expressing EGFP-eDHFR_L28C-CAAX without any treatment. A typical FLIM image is given in Figure 4c (#1) with its confocal image (Figure 4c, #2). In contrast, in cells treated with TMP-AcAz/BCN-TAMRA, a significant reduced EGFP fluorescence lifetime of 2.26 ± 0.01 ns (Figure 4d) was observed, suggesting energy transfer from EGFP to TAMRA. A typical FLIM image is given in Figure 4b (#3). These results demonstrated that the "tagging-then-labeling" approach can be very useful for FRET studies in live cells (Figure S10).

Conclusions

In conclusion, we have developed a versatile "tagging-then-labeling" approach for chemical labeling of proteins in live cells. Both the tagging and labeling reactions are fast. TMP-AcAz tagged engineered eDHFR proteins with a half-life of ca. 4 minute via affinity conjugation. The azido-tagged protein can undergo strain-promoted cycloadditions with various cyclooctyne probes, including BCN- and DBCO-TAMRA and BCN-ATTO647N. We demonstrated that the application of the approach for specific labeling of intracellular proteins and FRET studies in live cells. We expect that the "tagging-then-labeling" approach can be widely used for chemical protein labeling in cells to study protein trafficking, protein-protein interactions, proteomics and to elucidate various cellular processes.

ACKNOWLEDGMENT

We thank Philippe Bastiaens for the support in FLIM measurements. We thank James C. Hu for the kind gift of the *E. coli.* DHFR plasmid. This work was supported in part by DFG grants (grant No.: SPP 1623 and SFB 642 to Y.W.W.).

Notes and references

[a] *Chemical Genomics Centre of the Max Planck Society, Otto-Hahn-Str. 15, 44227 Dortmund, Germany. E-mail: yaowen.wu@mpi-dortmund.mpg.de*

[b] *Max-Planck Institute for Molecular Physiology, Otto-Hahn-Str. 11, 44227 Dortmund, Germany.*

† Electronic Supplementary Information (ESI) available: organic synthesis, supporting figures and others. See DOI: 10.1039/b000000x/

1. B. N. Giepmans, S. R. Adams, M. H. Ellisman and R. Y. Tsien, *Science*, 2006, **312**, 217-224.
2. H. M. O'Hare, K. Johnsson and A. Gautier, *Curr. Opin. Struc. Biol.*, 2007, **17**, 488-494.
3. J. A. Prescher and C. R. Bertozzi, *Nat. Chem. Biol.*, 2005, **1**, 13-21.
4. Y. W. Wu and R. S. Goody, *J. Pept. Sci*., 2010, **16**, 514-523.
5. R. Wombacher and V. W. Cornish, *J. Biophotonics*, 2011, **4**, 391-402.
6. K. M. Marks and G. P. Nolan, *Nat. Methods*, 2006, **3**, 591-596.
7. B. A. Griffin, S. R. Adams and R. Y. Tsien, *Science*, 1998, **281**, 269-272.
8. E. G. Guignet, R. Hovius and H. Vogel, *Nat. Biotechnol.*, 2004, **22**, 440-444.
9. L. W. Miller, Y. Cai, M. P. Sheetz and V. W. Cornish, *Nat. Methods*, 2005, **2**, 255-257.
10. A. Keppler, H. Pick, C. Arrivoli, H. Vogel and K. Johnsson, *P. Natl. Acad. Sci. USA*, 2004, **101**, 9955-9959.
11. A. Gautier, E. Nakata, G. Lukinavicius, K. T. Tan and K. Johnsson, *J. Am. Chem. Soc*., 2009, **131**, 17954-17962.
12. M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner and H. L. Ploegh, *Nat. Chem. Biol.*, 2007, **3**, 707-708.
13. G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit and K. V. Wood, *ACS Chem. Biol.*, 2008, **3**, 373-382.
14. M. Fernandez-Suarez, H. Baruah, L. Martinez-Hernandez, K. T. Xie, J. M. Baskin, C. R. Bertozzi and A. Y. Ting, *Nat. Biotechnol.*, 2007, **25**, 1483-1487.
15. Y. Hori, H. Ueno, S. Mizukami and K. Kikuchi, *J. Am. Chem. Soc.*, 2009, **131**, 16610-16611.
16. B. Zakeri, J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy and M. Howarth, *P. Natl. Acad. Sci. USA*, 2012, **109**, E690-E697.
17. L. F. Kuyper, B. Roth, D. P. Baccanari, R. Ferone, C. R. Beddell, J. N. Champness, D. K. Stammers, J. G. Dann, F. E. Norrington, D. J. Baker and P. J. Goodford, *J. Med. Chem.*, 1982, **25**, 1120-1122.
18. W. Liu, F. Li, X. Chen, J. Hou, L. Yi and Y. W. Wu, *J. Am. Chem. Soc.*, 2014, **136**, 4468-4471.
19. N. J. Agard, J. A. Prescher, and C. R. Bertozzi, *J Am Chem Soc*, 2004 **126**, 15046-15047.
20. P. D. Reiss, P. F. Zuurendonk and R. L. Veech, *Anal. Biochem.*, 1984, **140**, 162-171.
21. R. L. Veech, L. V. Eggleston and H. A. Krebs, *Biochem. J.*, 1969, **115**, 609-619.
22. Z. Chen, C. Jing, S. S. Gallagher, M. P. Sheetz and V. W. Cornish, *J. Am. Chem.*, 2012, **134**, 13692-13699.
23. S. Uchinomiya, H. Nonaka, S. Wakayama, A. Ojida and I. Hamachi, *Chem. Commun.*, 2013, **49**, 5022-5024.
24. J. Dommerholt, S. Schmidt, R. Temming, L. J. Hendriks, F. P. Rutjes, J. C. van Hest, D. J. Lefeber, P. Friedl and F. L. van Delft, *Angew. Chem.*, 2010, **49**, 9422-9425.
25. C. G. Gordon, J. L. Mackey, J. C. Jewett, E. M. Sletten, K. N. Houk and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2012, **134**, 9199-9208.
26. S. Kosugi, M. Hasebe, N. Matsumura, H. Takashima, E. Miyamoto-Sato, M. Tomita and H. Yanagawa, *J. Biol. Chem.*, 2009, **284**, 478-485.
27. S. J. Falk, L. Y. Guo, N. Sekulic, E. M. Smoak, T. Mani, G. A. Logsdon, K. Gupta, L. E. Jansen, G. D. Van Duyne, S. A. Vinogradov, M. A. Lampson, and B. E. Black, *Science*, 2015**, 348**, 699-703.
28. D. Normanno, L. Boudarene, C. Dugast-Darzacq, J. Chen, C. Richter, F. Proux, O. Benichou, R. Voituriez, X. Darzacq, M. Dahan, *Nat. Commun.* 2015, **6**, 7357.