ELSEVIER

Selective chemical labeling of proteins in living cells

Lawrence W Miller and Virginia W Cornish

Labeling proteins with fluorophores, affinity labels or other chemically or optically active species is immensely useful for studying protein function in living cells or tissue. The use of genetically encoded green fluorescent protein and its variants has been particularly valuable in this regard. In an effort to increase the diversity of available protein labels, various efforts to append small molecules to selected proteins *in vivo* have been reported. This review discusses recent advances in selective, *in vivo* protein labeling based on small molecule ligand–receptor interactions, intein-mediated processes, and enzyme-catalyzed protein modifications.

Addresses
Department of Chemistry, Columbia University, Havemeyer Hall, MC 3153, 3000 Broadway, New York, New York 10027, USA

Corresponding author: Miller, LW (lwm2001@columbia.edu)

Current Opinion in Chemical Biology 2005, **9**:56–61

This review comes from a themed issue on
Proteomics and genomics
Edited by Benjamin F Cravatt and Thomas Kodadek

Available online 8th January 2005

1367-5931/$ – see front matter

DOI 10.1016/j.cbpa.2004.12.007

Abbreviations

ACP	acyl carrier protein
CoA	coenzyme A
DHFR	dihydrofolate reductase
EDT	1,2-ethanedithiol
FlAsH	4′,5′-bis(1,3,2-dithioarsolan-2-yl)fluorescein
GFP	green fluorescent protein
hAGT	human O^6-alkylguanine-DNA alkyltransferase
Mtx	methotrexate
NTA	nitrilotriacetate
PPTase	phosphopantetheine transferase
PTD	transduction domain
SLF′	synthetic ligation factor

Introduction

Elucidating the distribution, dynamics and chemical environment of proteins inside living cells is critical for understanding the biomolecular mechanisms of cellular function. Labeling of proteins with fluorescent probes or affinity reagents has facilitated *in vitro* studies of protein structure, dynamics and protein–protein interactions [1]. However, traditional methods of protein labeling are often inadequate for *in vivo* studies because they require purification of the protein, chemical labeling, repurification and reintroduction into cells by invasive methods such as microinjection. These limitations have spawned efforts to non-invasively and site-specifically label proteins in living cells or tissue.

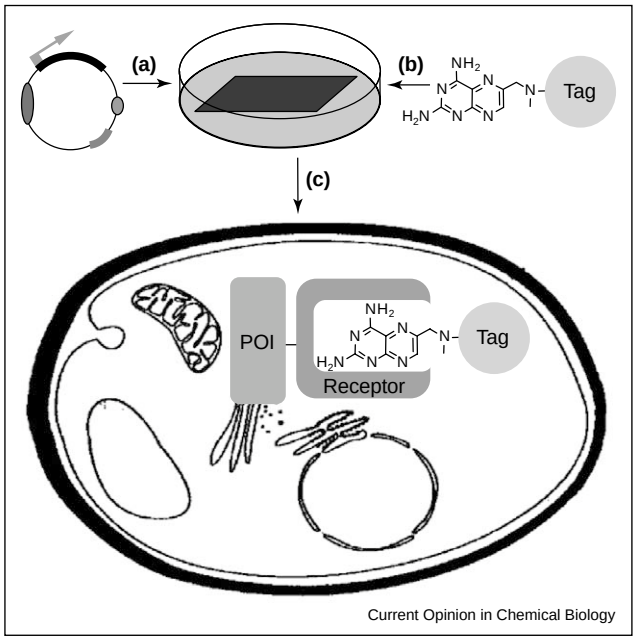
The most prominent method of protein labeling is to genetically encode green fluorescent protein (GFP) or one of its variants as a fusion to the protein of interest [2••,3]. The resulting gene fusion is expressed, and the autofluorescent GFP fusion is detected microscopically. The relatively small size (ca. 27 kDa) of GFP and its compact, single-domain structure allow it to be fused to a wide variety of target proteins with little or no interference in native protein functionality. Although GFP variants have proven to be extremely useful for *in vivo* studies of protein function, their utility is inherently limited because their spectral and structural characteristics are interdependent. Whereas mutagenesis has led to the development of differently colored GFPs, including cyan, green, yellow and blue variants, and a red-emitting protein has been cloned from *Discosoma* [2••], it has been difficult to engineer GFP variants with well-resolved absorption and emission spectra and to obtain a well-behaved red variant. To increase the diversity of protein labels, approaches are needed that combine the ability to genetically encode the label as for GFP with the flexibility of small-molecule labels.

The need for chemically diverse protein labels has led researchers to develop ways to label fusion proteins with small molecule probes. The general strategy of *in vivo*, site-specific protein labeling entails genetically fusing a target protein to a receptor protein, protein-domain or peptide sequence (see Figure 1). The small-molecule probe consists of a receptor-binding ligand coupled to a fluorophore or other functional moiety. When added to cells growing in culture, the probe enters the cell and binds specifically and stably to the receptor fusion. The success of this strategy depends on identifying or developing a receptor that is specific for the small molecule and that doesn't interfere with the function of the target protein. The small-molecule probe needs to be cell-permeable and non-toxic. The approaches reviewed herein adhere to this general strategy — exploiting small-molecule probes that bind receptor fusions via spontaneous non-covalent or covalent interaction, intein-mediated rearrangement and subsequent small-molecule attachment, or enzyme-catalyzed covalent labeling.

Direct chemical labeling of receptor domains with small molecules

Most of the approaches to *in vivo* chemical labeling of proteins exploit a specific, high-affinity non-covalent or

Figure 1



General strategy for site-specific chemical labeling of proteins *in vivo*. **(a)** Living cells are transfected with DNA encoding a protein of interest (POI) fused to a receptor domain. **(b)** Upon expression of the receptor fusion, a cell-permeable small molecule probe consisting of a ligand coupled to a detectable tag is added to cell growth medium. **(c)** Protein function is analyzed in the living cells via fluorescent microscopy or other detection methods.

covalent interaction between a synthetic ligand and its corresponding receptor. The ligand–receptor pairs include hapten–antibody, biotin–avidin, various enzyme–inhibitor combinations, nitrilotriacetate (NTA)-oligohistidine sequence, and biarsenical fluorophores that bind cysteine-rich peptide sequences. Each approach has strengths and limitations that we explore below.

One of the first reported ligand–receptor pairs for general *in vivo* protein labeling was reported by Farinas and Verkman, who labeled a single-chain antibody fusion with a fluorescein-conjugated hapten and optically measured the pH value of Golgi bodies in live cells [4]. A similar strategy to measure the pH values of organelles by exploiting the biotin–avidin interaction was reported by Wu *et al.* [5]. However, neither antibodies nor avidin make good receptors for general intracellular protein labeling. Farinas and Verkman reported that their antibody did not express well in cellular reducing environments, and avidin expresses as a 63-kDa tetramer, the large size of which is likely to interfere with target protein function.

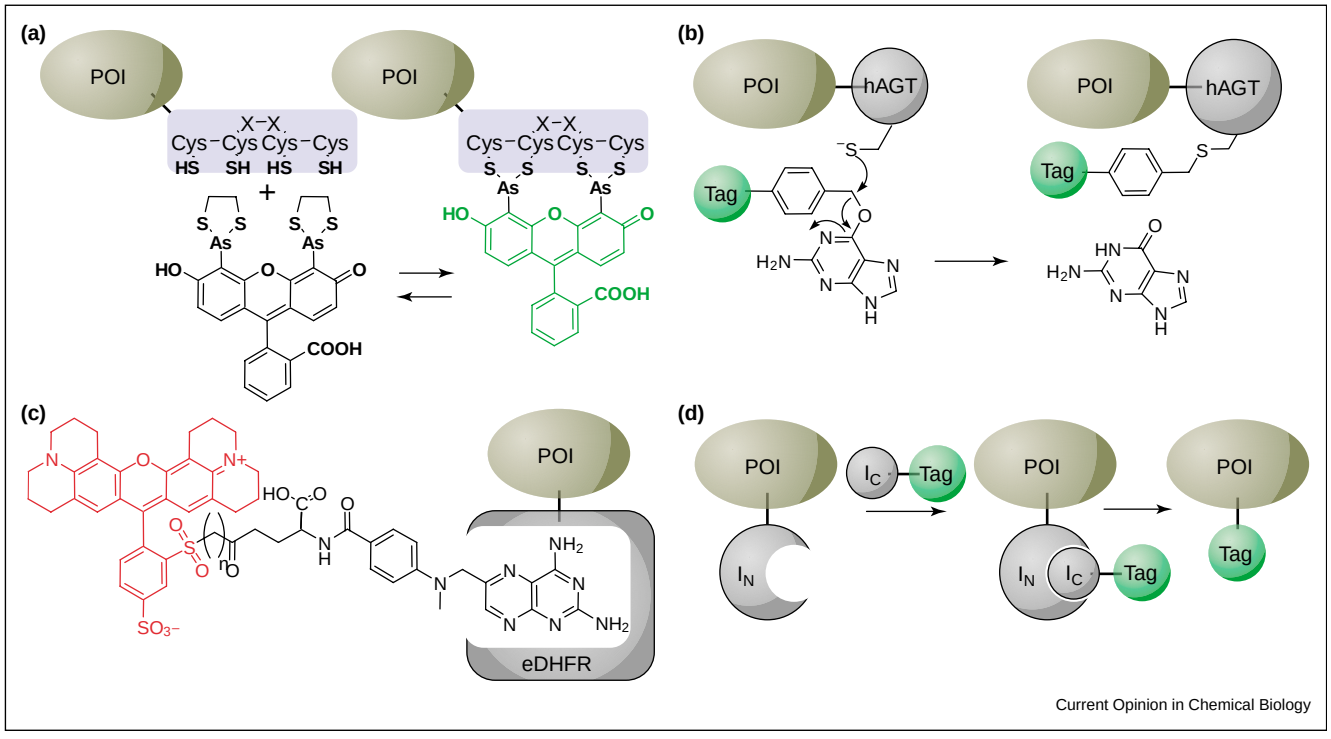
The prototypical system for the specific chemical labeling of proteins *in vivo* is the biarsencial ligand/tetracysteine motif interaction pair developed in Roger Tsien's laboratory (Figure 2a) [6,7,8•,9,10]. This approach relies upon the subnanomolar affinity between a short tetracysteine peptide (CCXXCC, where X is any amino acid except cysteine) and a biarsenical compound such as 4′,5′-bis(1,3,2-dithioarsolan-2-yl)fluorescein (FlAsH). In its application, a target protein of interest is expressed in living cells with the tetracysteine motif. The FlAsH dye is administered to cells in the presence of an excess of 1,2-ethanedithiol (EDT). EDT outcompetes endogenous proteins with closely spaced cysteine pairs, thus minimizing non-specific binding and toxicity. In addition to the green FlAsH biarsenical, blue (ChoXAsH), red (ReAsH) [8] and a biarsenical derivative of nile red have been synthesized [11].

The FlAsH-based protein labeling system has several advantages. First, the biarsenical derivatives show a dramatic increase in fluorescence upon binding to their target, minimizing background noise in labeling experiments. Second, the tetracysteine motif is sufficiently small that it can be fused not only to the N or C terminus of a protein, but it can also be incorporated into loops or on the outer surface of α helices with little chance of the tag interfering with target protein function. The stability of the biarsenical-tetracysteine motif interaction and the availability of different colors enables the consecutive labeling of fusion proteins in pulse-chase experiments [9]. A further advantage of the FlAsH system is that it allows the fluorescence detection to be confirmed by electron microscopy. ReAsH localized to the tetracysteine fusion protein photoconverts diaminobinzidine into an electron-rich precipitate [9]. Recently, it was demonstrated that FlAsH and ReAsH can be used as mediators of chromophore-assisted light inactivation (CALI) [10,12]. In one such example, Tour *et al.* labeled cells expressing a connexin43-tetracysteine fusion with ReAsH [10]. The electrical coupling of two cells was monitored by patch-clamp method. Within 30 s of exposure to high-intensity (17 W/cm^2) epi-illumination at the ReAsH absorption wavelength, the gap junctions between the patch-clamped cells were almost completely inactivated, probably due to the localized photogeneration of singlet oxygen.

The FlAsH system has several limitations, however. Even with the addition of excess EDT, the background fluorescence in biarsenical-labeled cells is high due to the non-specific labeling of cysteine-rich proteins [13]. Furthermore, the cysteines of the receptor tag must be in reduced form, making labeling of proteins in cellular oxidizing environments difficult. Finally, biarsenicals are not modular, in the sense that the tetracysteine-binding moiety is coupled directly to the fluorophore. This makes the development of a diverse collection of biarsenical probes a synthetically challenging proposition.

Since the development of FlAsH, a variety of other ligand–receptor strategies for *in vivo* protein labeling have been

Figure 2



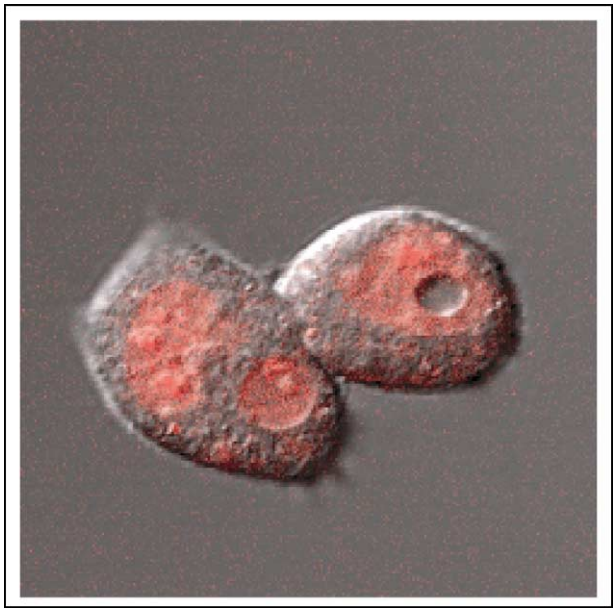
Various methods of chemically labeling fusion proteins *in vivo*. **(a)** Fluorescent biarsenical protein labeling (FlAsH). A biarsenical fluorescein derivative binds to a short tetracysteine peptide fused to the protein of interest (POI). **(b)** Covalent labeling of human O^6-alkylguanine DNA alkyl transferase (hAGT) with benzyl guanine derivatives. **(c)** Schematic of methotrexate-Texas Red™ bound to an *E. coli* dihydrofolate reductase (eDHFR) fusion protein. **(d)** Split intein labeling. The N-terminal portion of a split intein is expressed as a fusion to the POI. The C-terminal half of the split intein, covalently labeled with a chemical tag, binds to the N-terminal half and self-splices, yielding the tagged POI.

reported. Kai Johnnson's laboratory has developed a strategy for covalently labeling fusion proteins *in vivo*. Fluorescently labeled O^6-benzylguanine derivatives irreversibly and specifically label human O^6-alkylguanine-DNA alkyltransferase (hAGT) fused to target proteins and expressed in mammalian cells (Figure 2b) [14••,15•,16•,17–19]. The authors first showed that a green benzyl guanine-alexafluor488™ probe rapidly (<5 min) permeated the membrane of hAGT-deficient Chinese hamster ovary (CHO) cells and specifically labeled an overexpressed hAGT fusion targeted to the nucleus [15•]. In subsequent work, various fluorescent benzyl guanine derivatives were used to label several hAGT fusions in different subcellular locations. The covalent interaction between benzyl guanine and hAGT enables the visualization of labeled proteins over the course of several hours — limited only by the probe-induced, intracellular degradation of the fusion protein [16•]. hAGT, at ca. 25 kDa, is comparable in size to GFP, and should thus be similarly non-interfering of target protein function. One significant drawback of this approach is that it can only be applied for labeling proteins in hAGT-deficient cell lines, as the benzyl guanine substrates react with endogenous AGT.

A variety of non-covalent ligand–receptor interactions has been leveraged for protein labeling. Guignet *et al.* exploited the NTA–oligohistidine interaction commonly used for protein purification to investigate the structure of the transmembrane $5HT_3$ serotonin receptor by fluorescence resonant energy transfer (FRET) [20]. The authors reported a dissociation constant of ca. 0.2 μM between NTA–rhodamine and His_{10}-labeled $5HT_3$. The relatively low affinity of NTA probes for His-labeled proteins probably makes stable labeling and imaging problematic. Marks *et al.* introduced a method of labeling proteins *in vivo* with fluorescein for the purpose of localized singlet oxygen photogeneration [21]. This method relies upon the high-affinity interaction (K_D = ca. 0.1 nM) between a mutant of FKBP12 and a derivative of synthetic ligation factor (SLF′) originally developed by Clackson *et al.* [22]. The authors reported light-induced inactivation of a β-galactosidase-FKBP12 fusion expressed in 3T3 fibroblast cells and bound to fluorescein-SLF′.

In the Cornish laboratory, we have developed a protein labeling strategy that relies on the non-covalent binding

Figure 3



Confocal micrograph of DHFR-deficient CHO cells expressing nucleus-localized eDHFR. An overlay of differential image contrast and fluorescence channels is shown (excitation at 568 nm). The cells were incubated in media containing 2 μM methotrexate-Texas Red™ for 16 h and washed with phosphate buffered saline before imaging.

of dihydrofolate reductase (DHFR), an 18 kDa monomeric enzyme and fluorescently labeled DHFR inhibitors (Figure 2c). A fluorescinated derivative of the DHFR inhibitor methotrexate (Mtx) has long been used to quantify intracellular DHFR levels [23]. Mtx-fluorescein was subsequently used to microscopically detect the subcellular location of fusions to the mammalian form of DHFR [24–26]. Miller *et al*. reported a general strategy of protein labeling wherein fusions of the *Escherichia coli* form of DHFR (eDHFR) were labeled with fluorescent Mtx. DHFR-deficient CHO cells transiently expressing fusions of eDHFR localized to the plasma membrane or nucleus could be effectively labeled by incubating the cells overnight in the presence of micromolar concentrations of Mtx-Texas Red™ (Figure 3) [27•].

Indirect approaches to chemically modifying proteins *in vivo*

Besides direct chemical labeling of receptor domains with small-molecule probes, various approaches that co-opt naturally occurring biochemical processes to modify proteins in living cells have been reported. One such approach relies on the specific incorporation of unnatural amino acids based on suppressor tRNA technology [28,29,30•]. If this method can be developed to the extent that a wide variety of complex chemical labels can be incorporated into proteins in living mammalian cells, it will prove to be enormously useful.

Some recent efforts have adapted protein splicing for labeling proteins with small molecules [31–33]. Protein splicing is a naturally occurring post-translational modification wherein a protein autocatalytically rearranges to excise an internal segment (an intein) and ligate the flanking N- and C-terminal sequences, or exteins [34]. The method of native or expressed protein ligation exploits intein-based splicing to yield purified proteins with C-terminal thioesters. The thioester is then reacted with derivatized cysteines to append peptides, affinity labels or fluorophores to the protein of interest [35].

Giriat and Muir successfully adopted intein splicing to selectively label proteins in living cells by expressing a target protein with the first half of the naturally occurring Ssp DnaE split intein fused to its C-terminus (Figure 2d) [31]. The second half of the split intein covalently linked to a small-molecule probe and a protein transduction domain (PTD) peptide is added to the cell growth media. Upon entering the cell, the PTD, linked to the probe-derivatized intein half via a disulfide bond, is released. The split intein halves combine and self-splice, leaving the protein of interest labeled at its C-terminus with the probe. One potential drawback of intein-based protein labeling is that the splicing kinetics are such that complete labeling may require time scales on the order of hours.

Yet another approach to labeling fusion proteins relies on specific third-party ligases to transfer a functional small molecule to an acceptor domain. Until recently, the only practical example of this method was the *in vivo* biotinylation of proteins fused to biotin acceptor domains by biotin ligases [36,37]. Two recent papers report a labeling strategy that uses a phosphopantetheine transferase (PPTase) to catalyze the transfer of 4′-phosphopantetheine from coenzyme A (CoA) to a serine residue of an acyl carrier protein (ACP) [38,39]. In one example, George *et al*. chose the ACP/PPTase pair from *E. coli* to label the exoplasmic N-terminus of the G-protein coupled receptor neurokinin-1 (NK_1). NK_1 expressed in HEK293 cells was labeled with Cy3, Cy5 or biotin-derivatized CoA [38]. While the ACP/PPTase/CoA strategy is promising for purification or cell-surface labeling applications, the highly-charged CoA derivates are likely to be relatively membrane-impermeant.

Conclusion

In vivo studies of protein function require greater spatial, temporal and compositional resolution than is currently available. To meet this need, a diverse set of chemical-based protein labeling technologies must be developed. The approaches reported to date point the way toward effective design criteria for site-specific protein labeling in living systems. Receptor moieties must be amenable to genetic encoding as fusions to the protein of interest. The receptors must be relatively small so as to not perturb

protein function — an ideal receptor would be a short peptide sequence that could be inserted into various locations within the protein. Chemical probes should bind receptors with high specificity and stability so as to enable functional studies over a time-scale of hours with no background noise. Probes should be designed in a modular fashion so that a wide variety of fluorophores, affinity labels or other functional moieties can be easily linked. The kinetics of cell loading and receptor binding should be fast enough (on the order of minutes) to facilitate the most time-sensitive biological assays. Finally, a variety of complementary probe–receptor pairs will be needed to enable the simultaneous study of multiple target proteins. Meeting these design criteria will provide exciting challenges for chemical biologists and protein chemists in the near future.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Hermanson GT: *Bioconjugate Techniques*. San Diego: Academic Press; 1996.

2. •• Zhang J, Campbell RE, Ting AY, Tsien RY: **Creating new fluorescent probes for cell biology**. *Nat Rev Mol Cell Biol* 2002, **3**:906-918.
The authors comprehensively review recent advances in the development of fluorescent proteins and their use in the study of protein function.

3. Tsien RY: **The green fluorescent protein**. *Annu Rev Biochem* 1998, **67**:509-544.

4. Farinas J, Verkman AS: **Receptor-mediated targeting of fluorescent probes in living cells**. *J Biol Chem* 1999, **274**:7603-7606.

5. Wu MM, Llopis J, Adams S, McCaffery JM, Kulomaa MS, Machen TE, Moore HP, Tsien RY: **Organelle pH studies using targeted avidin and fluorescein-biotin**. *Chem Biol* 2000, **7**:197-209.

6. Griffin BA, Adams SR, Jones J, Tsien RY: **Fluorescent labeling of recombinant proteins in living cells with FlAsH**. *Methods Enzymol* 2000, **327**:565-578.

7. Griffin BA, Adams SR, Tsien RY: **Specific covalent labeling of recombinant protein molecules inside live cells**. *Science* 1998, **281**:269-272.

8. • Adams SR, Campbell RE, Gross LA, Martin BR, Walkup GK, Yao Y, Llopis J, Tsien RY: **New biarsenical ligands and tetracysteine motifs for protein labeling *in vitro* and *in vivo*: synthesis and biological applications**. *J Am Chem Soc* 2002, **124**:6063-6076.
Synthesis and biological characterization of fluorescent biarsenical ligands for labeling tetracysteine peptide motifs.

9. Gaietta G, Deerinck TJ, Adams SR, Bouwer J, Tour O, Laird DW, Sosinsky GE, Tsien RY, Ellisman MH: **Multicolor and electron microscopic imaging of connexin trafficking**. *Science* 2002, **296**:503-507.

10. Tour O, Meijer RM, Zacharias DA, Adams SR, Tsien RY: **Genetically targeted chromophore-assisted light inactivation**. *Nat Biotechnol* 2003, **21**:1505-1508.

11. Nakanishi J, Nakajima T, Sato M, Ozawa T, Tohda K, Umezawa Y: **Imaging of conformational changes of proteins with a new environment-sensitive fluorescent probe designed for site-specific labeling of recombinant proteins in live cells**. *Anal Chem* 2001, **73**:2920-2928.

12. Marek KW, Davis GW: **Transgenically encoded protein photoinactivation (FlAsH-FALI): acute inactivation of synaptotagmin I**. *Neuron* 2002, **36**:805-813.

13. Stroffekova K, Proenza C, Beam KG: **The protein-labeling reagent FLASH-EDT2 binds not only to CCXXCC motifs but also non-specifically to endogenous cysteine-rich proteins**. *Pflugers Arch* 2001, **442**:859-866.

14. •• Johnsson N, Johnsson K: **A fusion of disciplines: chemical approaches to exploit fusion proteins for functional genomics**. *ChemBioChem* 2003, **4**:803-810.
An excellent review on the chemical modification of fusion proteins for the study of protein function in living systems.

15. • Keppler A, Gendreizig S, Gronemeyer T, Pick H, Vogel H, Johnsson K: **A general method for the covalent labeling of fusion proteins with small molecules *in vivo***. *Nat Biotechnol* 2003, **21**:86-89.
The authors introduce benzyl guanine derivatives for the covalent labeling of hAGT fusion proteins in living cells.

16. • Keppler A, Pick H, Arrivoli C, Vogel H, Johnsson K: **Labeling of fusion proteins with synthetic fluorophores in live cells**. *Proc Natl Acad Sci USA* 2004, **101**:9955-9959.
Further development and characterization of the benzyl guanine-hAGT protein labeling method.

17. Keppler A, Kindermann M, Gendreizig S, Pick H, Vogel H, Johnsson K: **Labeling of fusion proteins of O6-alkylguanine-DNA alkyltransferase with small molecules *in vivo* and *in vitro***. *Methods* 2004, **32**:437-444.

18. Kindermann M, Sielaff I, Johnsson K: **Synthesis and characterization of bifunctional probes for the specific labeling of fusion proteins**. *Bioorg Med Chem Lett* 2004, **14**:2725-2728.

19. Juillerat A, Gronemeyer T, Keppler A, Gendreizig S, Pick H, Vogel H, Johnsson K: **Directed evolution of O(6)-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules *in vivo***. *Chem Biol* 2003, **10**:313-317.

20. Guignet EG, Hovius R, Vogel H: **Reversible site-selective labeling of membrane proteins in live cells**. *Nat Biotechnol* 2004, **22**:440-444.

21. Marks KM, Braun PD, Nolan GP: **A general approach for chemical labeling and rapid, spatially controlled protein inactivation**. *Proc Natl Acad Sci USA* 2004, **101**:9982-9987.

22. Clackson T, Yang W, Rozamus LW, Hatada M, Amara JF, Rollins CT, Stevenson LF, Magari SR, Wood SA, Courage NL *et al.*: **Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity**. *Proc Natl Acad Sci USA* 1998, **95**:10437-10442.

23. Kaufman RJ, Bertino JR, Schimke RT: **Quantitation of dihydrofolate reductase in individual parental and methotrexate-resistant murine cells. Use of a fluorescence activated cell sorter**. *J Biol Chem* 1978, **253**:5852-5860.

24. Israel DI, Kaufman RJ: **Dexamethasone negatively regulates the activity of a chimeric dihydrofolate reductase/glucocorticoid receptor protein**. *Proc Natl Acad Sci USA* 1993, **90**:4290-4294.

25. Remy I, Michnick SW: **Clonal selection and *in vivo* quantitation of protein interactions with protein-fragment complementation assays**. *Proc Natl Acad Sci USA* 1999, **96**:5394-5399.

26. Remy I, Michnick SW: **Visualization of biochemical networks in living cells**. *Proc Natl Acad Sci USA* 2001, **98**:7678-7683.

27. • Miller LW, Sable J, Goelet P, Sheetz MP, Cornish VW: **Methotrexate Conjugates: A Molecular *in vivo* Protein Tag**. *Angew Chem Int Ed Engl* 2004, **43**:1672-1675.
The authors introduce methotrexate and *E. coli* DHFR as a generic ligand–receptor pair for *in vivo* protein labeling.

28. Wang L, Brock A, Herberich B, Schultz PG: **Expanding the genetic code of *Escherichia coli***. *Science* 2001, **292**:498-500.

29. Chin JW, Cropp TA, Anderson JC, Mukherji M, Zhang Z, Schultz PG: **An expanded eukaryotic genetic code**. *Science* 2003, **301**:964-967.

30. • Zhang Z, Alfonta L, Tian F, Bursulaya B, Uryu S, King DS, Schultz PG: **Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells**. *Proc Natl Acad Sci USA* 2004, **101**:8882-8887.

The first successful incorporation of an unnatural amino acid into a protein in mammalian cells using suppressor tRNA technology.

31. Giriat I, Muir TW: **Protein semi-synthesis in living cells**. *J Am Chem Soc* 2003, **125**:7180-7181.

32. Buskirk AR, Ong YC, Gartner ZJ, Liu DR: **Directed evolution of ligand dependence: small-molecule-activated protein splicing**. *Proc Natl Acad Sci USA* 2004, **101**:10505-10510.

33. Yeo DS, Srinivasan R, Uttamchandani M, Chen GY, Zhu Q, Yao SQ: **Cell-permeable small molecule probes for site-specific labeling of proteins**. *Chem Commun (Camb)* 2003:2870-2871.

34. Noren CJ, Wang J, Perler FB: **Dissecting the chemistry of protein splicing and its applications**. *Angew Chem Int Ed Engl* 2000, **39**:450-466.

35. Muir TW: **Semisynthesis of proteins by expressed protein ligation**. *Annu Rev Biochem* 2003, **72**:249-289.

36. Cull MG, Schatz PJ: **Biotinylation of proteins *in vivo* and *in vitro* using small peptide tags**. *Methods Enzymol* 2000, **326**:430-440.

37. Cronan JE Jr, Reed KE: **Biotinylation of proteins *in vivo*: a useful posttranslational modification for protein analysis**. *Methods Enzymol* 2000, **326**:440-458.

38. 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.

39. Yin J, Liu F, Li X, Walsh CT: **Labeling proteins with small molecules by site-specific posttranslational modification**. *J Am Chem Soc* 2004, **126**:7754-7755.