Review

Journal of PeptideScience

Received: 15 July 2010 Accepted: 20 July 2010 Published online in Wiley Online Library:

(wileyonlinelibrary.com) DOI 10.1002/psc.1287

Probing protein function by chemical modification‡

Yao-Wen Wu[a,b]* and Roger S. Goody[a]

Labeling proteins with synthetic probes, such as fluorophores, affinity tags, and other functional labels is enormously useful for characterizing protein function *in vitro*, in live cells, or in whole organisms. Recent advancements of chemical methods have substantially expanded the tools that are applicable to modify proteins. In this review, we discuss some important chemical methods for site-specific protein modification and highlight the application of established techniques to tackle biological questions.

Keywords: protein labeling; fluorescent proteins; chemical probes; native chemical ligation; expressed protein ligation; protein *trans*-splicing; chemoselective reactions; bioorthogonal reactions; *in vivo* chemical labeling; Rab GTPases; prenylation; post-translational modification

Introduction

Studying protein function *in vitro* or in the context of live cells and organisms is of vital importance in biological research. Genetic tags such as fluorescent proteins (FPs) are widely used to detect proteins. The 2008 Nobel Prize in Chemistry rewarded the discovery and use of GFP as a tagging tool in biological science. However, compared to chemical tags, FPs have several limitations. First, FPs are generally not sensitive to environmental parameters such as hydrophobicity, pH, and ion concentrations, because the chromophore is buried in a barrel structure [1]. Second, although different variants of FP are available through mutagenesis, they cannot compete with organic dyes in the flexibility of modification and spectral range [2]. Moreover, FPs have disadvantages in brightness and photostability and are therefore are not ideal for most single molecule studies. Third, FPs can only provide an optical read-out, whereas other detection modalities such as spin labels for electron paramagnetic resonance (EPR) spectroscopy have unique properties and can be used to monitor conformational change of proteins [3]. Fourth, genetic reporters are less amenable for temporal control. In living systems, the proteins are produced on the timescale of gene expression and mature gradually, presenting a disadvantage for monitoring rapid intracellular biochemical events. In contrast, protein activity can be precisely controlled by light using photocleavable chemical probes [4]. Finally, genetic tags cannot be used to label nucleic acids, glycans, lipids, and protein post-translational modifications. In contrast, chemical probes are able to achieve properties that are not readily possible when using fluorescent proteins, such as fluorophore-assisted light inactivation, real-time detection of protein synthesis, and multi-color pulse-chase labeling [5]. Many organic dyes are superior to fluorescent proteins in terms of brightness, photostability, far red-emission, environmental sensitivity, and potential for modifications to their spectral and biochemical properties. Moreover, chemical modification is widely used for protein immobilization on microarrays [6]. Altogether, chemical modification of proteins has become an important strategy for the study of protein function. The emerging chemical labeling techniques have significantly broadened the range of manipulating protein structures. Here, we summarize the advancement of chemical labeling methods and their application in biological research.

Native Chemical Ligation and Expressed Protein Ligation

Synthetic chemistry provides almost unlimited possibilities for modulation of the structure of a polypeptide chain in order to understand protein function. However, as proteins are large molecules, total chemical synthesis of proteins is a considerable challenge [7]. In the early 1990s, Kent and coworkers introduced the breakthrough approach of native chemical ligation (NCL), which is now a general method for chemical protein synthesis [8]. In this method, two unprotected synthetic peptide fragments are coupled together under neutral aqueous conditions with the formation of a native peptide bond at the ligation site.

The principle of NCL is depicted in Figure 1. The approach is based on the chemoselective reaction between a peptide containing a *C*-terminal thioester and another peptide containing an *N*-terminal cysteine. The initial chemoselective transthioesterification in NCL is essentially reversible, whereas the subsequent S → N acyl shift is spontaneous and irreversible. Thus, the reaction is driven to form an amide bond specifically at the ligation site, even in the presence of unprotected internal cysteine residues. NCL

* *Correspondence to: Yao-Wen Wu, Department of Physical Biochemistry, Max Planck Institute of Molecular Physiology, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany. E-mail: wu@mpi-dortmund.mpg.de*

a *Department of Physical Biochemistry, Max Planck Institute of Molecular Physiology, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany*

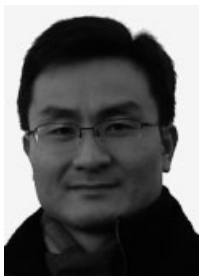
b *Cardiovascular Division and Randall Division for Cell and Molecular Biophysics, King's College London, London SE1 1UL, United Kingdom*

‡ *Special issue devoted to contributions presented at the Symposium 'Probing Protein Function through Chemistry', 20–23 September 2009, Ringberg Castle, Germany.*

Journal of PeptideScience

Biography

Roger Goody was born in 1944 in Northampton, England and studied at the University of Birmingham, England. He obtained his BSc in chemistry in 1965 and his PhD in 1968. After two postdoctoral positions, he completed his habilitation in biochemistry/biophysics at the University of Heidelberg, Germany and was nominated professor in 1990. Since 1993, he has been Director at the MPI of Molecular Physiology, Dortmund, Germany. In 2004, he became Professor of Biochemistry at the Ruhr-University, Bochum, Germany. His interests are in the area of phosphate-transferring proteins.



Yao-Wen Wu was born in 1980 in China. He received his BS in Chemistry from Sun Yat-sen University in 2001 and his MS in Organic Chemistry from Tsinghua University in 2004. After graduating as Dr rer. nat. (2008) at the Technical University of Dortmund working at the Max Planck Institute of Molecular Physiology in Dortmund and a postdoctoral period in cell biology at King's College London, he was appointed leader of an Otto Hahn group at the Max Planck Institute in Dortmund. His research interests are in chemical biology and biochemistry of membrane trafficking.

permits connecting two or more synthetic peptides to introduce functional tags into proteins. A number of refinements and extension in ligation methodology and strategy have been developed, such as extension of ligation site other than cysteine, auxiliary group-facilitated ligation [9,10], catalytic thiol cofactors [11,12], and kinetically controlled ligation (KCL) [13,14] (for a recent review see Ref. 15).

The scope of application of NCL was significantly widened upon introduction of the approach referred to as expressed protein ligation (EPL) from the Muir laboratory [16]. In this method, both fragments containing *C*-terminal thioester and *N*-terminal cysteine, respectively, can be produced recombinantly. Expressed protein ligation (EPL) emerged as a result of the advances in self-cleavable affinity tags for recombinant protein purification using intein chemistry [17]. Inteins are protein insertion sequences flanked by host protein sequences (*N*- and *C*-exteins) and are eventually removed by a post-translational process termed protein splicing [18]. Mutated inteins containing a *C*-terminal Asn to Ala substitution have been designed to keep their ability in the initial N $\rightarrow$ S acyl shift without further going through later steps of protein splicing [19]. Therefore, proteins fused to the *N*-termini of these engineered inteins can be cleaved by thiol reagents (such as 2-mercaptoethanesulfonate, MESNA) via an intermolecular transthioesterification reaction, releasing the α-thioester-tagged proteins (Figure 2).

The recombinant polypeptide α-thioesters can then be ligated with a synthetic peptide or recombinant protein containing *N*-terminal cysteine. EPL thereby allows site-specific introduction of probes (such as fluorophores and isotopes), post-translational modifications (such as prenylation, glycosylation, phosphorylation, and ubiquitination), incorporation of unnatural amino acids, and immobilization of proteins or peptides onto a chip (reviewed in Ref. 20). Moreover, sequential EPL strategies make it possible to introduce modification at any position in the protein sequence [21,22]. Therefore, EPL provides a platform that allows the application of powerful synthetic chemistry tools to proteins.

Recent examples of using EPL include the preparation of fluorescent prenylated Rab GTPases, key proteins involved in organizing intracellular vesicular transport. Rab GTPases are post-translationally modified by (usually) two geranylgeranyl groups at their *C*-terminus, which enables them to associate with membranes. Earlier, there were insurmountable difficulties in recombinant preparation of prenylated proteins and significant challenges in obtaining prenylated Rab GTPases in defined nucleotide-bound states. Therefore, it was technically difficult to analyze the interaction between GDP/GTP-bound prenylated Rab and its regulators Rab escort protein (REP) and GDP dissociation inhibitor (GDI). The EPL technique allowed the production of fluorescent prenylated Rab proteins by ligating synthetic prenylated peptides containing a fluorophore on the lipid or on the side chains of amino acids with recombinant Rab thioesters [23,24] (Figure 3). Dansyl and NBD fluorophores were introduced as reporters. This approach also enabled precise installation of GDP/GTP(or analog GppNHp) into Rab proteins to generate the 'off' and 'on' states, yielding homogeneous preparations of functionalized prenylated proteins in well-defined nucleotide-bound states, which was previously unfeasible. Such semi-synthetic Rab protein probes displayed significant changes in fluorescence intensity upon binding to REP and GDI, therefore enabling quantitative analysis of the interactions. These analyses led to the establishment of a thermodynamic model of Rab membrane recycling and a novel model of guanine nucleotide exchange factors (GEFs)-mediated Rab membrane targeting [25,26].

Inteins can also be split into two parts that function only when they interact with each other [27,28]. The precursor fragments can be fused to parts of a split intein, so that when these two pieces interact with each other, the resulting intein activity can mediate a *trans*-splicing reaction (Figure 4). The discovery of the naturally occurring split *Synechocystis (Ssp)* DnaE intein makes it possible to bypass denaturation and renaturation of the isolated precursor fragments. Protein *trans*-splicing is a useful approach to selectively ligate two polypeptides, thereby providing a valuable tool for protein engineering [29,30]. In particular, the split *Ssp* DnaB Int^N fragment with a length as short as 11 amino acids was capable of protein *trans*-splicing, which significantly facilitates the preparation of the modified *N*-terminal fragment by solid-phase peptide synthesis (SPPS) [29,31]. Protein *trans*-splicing has been employed to generate head-to-tail cyclic peptides and proteins [32,33], to selectively incorporate isotopes for NMR analysis [34,35], to fluorescently label proteins *in vitro* or in live cells [36,37], to assay protein–protein interaction *in vivo* [38–41], to detect biological events *in vivo* [42–44], and to selectively immobilize proteins onto a chip [45].

Chemoselective Chemistry

In general, the genetic code specifies 20 standard amino acids for proteins. Some of the residues are amenable for chemoselective modification under physiological conditions, for example,

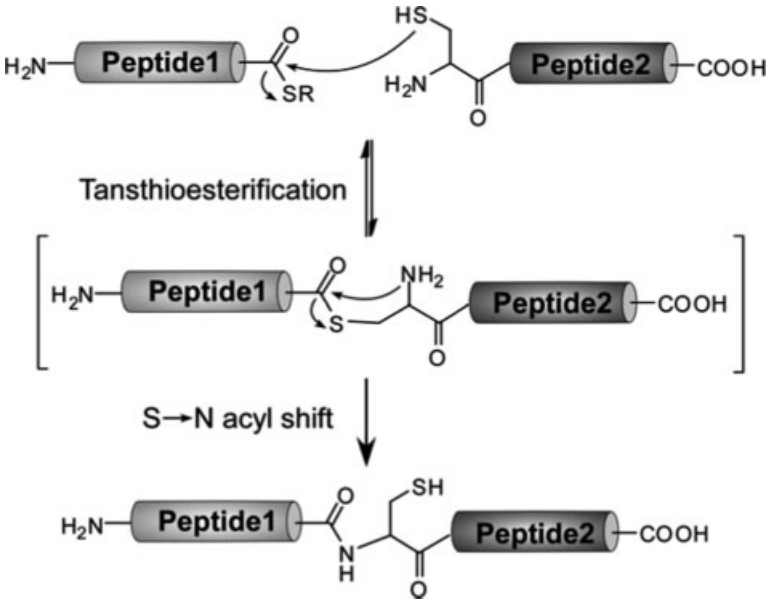


Figure 1. Principle of native chemical ligation.

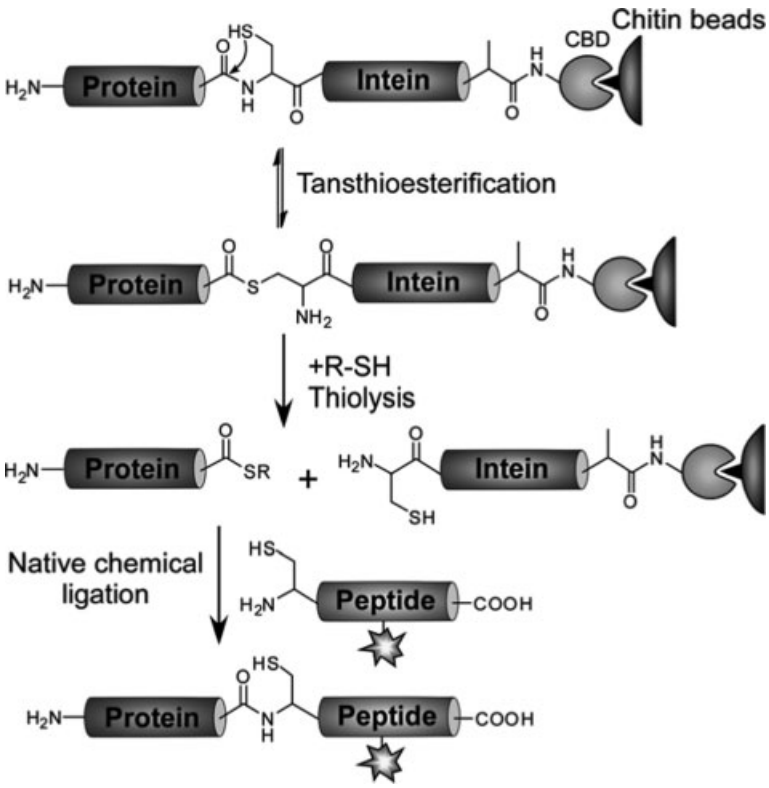


Figure 2. Principle of expressed protein ligation.

the sulfhydryl reagents iodoacetamide and maleimide modify cysteine, and succinimidyl esters react with lysine side chains or *N*-terminal primary amines. Such residue-specific reactions are widely used in tissue immunostaining via the conjugation of organic dyes to antibodies. Recent progress in bioconjugation techniques has expanded the range of modification residues to include tryptophan and tyrosine [46–49]. However, these residue-specific bioconjugation approaches are not selective when the object of study is biological systems instead of an individual protein.

Bioorthogonal chemistry allows specific, covalent attachment of chemical tags to biomolecules, without displaying cross-reactivity with other biomolecules *in vivo*. The term 'bioorthogonality' was introduced by Bertozzi's Lab when they first developed Staudinger ligation for selective labeling of glycans at the cell surface [50]. There are several advantages for bioorthogonal chemistry: (a) it is versatile and applicable to all kinds of biomolecules; (b) it can be used to label targets of interest in living cells and organisms; (c) the functional groups are small enough to

Journal of PeptideScience

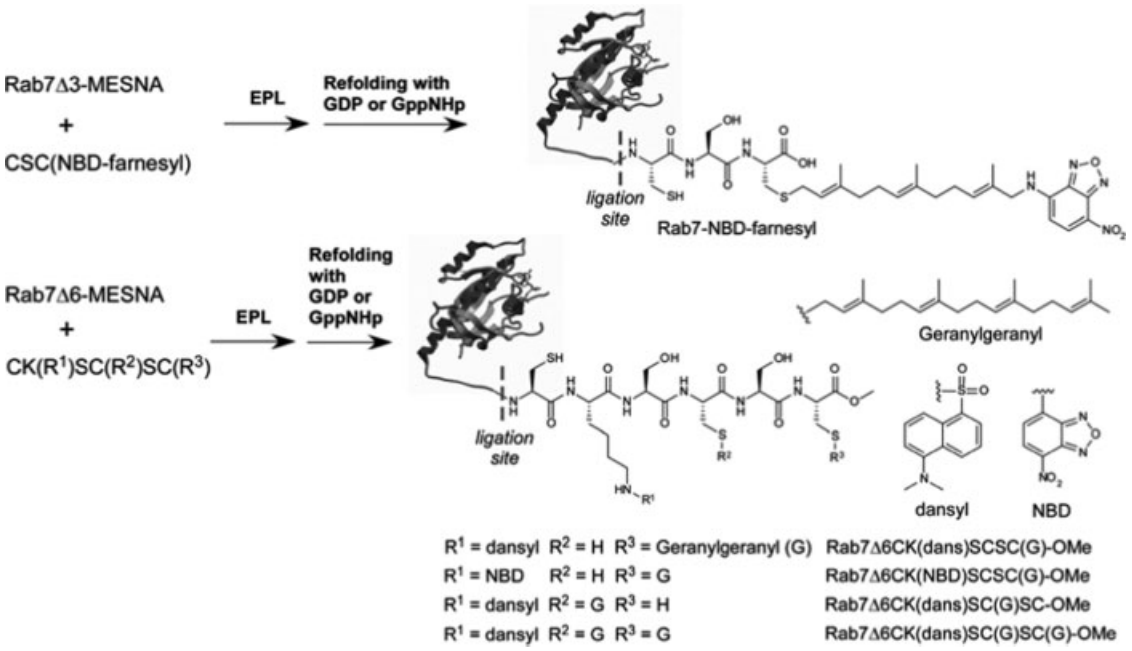


Figure 3. Construction of prenylated Rab proteins with a fluorophore on the lipid or on the side chain of amino acid.

Figure 4. Protein *trans*-splicing to ligate two fragments. The split intein *N*-terminal part (I_N) is fused to a tag and the split intein *C*-terminal part (I_C) is fused to the *C*-terminal protein fragment. The two pieces of the intein associate to reconstitute splicing activity.

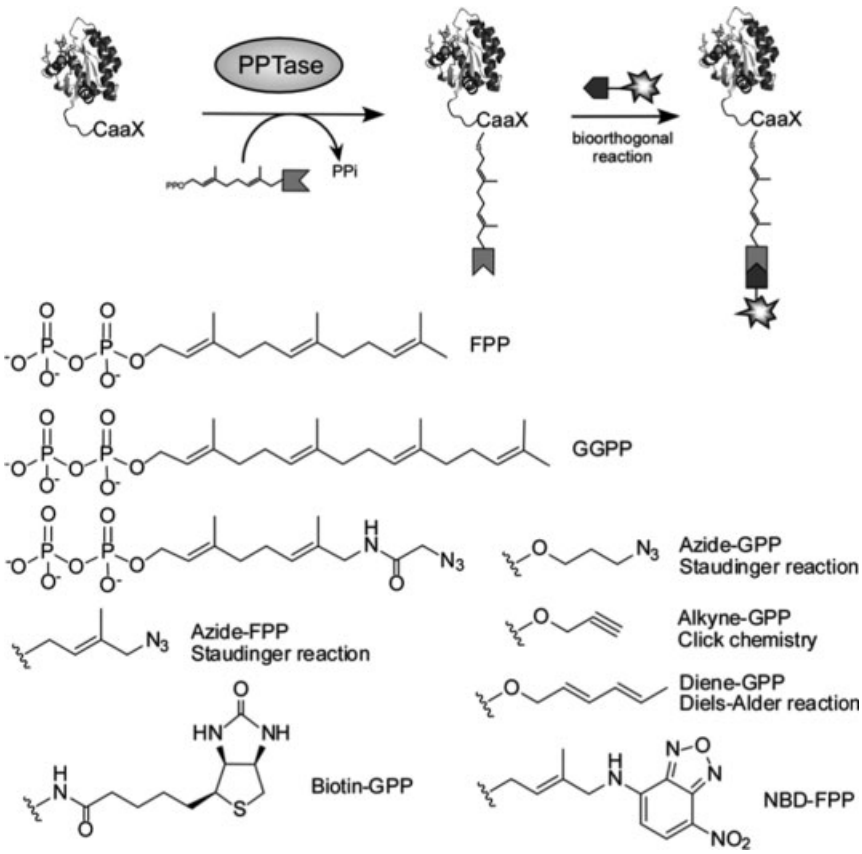
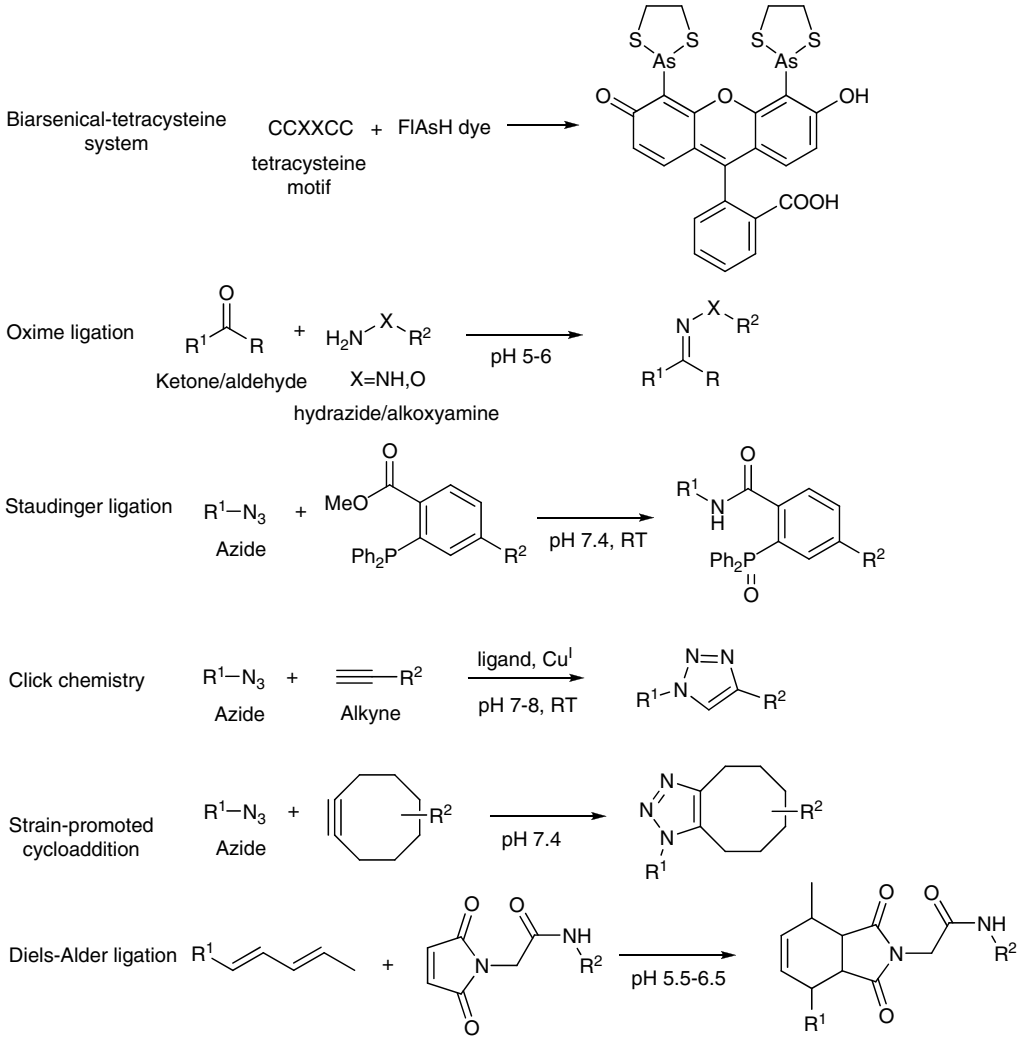


Figure 5. Labeling proteins through prenylation using modified isoprenoids.



Scheme 1. Bioorthogonal reactions: biarsenical-tetracysteine system [66], oxime ligation [67], Staudinger ligation [50], click chemistry [68], strain-promoted cycloaddition [69], and Diels–Alder ligation [70].

be tolerated by biosynthetic enzymes and could therefore be incorporated into metabolites or substrates in biological systems. Therefore, bioorthogonal chemistry has been successfully used in genome-wide protein profiling, including proteomic analysis of glycoproteins and glycosidases [51–53] and activity-based protein profiling (ABPP) for the identification of hydrolases, proteases, kinases, phosphotases, and glycosidases [54,55].

Widely used bioorthogonal reactions are listed in Scheme 1. Recent progress in bioorthogonal chemistry has been reviewed elsewhere [56,57]. Applying bioorthogonal chemistry for protein labeling involves the introduction of one of the functional pairs into a biomolecule and subsequent selective attachment of tags. This can be achieved by using the cellular biosynthetic machinery to incorporate unnatural amino acids [58] and modified monosaccharides [59], by nonsense suppression methodology [60,61], by enzyme-catalyzed modifications [62–64] (See Section on Chemical Labeling in Live Cells), and by semi-synthetic approaches as described above. A recent example includes the intein-mediated incorporation of an oxyamine moiety into the *C*-termini of proteins, which are amenable for subsequent conjugation with a fluorophore containing a ketone functionality [65].

One of the strategies for introducing functional groups to proteins includes using post-translational prenylation. Protein prenyltransferases (PPTases) are responsible for coupling soluble farnesyl pyrophosphate (FPP) and geranylgeranyl pyrophosphate (GGPP) moieties to the *C*-terminal cysteine(s) of proteins, leading to the formation of a stable thioether bond. Farnesyltransferase (FTase) and protein geranylgeranyltransferase type I (GGTase-I) recognize the *C*-terminal CaaX motif (C is cysteine, a is usually, but not necessarily, an aliphatic amino acid, and X can be one of the varieties of amino acids) in their substrates. Interestingly, PPTases can tolerate diverse modifications of their lipid substrates. Therefore, bioorthogonal groups or probes can be incorporated into proteins containing a CaaX tag via prenylation [71–73] (Figure 5). This strategy has been employed for proteomic analysis of prenylated substrates in cells [71,74], for identification of PPTases inhibitors [75,76], for protein immobilization [77,78], and for selective protein labeling [79,80].

Chemical Labeling in Live Cells

Biological systems are composed of interacting networks of biomolecules. Observation of many biologically relevant

Journal of PeptideScience

Table 1. Methods for chemical labeling in live cells

Tag (size in amino acids)	Small molecule probe	Enzyme required	Cellular labeling	Comments	References
Fluorogenic biarsenical and bisboronic reagents					
CCXXCC (6) tetracysteine	Biarsenical: FlAsH, ReAsH	No	Intracellular	Fluorogenic label; Nonspecific reactions and interactions; Cytotoxic in some cases.	66
SSXXSS (6) tetraserine	Bisboronic RhoBo	No	Cell surface	Fluorogenic Low; Cytotoxic; Off-target labeling of endogenous tetraserine motif.	86
Metal chelation					
6H (6)	Ni-NTA	No	Cell surface	Ni^{2+} is toxic to cells and quenches fluorescence.	87–89
6H (6)	Zn^{2+} complex HisZiFit	No	Cell surface	Noncytotoxic.	90
DDDD (4)	Zn^{2+} complex	No	Cell surface	Spectroscopic change upon chelation.	91–94
Enzymatic modification					
ACP (77)/A1 (12) or PCP(80)/ybbR (11) tag	Coenzyme A (CoA) derivatives	Phosphopantetheinyl transferases (AcpS or Sfp)	Cell surface	AcpS or Sfp catalyzes the attachment of a CoA-activated probe to the serine residue of the peptide tag. $k_{cat} = 0.03–0.3\ s^{-1}$, $k_{cat}/K_m = 0.25–4.2 \times 10^4\ M^{-1}\ s^{-1}$	95–98
AP tag (15)	Biotin ketone analog and subsequent oxime ligation	Biotin ligase (BirA)	Cell surface	Ketobiotin functions as a substrate for BirA and is incorporated into the lysine residue of the peptide tag; Minimal labeling time 20 min.	63
Q-tag (7)	Cadaverine derivatives	Transglutaminase (TGase)	Cell surface	TGase mediates the formation of amide bond between the amine of cadaverine and the glutamate residue of the peptide tag.	62,99
LAP tag (13–22)	Azido lipoic acids and subsequent strain-promoted cycloaddition, or coumarin	Lipoic acid ligase (LplA) and LplA mutant	Cell surface and intracellular	$k_{cat} = 0.048\ s^{-1}$ for the conjugation of azido lipoic acid and $k = 4.3 \times 10^{-3}\ M^{-1}\ s^{-1}$ for cycloaddtion; Coumarin addition by ^{W37V}LplA: $k_{cat} = 0.019\ s^{-1}$, $K_m = 50\ \mu M$; ^{W37I}LplA: $k_{cat} = 0.016\ s^{-1}$, $K_m = 261\ \mu M$.	64,100
CXPXR (6)	Oxyamine or hydrazide probes	Formylglycine-generating enzyme (FGE)	Cell surface	FGE catalyzes the transformation of a Cys of the peptide tag to a formylglycine.	62,101
LPXTG (5)	Polyglycine peptide probes	Sortase A (Srt A)	Cell surface	SrtA cleaves the peptide at the Thr-Gly site, and attaches a labeled polyglycine peptide; No limitation of the size of modification introduced.	102
Self-labeling enzymatic domains					
SNAP, CLIP tags (182)	Benzylguanine derivatives	Mutated O^6-alkylguanine-DNA alkyltransferase (hAGT)	Intracellular and cell surface	hAGT is alkylated through transferring the labeled benzyl group from O^6-benzylguanosine derivatives. $k_{obs} = 0.1–2.8 \times 10^4\ M^{-1}\ s^{-1}$	103,104
Halo tag (296)	Alkyl chloride probes	Mutated haloalkane dehalogenase (DhaA)	Intracellular and cell surface	The mutated DhaA is covalently modified by alkyl chloride probes. $k_{obs} = 2.7–8.5 \times 10^6\ M^{-1}\ s^{-1}$	105
Ligand-binding domains					
eDHFR (159)	Trimethylprim (TMP) conjugates	No	Intracellular	Tight and specific binding between *Escherichia coli* dihyrofolate reductase (eDHFR) and TMP ligand; Efficient labeling in cells within 10 min; Reversible labeling.	106
FKBP12 (108)	SLF′ conjugates	No	Intracellular	Tight and specific binding between FKBP12 mutant and SLF′ ligand.	107

events cannot be accomplished by merely studying individual biomolecules, but requires investigation in a functioning biological system. Thus, tremendous efforts have been made to visualize biological processes in live cells and whole organisms, usually requiring tracking of biomolecules in such environments. In addition to fluorescent proteins, chemical tagging methods have been shown to be valuable for labeling biomolecules in live cells and organisms [5,59,81,82]. For example, pulse-labeling of a protein of interest in cells can be easily achieved by applying a small molecule probe into the cell medium and subsequent wash-out. The newly synthesized proteins can be detected by adding a probe of different color. Such measurements have enabled the real-time observation of protein assembly, trafficking, and degradation [83–85]. An ideal chemical probe for *in vivo* labeling should follow several criteria: high sensitivity (S/N ratio) and specificity (bioorthogonality), fast reaction rate, low interference with function and localization of the target, and minimal cellular perturbation.

Although there have been significant advances in the methods for chemical labeling proteins *in vitro*, only few of them are applicable for *in vivo* labeling. This represents a great challenge for selective chemical labeling of proteins in their native environments. Recent advancements in *in vivo* chemical labeling techniques involve the combination of the specificity of genetically encoded tags and the flexibility of small molecule probes. Such amino acid sequences include tetracysteine/tetraserine motifs, metal chelation motifs, peptide tags for enzymatic modification, ligand-binding domains, and self-labeling enzymatic domains (Table 1). As summarized in Table 1, all these *in vivo* chemical labeling methods have their pros and cons in terms of the tag size, cytotoxicity, specificity, reaction kinetics, signal-to-noise ratio, and accessibility for intracellular labeling. Short peptide tags are possible; however, the specificity has to be conferred by using chelation or enzyme-catalyzed modifications, which usually leads to the problems of cytotoxicity and relatively slow reaction rates, respectively. Tagging systems using ligand-binding domains and self-labeling enzymatic domains are more straightforward and suited for intracellular labeling, whereas these relatively large tags may interfere with protein functions.

Perspectives

The methodologies of protein chemical modification have grown rapidly in the past few years. They have opened up new opportunities to tackle fundamental biological questions. However, there is still a great demand for labeling reagents that display fast kinetics and high specificity. Rapid reactions are important for the detection of biological events that usually occur on a short timescale. Future directions could include the improvement of conjugation chemistry and the discovery of efficient enzymes for modification. Such efforts may involve discovering new bioorthogonal chemistry in water, directed-evolution, and rational design of mutant enzymes.

Chemical labeling in live cells and organisms, albeit challenging, is essential for cell biological studies. Intracellular chemical labeling is more demanding than labeling at the cell surface due to the stringent conditions for chemical reactions inside cells and the problem of cell penetration. Many chemical labeling reagents are cytotoxic or have low cell-permeability, and are thus not suited for intracellular labeling. Therefore, chemical reagents for intracellular labeling with low cytotoxicity, good cell-permeability, fast reaction rates, and low background staining are required in the future. Further development of fluorescent probes may include photo-controllable reporters, ion sensors, two-photon fluorophores, and probes for small biomolecules such as cholesterol [108] and phosphatidylinositol 3-phosphate (PI3P) [109]. Besides organic dyes, many other imaging modes emerge as an interesting avenue, particularly for the diagnosis in whole animals and single molecule studies in cells. These include long-lifetime luminescent lanthanides [110], quantum dots [111], magnetic resonance imaging (MRI) contrasting reagents [112] and radionuclide-filled single-walled carbon nanotubes (SWNT) [113]. Chemical modification of proteins or other biomolecules is a fast-growing field that offers ample opportunities for chemical biologists to contribute to the expansion of novel chemical tools for biological studies.

References

1 Zhang J, Campbell RE, Ting AY, Tsien RY. Creating new fluorescent probes for cell biology. *Nat. Rev. Mol. Cell Biol.* 2002; **3**: 906–918.
2 Shaner NC, Steinbach PA, Tsien RY. A guide to choosing fluorescent proteins. *Nat. Methods* 2005; **2**: 905–909.
3 Wegener AA, Klare JP, Engelhard M, Steinhoff HJ. Structural insights into the early steps of receptor-transducer signal transfer in archaeal phototaxis. *EMBO J.* 2001; **20**: 5312–5319.
4 Schlichting I, Almo SC, Rapp G, Wilson K, Petratos K, Lentfer A, Wittinghofer A, Kabsch W, Pai EF, Petsko GA, Goody RS. Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis. *Nature* 1990; **345**: 309–315.
5 Giepmans BN, Adams SR, Ellisman MH, Tsien RY. The fluorescent toolbox for assessing protein location and function. *Science* 2006; **312**: 217–224.
6 Wong LS, Khan F, Micklefield J. Selective covalent protein immobilization: strategies and applications. *Chem. Rev.* 2009; **109**: 4025–4053.
7 Dawson PE, Kent SB. Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* 2000; **69**: 923–960.
8 Dawson PE, Muir TW, Clark-Lewis I, Kent SB. Synthesis of proteins by native chemical ligation. *Science* 1994; **266**: 776–779.
9 Offer J, Boddy CNC, Dawson PE. Extending synthetic access to proteins with a removable acyl transfer auxiliary. *J. Am. Chem. Soc.* 2002; **124**: 4642–4646.
10 Meutermans WDF, Golding SW, Bourne GT, Miranda LP, Dooley MJ, Alewood PF, Smythe ML. Synthesis of difficult cyclic peptides by inclusion of a novel photolabile auxiliary in a ring contraction strategy. *J. Am. Chem. Soc.* 1999; **121**: 9790–9796.
11 Johnson ECB, Kent SBH. Insights into the mechanism and catalysis of the native chemical ligation reaction. *J. Am. Chem. Soc.* 2006; **128**: 6640–6646.
12 Dawson PE, Churchill MJ, Ghadiri MR, Kent SBH. Modulation of reactivity in native chemical ligation through the use of thiol additives. *J. Am. Chem. Soc.* 1997; **119**: 4325–4329.
13 Bang D, Pentelute BL, Kent SB. Kinetically controlled ligation for the convergent chemical synthesis of proteins. *Angew. Chem., Int. Ed.* 2006; **45**: 3985–3988.
14 Durek T, Torbeev VY, Kent SB. Convergent chemical synthesis and high-resolution x-ray structure of human lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* 2007; **104**: 4846–4851.
15 Hackenberger CPR, Schwarzer D. Chemoselective Ligation and Modification Strategies for Peptides and Proteins. *Angew. Chem., Int. Ed.* 2008; **47**: 10030–10074.
16 Muir TW, Sondhi D, Cole PA. Expressed protein ligation: a general method for protein engineering. *Proc. Natl. Acad. Sci. U.S.A.* 1998; **95**: 6705–6710.
17 Chong S, Mersha FB, Comb DG, Scott ME, Landry D, Vence LM, Perler FB, Benner J, Kucera RB, Hirvonen CA, Pelletier JJ, Paulus H, Xu MQ. Single-column purification of free recombinant proteins using a self- cleavable affinity tag derived from a protein splicing element. *Gene* 1997; **192**: 271–281.

18 Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebl M, Stevens TH. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* 1990; **250**: 651–657.
19 Xu MQ, Perler FB. The mechanism of protein splicing and its modulation by mutation. *EMBO J.* 1996; **15**: 5146–5153.
20 Muir TW. Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* 2003; **72**: 249–289.
21 Cotton GJ, Muir TW. Generation of a dual-labeled fluorescence biosensor for Crk-II phosphorylation using solid-phase expressed protein ligation. *Chem. Biol.* 2000; **7**: 253–261.
22 Cotton GJ, Ayers B, Xu R, Muir TW. Insertion of a synthetic peptide into a recombinant protein framework: a protein biosensor. *J. Am. Chem. Soc.* 1999; **121**: 1100–1101.
23 Alexandrov K, Heinemann I, Durek T, Sidorovitch V, Goody RS, Waldmann H. Intein-mediated synthesis of geranylgeranylated Rab7 protein in vitro. *J. Am. Chem. Soc.* 2002; **124**: 5648–5649.
24 Durek T, Alexandrov K, Goody RS, Hildebrand A, Heinemann I, Waldmann H. Synthesis of fluorescently labeled mono- and diprenylated Rab7 GTPase. *J. Am. Chem. Soc.* 2004; **126**: 16368–16378.
25 Wu YW, Tan KT, Waldmann H, Goody RS, Alexandrov K. Interaction analysis of prenylated Rab GTPase with Rab escort protein and GDP dissociation inhibitor explains the need for both regulators. *Proc. Natl. Acad. Sci. U.S.A.* 2007; **104**: 12294–12299.
26 Wu YW, Oesterlin LK, Tan KT, Waldmann H, Alexandrov K, Goody RS. Membrane targeting mechanism of Rab GTPases elucidated by semisynthetic protein probes. *Nat. Chem. Biol.* 2010; **6**: 534–540.
27 Southworth MW, Adam E, Panne D, Byer R, Kautz R, Perler FB. Control of protein splicing by intein fragment reassembly. *EMBO J.* 1998; **17**: 918–926.
28 Mills KV, Lew BM, Jiang S, Paulus H. Protein splicing in trans by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA intein. *Proc. Natl. Acad. Sci. U.S.A.* 1998; **95**: 3543–3548.
29 Xu MQ, Evans TC, Jr. Recent advances in protein splicing: manipulating proteins in vitro and in vivo. *Curr. Opin. Biotechnol.* 2005; **16**: 440–446.
30 Mootz HD. Split Inteins as Versatile Tools for Protein Semisynthesis. *Chembiochem* 2009; **10**: 2579–2589.
31 Sun W, Yang J, Liu XQ. Synthetic two-piece and three-piece split inteins for protein trans-splicing. *J. Biol. Chem.* 2004; **279**: 35281–35286.
32 Scott CP, bel-Santos E, Wall M, Wahnon DC, Benkovic SJ. Production of cyclic peptides and proteins in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 1999; **96**: 13638–13643.
33 Williams NK, Prosselkov P, Liepinsh E, Line I, Sharipo A, Littler DR, Curmi PM, Otting G, Dixon NE. In vivo protein cyclization promoted by a circularly permuted Synechocystis sp. PCC6803 DnaB mini-intein. *J. Biol. Chem.* 2002; **277**: 7790–7798.
34 Xu R, Ayers B, Cowburn D, Muir TW. Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proc. Natl. Acad. Sci. U.S.A.* 1999; **96**: 388–393.
35 Otomo T, Ito N, Kyogoku Y, Yamazaki T. NMR observation of selected segments in a larger protein: central-segment isotope labeling through intein-mediated ligation. *Biochemistry* 1999; **38**: 16040–16044.
36 Ludwig C, Pfeiff M, Linne U, Mootz HD. Ligation of a synthetic peptide to the N terminus of a recombinant protein using semisynthetic protein trans-splicing. *Angew. Chem., Int. Ed.* 2006; **45**: 5218–5221.
37 Giriat I, Muir TW. Protein semi-synthesis in living cells. *J. Am. Chem. Soc.* 2003; **125**: 7180–7181.
38 Ozawa T, Nogami S, Sato M, Ohya Y, Umezawa Y. A fluorescent indicator for detecting protein-protein interactions in vivo based on protein splicing. *Anal. Chem.* 2000; **72**: 5151–5157.
39 Ozawa T, Kaihara A, Sato M, Tachihara K, Umezawa Y. Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing. *Anal. Chem.* 2001; **73**: 2516–2521.
40 Kanno A, Ozawa T, Umezawa Y. Intein-mediated reporter gene assay for detecting protein-protein interactions in living mammalian cells. *Anal. Chem.* 2006; **78**: 556–560.
41 Paulmurugan R, Umezawa Y, Gambhir SS. Noninvasive imaging of protein-protein interactions in living subjects by using reporter protein complementation and reconstitution strategies. *Proc. Natl. Acad. Sci. U.S.A.* 2002; **99**: 15608–15613.
42 Kanno A, Ozawa T, Umezawa Y. Genetically encoded optical probe for detecting release of proteins from mitochondria toward cytosol in living cells and mammals. *Anal. Chem.* 2006; **78**: 8076–8081.
43 Kim SB, Ozawa T, Umezawa Y. Genetically encoded stress indicator for noninvasively imaging endogenous corticosterone in living mice. *Anal. Chem.* 2005; **77**: 6588–6593.
44 Kim SB, Ozawa T, Watanabe S, Umezawa Y. High-throughput sensing and noninvasive imaging of protein nuclear transport by using reconstitution of split Renilla luciferase. *Proc. Natl. Acad. Sci. U.S.A.* 2004; **101**: 11542–11547.
45 Kwon Y, Coleman MA, Camarero JA. Selective immobilization of proteins onto solid supports through split-intein-mediated protein trans-splicing. *Angew. Chem. Int. Ed. Engl.* 2006; **45**: 1726–1729.
46 Ban H, Gavrilyuk J, Barbas CF, III. Tyrosine bioconjugation through aqueous ene-type reactions: a click-like reaction for tyrosine. *J. Am. Chem. Soc.* 2010; **132**: 1523–1525.
47 Antos JM, McFarland JM, Iavarone AT, Francis MB. Chemoselective tryptophan labeling with rhodium carbenoids at mild pH. *J. Am. Chem. Soc.* 2009; **131**: 6301–6308.
48 Joshi NS, Whitaker LR, Francis MB. A three-component Mannich-type reaction for selective tyrosine bioconjugation. *J. Am. Chem. Soc.* 2004; **126**: 15942–15943.
49 Antos JM, Francis MB. Selective tryptophan modification with rhodium carbenoids in aqueous solution. *J. Am. Chem. Soc.* 2004; **126**: 10256–10257.
50 Saxon E, Bertozzi CR. Cell surface engineering by a modified Staudinger reaction. *Science* 2000; **287**: 2007–2010.
51 Hang HC, Yu C, Kato DL, Bertozzi CR. A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* 2003; **100**: 14846–14851.
52 Vocadlo DJ, Hang HC, Kim EJ, Hanover JA, Bertozzi CR. A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc. Natl. Acad. Sci. U.S.A.* 2003; **100**: 9116–9121.
53 Vocadlo DJ, Bertozzi CR. A strategy for functional proteomic analysis of glycosidase activity from cell lysates. *Angew. Chem., Int. Ed.* 2004; **43**: 5338–5342.
54 Cravatt BF, Wright AT, Kozarich JW. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.* 2008; **77**: 383–414.
55 Evans MJ, Cravatt BF. Mechanism-based profiling of enzyme families. *Chem. Rev.* 2006; **106**: 3279–3301.
56 Lim RKV, Lin Q. Bioorthogonal chemistry: recent progress and future directions. *Chem. Commun. (Camb)* 2010; **46**: 1589–1600.
57 Sletten EM, Bertozzi CR. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* 2009; **48**: 6974–6998.
58 Link AJ, Mock ML, Tirrell DA. Non-canonical amino acids in protein engineering. *Curr. Opin. Biotechnol.* 2003; **14**: 603–609.
59 Prescher JA, Bertozzi CR. Chemistry in living systems. *Nat. Chem. Biol.* 2005; **1**: 13–21.
60 Wang L, Brock A, Herberich B, Schultz PG. Expanding the genetic code of Escherichia coli. *Science* 2001; **292**: 498–500.
61 Wang L, Schultz PG. Expanding the genetic code. *Angew. Chem. Int. Ed.* 2004; **44**: 34–66.
62 Carrico IS, Carlson BL, Bertozzi CR. Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* 2007; **3**: 321–322.
63 Chen I, Howarth M, Lin W, Ting AY. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods* 2005; **2**: 99–104.
64 Fernandez-Suarez M, Baruah H, Martinez-Hernandez L, Xie KT, Baskin JM, Bertozzi CR, Ting AY. Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat. Biotechnol.* 2007; **25**: 1483–1487.
65 Yi, *et al*. submitted.
66 Griffin BA, Adams SR, Tsien RY. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* 1998; **281**: 269–272.
67 Mahal LK, Yarema KJ, Bertozzi CR. Engineering chemical reactivity on cell surfaces through oligosaccharide biosynthesis. *Science* 1997; **276**: 1125–1128.

68 Wang Q, Chan TR, Hilgraf R, Fokin VV, Sharpless KB, Finn MG. Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* 2003; **125**: 3192–3193.
69 Agard NJ, Prescher JA, Bertozzi CR. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* 2004; **126**: 15046–15047.
70 de Araujo AD, Palomo JM, Cramer J, Kohn M, Schroder H, Wacker R, Niemeyer C, Alexandrov K, Waldmann H. Diels–Alder ligation and surface immobilization of proteins. *Angew. Chem. Int. Ed.* 2005; **45**: 296–301.
71 Kho Y, Kim SC, Jiang C, Barma D, Kwon SW, Cheng JK, Jaunbergs J, Weinbaum C, Tamanoi F, Falck J, Zhao YM. A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. *Proc. Natl. Acad. Sci. U.S.A.* 2004; **101**: 12479–12484.
72 Nguyen UT, Cramer J, Gomis J, Reents R, Gutierrez-Rodriguez M, Goody RS, Alexandrov K, Waldmann H. Exploiting the substrate tolerance of farnesyltransferase for site-selective protein derivatization. *Chembiochem* 2007; **8**: 408–423.
73 Xu J, DeGraw AJ, Duckworth BP, Lenevich S, Tann CM, Jenson EC, Gruber SJ, Barany G, Distefano MD. Synthesis and reactivity of 6,7-dihydrogeranylazides: reagents for primary azide incorporation into peptides and subsequent staudinger ligation. *Chem. Biol. Drug Des.* 2006; **68**: 85–96.
74 Nguyen UT, Guo Z, Delon C, Wu Y, Deraeve C, Franzel B, Bon RS, Blankenfeldt W, Goody RS, Waldmann H, Wolters D, Alexandrov K. Analysis of the eukaryotic prenylome by isoprenoid affinity tagging. *Nat. Chem. Biol.* 2009; **5**: 227–235.
75 Dursina BE, Reents R, Niculae A, Veligodsky A, Breitling R, Pyatkov K, Waldmann H, Goody RS, Alexandrov K. A genetically encodable microtag for chemo-enzymatic derivatization and purification of recombinant proteins. *Protein Expr. Purif.* 2005; **39**: 71–81.
76 Wu YW, Waldmann H, Reents R, Ebetino FH, Goody RS, Alexandrov K. A protein fluorescence amplifier: continuous fluorometric assay for rab geranylgeranyltransferase. *Chembiochem* 2006; **7**: 1859–1861.
77 Gauchet C, Labadie GR, Poulter CD. Regio- and chemoselective covalent immobilization of proteins through unnatural amino acids. *J. Am. Chem. Soc.* 2006; **128**: 9274–9275.
78 Duckworth BP, Xu J, Taton TA, Guo A, Distefano MD. Site-specific, covalent attachment of proteins to a solid surface. *Bioconjug. Chem.* 2006; **17**: 967–974.
79 Dursina B, Reents R, Delon C, Wu Y, Kulharia M, Thutewohl M, Veligodsky A, Kalinin A, Evstifeev V, Ciobanu D, Szedlacsek SE, Waldmann H, Goody RS, Alexandrov K. Identification and specificity profiling of protein prenyltransferase inhibitors using new fluorescent phosphoisoprenoids. *J. Am. Chem. Soc.* 2006; **128**: 2822–2835.
80 Duckworth BP, Zhang Z, Hosokawa A, Distefano MD. Selective labeling of proteins by using protein farnesyltransferase. *Chembiochem* 2007; **8**: 98–105.
81 Miller LW, Cornish VW. Selective chemical labeling of proteins in living cells. *Curr. Opin. Chem. Biol.* 2005; **9**: 56–61.
82 O'Hare HM, Johnsson K, Gautier A. Chemical probes shed light on protein function. *Curr. Opin. Struct. Biol.* 2007; **17**: 488–494.
83 Ju W, Morishita W, Tsui J, Gaietta G, Deerinck TJ, Adams SR, Garner CC, Tsien RY, Ellisman MH, Malenka RC. Activity-dependent regulation of dendritic synthesis and trafficking of AMPA receptors. *Nat. Neurosci.* 2004; **7**: 244–253.
84 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.
85 Jansen LE, Black BE, Foltz DR, Cleveland DW. Propagation of centromeric chromatin requires exit from mitosis. *J. Cell Biol.* 2007; **176**: 795–805.
86 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.
87 Guignet EG, Hovius R, Vogel H. Reversible site-selective labeling of membrane proteins in live cells. *Nat. Biotechnol.* 2004; **22**: 440–444.
88 Goldsmith CR, Jaworski J, Sheng M, Lippard SJ. Selective labeling of extracellular proteins containing polyhistidine sequences by a fluorescein-nitrilotriacetic acid conjugate. *J. Am. Chem. Soc.* 2006; **128**: 418–419.
89 Uchinomiya SH, Nonaka H, Fujishima SH, Tsukiji S, Ojida A, Hamachi I. Site-specific covalent labeling of His-tag fused proteins with a reactive Ni(II)-NTA probe. *Chem. Commun. (Camb.)* 2009; 5880–5882.
90 Hauser CT, Tsien RY. A hexahistidine-Zn2+-dye label reveals STIM1 surface exposure. *Proc. Natl. Acad. Sci. U.S.A.* 2007; **104**: 3693–3697.
91 Ojida A, Honda K, Shinmi D, Kiyonaka S, Mori Y, Hamachi I. Oligo-Asp tag/Zn(II) complex probe as a new pair for labeling and fluorescence imaging of proteins. *J. Am. Chem. Soc.* 2006; **128**: 10452–10459.
92 Honda K, Fujishima SH, Ojida A, Hamachi I. Pyrene excimer-based dual-emission detection of a oligoaspartate tag-fused protein by using a Zn(II)-DpaTyr probe. *Chembiochem* 2007; **8**: 1370–1372.
93 Honda K, Nakata E, Ojida A, Hamachi I. Ratiometric fluorescence detection of a tag fused protein using the dual-emission artificial molecular probe. *Chem. Commun. (Camb.)* 2006; 4024–4026.
94 Nonaka H, Tsukiji S, Ojida A, Hamachi I. Non-enzymatic covalent protein labeling using a reactive tag. *J. Am. Chem. Soc.* 2007; **129**: 15777–15779.
95 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.
96 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.
97 Yin J, Straight PD, McLoughlin SM, Zhou Z, Lin AJ, Golan DE, Kelleher NL, Kolter R, Walsh CT. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc. Natl. Acad. Sci. U.S.A.* 2005; **102**: 15815–15820.
98 Zhou Z, Cironi P, Lin AJ, Xu Y, Hrvatin S, Golan DE, Silver PA, Walsh CT, Yin J. Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem. Biol.* 2007; **2**: 337–346.
99 Lin CW, Ting AY. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J. Am. Chem. Soc.* 2006; **128**: 4542–4543.
100 Uttamapinant C, White KA, Baruah H, Thompson S, Fernandez-Suarez M, Puthenveetil S, Ting AY. A fluorophore ligase for site-specific protein labeling inside living cells. *Proc. Natl. Acad. Sci. U.S.A.* 2010; **107**: 10914–10919.
101 Wu P, Shui W, Carlson BL, Hu N, Rabuka D, Lee J, Bertozzi CR. Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag. *Proc. Natl. Acad. Sci. U.S.A.* 2009; **106**: 3000–3005.
102 Popp MW, Antos JM, Grotenbreg GM, Spooner E, Ploegh HL. Sortagging: a versatile method for protein labeling. *Nat. Chem. Biol.* 2007; **3**: 707–708.
103 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.
104 Gautier A, Juillerat A, Heinis C, Correa IR, *Jr.* Kindermann M, Beaufils F, Johnsson K. An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.* 2008; **15**: 128–136.
105 Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, Wood MG, Learish R, Ohana RF, Urh M, Simpson D, Mendez J, Zimmerman K, Otto P, Vidugiris G, Zhu J, Darzins A, Klaubert DH, Bulleit RF, Wood KV. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* 2008; **3**: 373–382.
106 Miller LW, Cai Y, Sheetz MP, Cornish VW. In vivo protein labeling with trimethoprim conjugates: a flexible chemical tag. *Nat. Methods* 2005; **2**: 255–257.
107 Marks KM, Braun PD, Nolan GP. A general approach for chemical labeling and rapid, spatially controlled protein inactivation. *Proc. Natl. Acad. Sci. U.S.A.* 2004; **101**: 9982–9987.
108 Li Z, Mintzer E, Bittman R. First synthesis of free cholesterol-BODIPY conjugates. *J. Org. Chem.* 2006; **71**: 1718–1721.
109 Kanai F, Liu H, Field SJ, Akbary H, Matsuo T, Brown GE, Cantley LC, Yaffe MB. The PX domains of p47phox and p40phox bind to lipid products of PI(3)K. *Nat. Cell Biol.* 2001; **3**: 675–678.
110 Maurel D, Comps-Agrar L, Brock C, Rives ML, Bourrier E, Ayoub MA, Bazin H, Tinel N, Durroux T, Prezeau L, Trinquet E, Pin JP. Cell-surface protein-protein interaction analysis with time-

resolved FRET and snap-tag technologies: application to GPCR oligomerization. *Nat. Methods* 2008; **5**: 561–567.
111 Pinaud F, Clarke S, Sittner A, Dahan M. Probing cellular events, one quantum dot at a time. *Nat. Methods* 2010; **7**: 275–285.
112 Olson ES, Jiang T, Aguilera TA, Nguyen QT, Ellies LG, Scadeng M, Tsien RY. Activatable cell penetrating peptides linked to nanoparticles as dual probes for in vivo fluorescence and MR imaging of proteases. *Proc. Natl. Acad. Sci. U.S.A.* 2010; **107**: 4311–4316.
113 Hong SY, Tobias G, Al-Jamal KT, Ballesteros B, Ali-Boucetta H, Lozano-Perez S, Nellist PD, Sim RB, Finucane C, Mather SJ, Green ML, Kostarelos K, Davis BG. Filled and glycosylated carbon nanotubes for in vivo radioemitter localization and imaging. *Nat. Mater.* 2010; **9**: 485–490.

