

# Genetically Encoding Bioorthogonal Functional Groups for Site-selective Protein Labeling

Wei Wan, Yane-Shih Wang and Wenshe R. Liu\*

Department of Chemistry, Texas A&M University, College Station, TX 77843, USA

## Abstract

Site-selective protein labeling is an indispensable approach in the currently intense chemical biology research area. Studies involving site-selective protein labeling span from the protein dynamic analysis *in vitro* to the protein-protein interaction investigation in living cells. In the past decade, multiple methods have been introduced to achieve site-selective protein labeling. These include genetic fusion of green fluorescent protein and its derivatives, selective chemical labeling of proteins with fusion tags, and site-specific modification of noncanonical amino acids that are genetically encoded. Using evolved orthogonal aminoacyl-tRNA synthetase-nonsense suppressor tRNA pairs, noncanonical amino acids with bioorthogonal functional groups such as azide, alkyne, tetrazine, alkene, keto, phenylhalide, etc. have been genetically incorporated into proteins in *E. coli*, yeast, and mammalian cells. Genetic encoding of these noncanonical amino acids enables multiple ways for site-selective protein labeling both *in vitro* and *in vivo*, allowing diverse strategies to interrogate protein functions. This review intends to provide a brief introduction to the genetic noncanonical amino acid incorporation technique and recent progresses in applying this technique to achieve site-selective protein labeling.

**Keywords:** Noncanonical amino acids; Nonsense suppression; Click chemistry; Bioorthogonal functional groups; Genetic code expansion

## Introduction

Protein labeling with fluorescent molecules that allows sensing and visualization of protein dynamics, localization, protein-ligand interactions, and protein-protein interactions, is an invaluable tool to understand protein functions in living cells. One of the most prominent methods of fluorescent protein labeling is to genetically encode green fluorescent protein (GFP) or one of its variants as a fusion to the protein of interest. This powerful technique has intrinsic important advantages such as high labeling specificity and simplicity [1-5]. In 2008, Nobel Prize in chemistry was awarded to three scientists Osamu Shimomura, Martin Chalfie and Roger Y. Tsien for their discovery and development of GFP, highlighting the great contribution of the GFP technique in advancing chemical and biological research. Although GFP variants have proved to be extremely useful for both *in vitro* and *in vivo* studies of protein functions, their utility is still limited because the molecular sizes of GFP variants (~27kDa) are large enough to potentially interfere with the structure and function of proteins to which they are fused and their spectral and structural characterization are interdependent [6-8]. To increase the diversity of protein labels, approaches comparable and complementary to the GFP technique have been developed that confer selectively fluorescent labeling of proteins with smaller chemical moieties.

Tag-based chemical labeling approaches have flourished recently. They require genetically fusing target proteins to peptide tags that specifically bind to or react with small molecule probes consisting of fluorophores. A great advantage of tag-based chemical labeling approaches is the flexibility in choosing fluorophores. A major advance in the tag-based protein labeling was achieved when biarsenical fluorescent dyes were used to label fusion proteins containing a tetracysteine (TC) motif [9-12]. Binding of a biarsenical dye, notably green-fluorescent FAsH or red fluorescent ReAsH, to the TC tag forms a stable fluorophore-protein complex. The small size of the tag is a proven advantage in several direct comparisons with GFP variants [8,13,14]. Following the introduction of the TC tag, other tags for chemical labeling of proteins have been developed. They fall into two

categories: peptide tags rationally designed or evolved for binding to chemical probes and peptide tags from natural biosynthetic pathways that serve as specific sites for covalent attachment of chemical groups by enzymes. The first category includes small fusion tags [15-19], DHFR-tag that non-covalently binds fluorescent trimethoprim derivatives [20,21], SNAP-tag that covalently reacts with fluorescent O<sup>6</sup>-benzylguanine substrates [22-26], and halo-tag that cleaves the carbon halogen bond of fluorescent ligands to become covalently labeled with the fluorophores [27-29]. Labeling of peptide tags in the second category all involve enzymes. In the ACP-tag system, a phosphopantetheine transferase enzyme is used to transfer a 4'-phosphopantetheine-linked probe from coenzyme A to a serine residue of acyl carrier protein (ACP) that is fused to other proteins [30-35]. In another system, biotin ligase has been used for covalent labeling of a 15-aa peptide tag with ketone-modified biotin molecules that can react with hydrazide or hydroxylamine fluorescent dyes [36-38]. The use of formylglycine-generating enzyme to generate formylglycine in a 13-aa peptide tag for labeling with hydrazide or hydroxylamine fluorescent dyes [39-41] and the exploitation of lipoic acid ligase to transfer an azide-containing lipoic acid probe to a 22-aa peptide tag for labeling with alkyne dyes [42,43] have also been successfully demonstrated. Although many tag-based chemical labeling techniques have advantages like flexibility in choosing fluorophores and less disturbance of structures and functions of target proteins, they have inherent limitations as well. First, a tag is generally fused to the N- or C-terminus of the protein of interest. Installation of a fusion tag at internal part of the protein

\*Corresponding author: Wenshe R. Liu, Department of Chemistry, Texas A&M University, College Station, TX 77843, USA, Tel: 979-845-1746; Fax: 979-845-4719; E-mail: [wliu@chem.tamu.edu](mailto:wliu@chem.tamu.edu)

Received July 25, 2012; Accepted July 25, 2012; Published July 30, 2012

Citation: Wan W, Wang YS, Liu WR (2012) Genetically Encoding Bioorthogonal Functional Groups for Site-selective Protein Labeling. Organic Chem Curr Res 1:e111. doi:10.4172/2161-0401.1000e111

Copyright: © 2012 Wan W, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

without disrupting protein structures and functions is difficult. Second, although most fusion tags are considerably smaller than GFP variants (the TC tag has only six residues), they are still not single residues. In comparison to single-residue modification, these fusion tags more likely adversely affect structures and functions of proteins they are fused to.

There are a few methods available for labeling proteins at single-residue levels. One broadly used approach exploits the reactivity of cysteine residues within proteins and labels them with appropriate thiol-reactive dyes [44]. Native chemical ligation and its extension, expressed protein ligation, have also been used to introduce various probes at single-residue levels [45-47]. These two approaches have, however, been generally limited to *in vitro* modification of purified proteins. Biological studies of such labeled proteins in cells require their reintroduction by invasive techniques such as microinjection or electroporation. Moreover, modification based on cysteine residues requires mutating all other non-targeted cysteine residues that could be critical to protein functions. For the native chemical ligation method, appropriate sites for ligation must be chosen carefully and modification of internal sites in large proteins is cumbersome. Therefore, to resolve issues associated with these aforementioned techniques, a single-residue labeling approach that is easy to perform *in vitro* and allow non-invasive labeling of proteins *in vivo* is necessary.

Genetically incorporating NAAs into proteins is an alternative powerful approach that allows site-selective labeling of proteins at single-residue levels. A general method for the genetic NAA incorporation approach in live cells was developed by Furter, Schultz, and their coworkers [48-55]. This method relies on the read-through of an in-frame stop codon in mRNA by a nonsense suppressor tRNA that is specifically acylated with a NAA by an evolved aminoacyl-tRNA synthetase. There is a naturally existing genetic NAA incorporation system. In some methanogenic archaea and a Gram-positive bacterium *Desulfitobacterium hafniense*, Pyl is co-translationally inserted into proteins by an in-frame amber codon [56-61]. Suppression of this amber codon is mediated by the Pyl amber suppressor tRNA (tRNA<sup>Pyl</sup><sub>CUA</sub>), which has a CUA anticodon and is acylated with Pyl by pyrrolysyl-tRNA synthetase (PylRS). The PylRS-tRNA<sup>Pyl</sup><sub>CUA</sub> pair in these organisms is orthogonal to other synthetase-tRNA pairs in cells, ensuring the fidelity of the Pyl incorporation. Similarly to the Pyl incorporation machinery, an orthogonal synthetase-nonsense suppressor tRNA pair can be developed in which the synthetase is evolved to specifically charge its cognate suppressor tRNA with a NAA. When expressed in cells, this orthogonal synthetase-suppressor tRNA pair enables the NAA to be site-specifically incorporated into a protein at the amber codon with high fidelity and efficiency. Using this approach, a variety of NAAs have been incorporated into proteins in bacteria, yeast and mammalian cells and used to study a large number of biological problems [48,62-66]. Three of these NAAs are fluorescent themselves [65,66]. Many others have chemically reactive groups such as phenylhalide, ketone, azide and alkyne, alkene, tetrazine, tetrazole, etc. These groups can be directly used to introduce fluorescent labels into proteins both *in vitro* and *in vivo* [62,63]. The genetic NAA incorporation approach for fluorescent protein labeling has essential advantages. First, it relies simply on the recombinant DNA technique so that it can be easily generalized. Large amount of modified proteins can be generated easily. Second, the labeling is site-directed and site-specific regardless of the incorporation site or protein size. Third, the localizable labeling in live cells can be achieved using chemically reactive NAAs and fluorescent dyes that have varied permeability to different organelles.

## Fluorescent NAAs

One great advantage of the GFP labeling technique is its efficiency and simplicity. When expressed in cells, a self-catalyzed process generates the GFP fluorophore and ensues quantitative labeling of a target protein that is fused to GFP. On the contrary, most chemical labeling strategies require further treatment after protein expression and therefore are more complicated [67-71]. An ideal chemical labeling approach that can achieve comparable simplicity and efficiency of the GFP labeling technique is to directly incorporate fluorescent NAAs into proteins. As of today, three fluorescent NAAs shown in Figure 1 have been genetically incorporated into proteins. Using an evolved tyrosyl-tRNA synthetase (MjTyrRS)-amber suppressing tRNA<sup>Tyr</sup><sub>CUA</sub> pair that was derived from *Methanocaldococcus jannaschii*, **1** was genetically installed into proteins in *E. coli* at amber mutation sites. **1** contains a fluorescent 7-hydroxycoumarin moiety that shows a high fluorescent quantum yield, a relatively large Stoke's shift, and sensitivity to pH and solvent polarity. These unique features of **1** have been applied to undergo a variety of protein function studies such as protein folding/unfolding, protein-protein interaction, and protein subcellular localization. Using myoglobin incorporated with **1**, Schultz and coworkers showed the sensitivity of fluorescence of **1** to the polarity of its environment could be used to track the unfolding process of myoglobin incorporated with **1** in urea [66]. This same physical property of **1** has also been used to visualize antibody-antigen interactions and the phosphorylation state of STAT3 [72,73]. The strong fluorescence of **1** has also been used to investigate subcellular localization of GroEL and FtsZ in *E. coli*. FtsZ is a bacterial tubulin homologue. FtsZ incorporated with **1** is seen at the cleavage furrow during cell division to form the Z-ring, providing the first example of a fully functional protein to be visualized in living cells using a genetically incorporated NAA [74,75]. Another fluorescent NAA that has been incorporated into proteins is **2** in Figure 1 that contains a fluorescent dansyl functional group. The incorporation of **2** into proteins in yeast and mammalian cells was achieved using an evolved leucyl-tRNA synthetase (EcLeuRS)-amber suppressing tRNA<sup>Tyr</sup><sub>CUA</sub> pair that was derived from *E. coli*. Like **1**, fluorescence of **2** is also sensitive to solvent polarity. Genetic incorporation of **2** into proteins can be applied to study protein folding/unfolding processes [65]. By genetically incorporating **2** into a voltage-dependent membrane lipid phosphatase, Wang and coworkers showed that **2** optically reports the conformational change of the voltage-sensitive domain in response to membrane depolarization [76]. Another fluorescent NAA that has been genetically incorporated into proteins is **3**. Like **1** and **2**, **3** is also an environment-sensitive fluorescent NAA. Using an evolved EcLeuRS-tRNA<sup>Tyr</sup><sub>CUA</sub> pair, **3** was successfully incorporated into glutamine-binding protein at the ligand binding

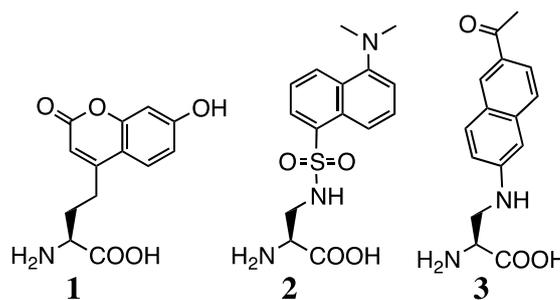


Figure 1: Three fluorescent NAAs genetically incorporated into proteins.

site in yeast. The high fluorescent sensitivity of **3** in glutamine-binding protein to the polarity of the environment allowed easy detection of the conformation rearrangement of glutamine-binding protein during its strong association with glutamine [77].

In comparison to protein-labeling approaches that need protein expression followed with additional chemical labeling, the direct incorporation of fluorescent NAAs into proteins is certainly a more optimal choice given its simplicity and efficiency. However, all three fluorescent NAAs that have been incorporated into proteins have relatively short fluorescent emission wavelengths. These narrow emission spectra and only a few available fluorophores limit the applications of the direct fluorescent NAA incorporation approach. Given the flexibility of PylRS and its variants in the recognition of different NAAs, other fluorescent NAAs could possibly be genetically encoded using evolved PylRS derivatives [78]. The orthogonal nature of the PylRS- $tRNA_{CUA}^{Tyr}$  pair in bacteria, yeast, and mammalian cells will also make fluorescent NAAs encoded using this pair more applicable for broad applications. However, before additional fluorescent NAAs that cover a wide spectral range are genetically encoded, protein-labeling approaches in which genetically encoded NAAs are chemically labeled with structurally diverse fluorophores are still necessary.

### Keto-containing NAAs

The concept of click reactions was introduced about a decade ago [79]. Reactions that can be classified as click reactions need to be selective, modular, and wide in scope. In order to undergo click-type protein-labeling reactions, bioorthogonal functional groups that do not exist in the biological system can be introduced into proteins followed by selective reactions with fluorophore-containing dyes. One of these bioorthogonal functional groups is the keto group. Strictly speaking, keto is not totally bioorthogonal. It exists in the cellular metabolites, cofactors, and a small group of proteins [80,81]. However, in general, keto is not found in proteins and DNA. Given that keto is the most versatile functional group in organic chemistry and selectively reacts with hydrazine- and hydroxylamine-containing molecules, its incorporation into proteins will make it possible to selectively label target proteins with hydrazine- and hydroxylamine-containing dyes. To genetically encode a keto functional group, Schultz and coworkers evolved several MjTyrRS variants for specific incorporation of a NAA **4** shown in Figure 2 into proteins in *E. coli* [62]. **4** has also been genetically encoded in yeast and mammalian cells using evolved tyrosyl-tRNA synthetase (EcTyrRS)-amber suppressing  $tRNA_{CUA}^{Tyr}$  pairs that were derived from *E. coli* [53,54]. Brustad et al. demonstrated a grand application of **4** in the protein folding dynamic analysis [82]. Using orthogonal reactions with a genetically encoded **4** and a cysteine residue, two fluorescent dyes that formed a Förster Resonance Energy Transfer (FRET) pair were introduced in T4 lysozyme for single-molecule FRET analysis of protein folding. Another two NAAs that have been genetically incorporated into proteins using evolved MjTyrRS variants are **5** and **6** [63,83]. By selectively targeting **5**, Zhang et al. showed that a cytoplasmic Z domain protein and outer membrane protein LamB could be selectively visualized with a hydrazide-containing fluorescent dye [63]. Although **4-6** have been proved useful in selective protein labeling, their labeling efficiency at the physiological pH is low. Close to quantitative labeling with hydroxylamine-containing dyes could only be achieved at pH 4 with overnight incubation and labeling with hydrazide-containing dyes exhibited very low efficiency at pH 4-10 [82]. The low reactivity of the keto group in **4-6** is possibly due to its conjugation with an aromatic phenyl ring. The conjugative electron-

donating effect of the phenyl group may reduce the electrophilicity of the keto carbonyl carbon and decrease its reactivity. To resolve this problem, Liu and coworkers designed another NAA **7** shown in Figure 2 [84]. The genetic incorporation of **7** was achieved using an evolved PylRS- $tRNA_{CUA}^{Tyr}$  pair that was original used for the genetic incorporation of *N*<sup>ε</sup>-acetyl-lysine [85]. **7** contains an aliphatic keto group and is in theory more reactive toward hydrazine- and hydroxylamine-containing dyes. As demonstrated, proteins incorporated with **7** could be quantitatively labeled with hydroxylamine-containing dyes or probes at the physiological conditions with 5 hrs incubation.

### Alkyne and Azide-containing NAAs

The copper catalyzed azide-alkyne Huisgen cycloaddition (CuAAC) reaction is a typical click reaction [86]. Both azide and alkyne are biologically inert. Azide itself is also absent in the biological system and alkyne doesn't exist in bio-macromolecules. Given the high reaction specificity and reactivity between azide and alkyne, specific installation of either azide or alkyne into a protein will confer labeling of this protein with a fluorescent dyes that contain a corresponding alkyne or azide functional group. The first NAA that was genetically incorporated into proteins for this purpose is **8** shown in Figure 3. The genetic incorporation of **8** into proteins was achieved in *E. coli* using evolved MjTyrRS- $tRNA_{CUA}^{Tyr}$  pairs and in yeast and mammalian cells using evolved EcTyrRS- $tRNA_{CUA}^{Tyr}$  pairs [53,54,87,88]. Using the CuAAC reaction, proteins incorporated with **8** have been selectively labeled with alkyne-containing fluorescent dyes with high labeling efficiency. It was also demonstrated that the CuAAC reaction worked efficiently to fluorescently label phage particles incorporated with **8** [89]. Deiters et al. also demonstrated that the same reaction could be applied to PEGylate proteins incorporated with **8** [88].

Although the CuAAC reaction has approved advantageous, the requirement to use Cu(I) as a catalyst does have some drawbacks. The Cu(I) catalyst can induce protein aggregation and oxidation, often obviating its application in living systems [90]. Phage particles incorporated with **8** were not viable when they were reacted with an alkyne-containing fluorescein dye in the presence of Cu(I) [91]. To improve biocompatibility of the CuAAC reaction, multiple new

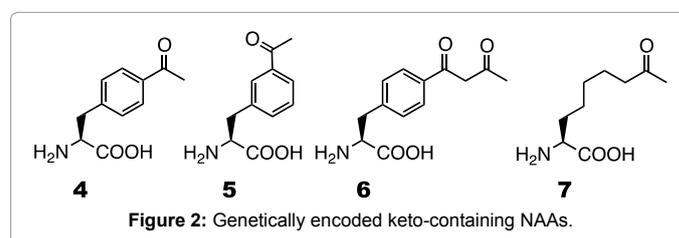


Figure 2: Genetically encoded keto-containing NAAs.

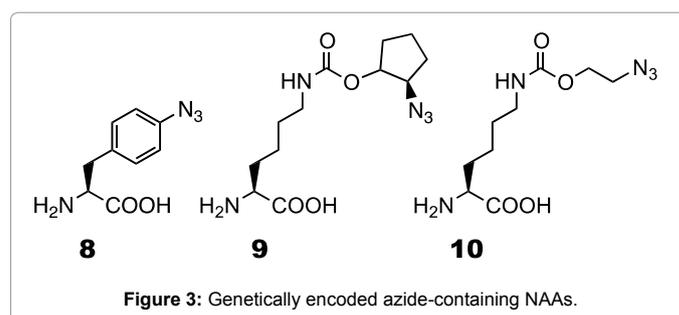


Figure 3: Genetically encoded azide-containing NAAs.

ligands of Cu(I) have been introduced [92,93]. One of these ligands not only largely shields the deleterious effects of Cu(I) and also increases its catalyzed reaction rate [92]. It has been applied in several studies for protein labeling in living cells [94]. There are also two alternative methods for chemically labeling azide-containing proteins without using catalysts. The Staudinger ligation reaction between an azide and a phosphine probe was developed by Bertozzi and coworkers and previously used to modify cell surface carbohydrates in both cellular and in vivo systems [95]. This reaction proceeds with excellent yields under the physiological conditions and is highly selective for azides. This reaction is also biocompatible. Phage particles incorporated with **8** were still viable after their reactions with phosphine-containing dyes [91]. The other reaction that can selectively label an azide-containing protein without the use of a catalyst is the azide-cyclooctyne cycloaddition reaction [96]. This reaction was also developed by Bertozzi and coworkers. A cyclooctyne that has a strain-promoted alkyne functional group undergoes a rapid reaction with an azide. Using a cyclooctyne-containing dye, Liu and coworkers showed that proteins incorporated with **8** could be selectively and efficiently labeled [90].

Beside **8**, two other azide-containing dyes **9** and **10** have also been incorporated into proteins in *E. coli* using evolved PylRS- tRNA<sup>Pyl</sup><sub>CUA</sub> pairs [78,97]. Chen and coworkers showed that HdeA incorporated with **10** in *E. coli* could be selectively labeled with an alkyne-containing and environment-sensitive fluorescent dye. HdeA is an acid-resistant chaperon that shows pH-mediated conformational changes under low pH conditions. One HdeA variants that was fluorescently labeled showed a strong fluorescence increase upon acidification [98]. Since the PylRS- tRNA<sup>Pyl</sup><sub>CUA</sub> pair is orthogonal in yeast and mammalian cells, **9** and **10** could be potentially incorporated into proteins in these cellular systems for selective protein labeling.

Alkyne-containing NAAs that have been genetically encoded are **11-16** in Figure 4. **11-13** contain a terminal alkyne. Proteins incorporated with these NAAs undergo the CuAAC reaction with azide-containing fluorescent dyes. Genetic encoding of **11** has been achieved in *E. coli*, yeast, and mammalian cells using evolved MjTyRS- tRNA<sup>Tyr</sup><sub>CUA</sub> pairs, evolved EcTyRS- tRNA<sup>Tyr</sup><sub>CUA</sub> pairs, and a designed PylRS mutant- tRNA<sup>Pyl</sup><sub>CUA</sub> pair [54,99-101]. **12** and **13** are genetically encoded in cells using the wild type PylRS- tRNA<sup>Pyl</sup><sub>CUA</sub> pair [78,102].

**14-16** contain a cyclooctyne moiety that undergoes the copper free azide-cyclooctyne cycloaddition reaction. All three NAAs have been genetically encoded in *E. coli* and mammalian cells using mutant PylRS- tRNA<sup>Pyl</sup><sub>CUA</sub> pairs [103,104]. Lemke et al. showed that a GFP variant mCherry incorporated with **14** in *E. coli* could be selectively lighted up with a coumarin azide.

### NAAs that Undergo Strain-Promoted Inverse-electron-demanding Diels-Alder Cycloaddition

Besides the strain-promoted azide-cyclooctyne cycloaddition reaction, **14-16** also undergo strain-promoted inverse-electron-demanding cycloaddition with tetrazine-containing molecules that can exhibit accelerated reaction rates using strained reactants and furthermore is irreversible because of the loss of N<sub>2</sub> [105]. This chemistry has been used in cells to label small molecules and is magnitudes faster than the classical CuAAC reaction. Lemke and coworkers showed that **14** and **15** efficiently reacted with a tetrazine-containing dye with a

second order reaction rate reaching to 400 M<sup>-1</sup>s<sup>-1</sup> [103,106]. Maltose binding protein incorporated with **14** in *E. coli* could be efficiently and rapidly labeled with a tetrazine-containing coumarin dye. In comparison to **14** and **15**, **16** has a higher reaction rate with a tetrazine dye with a second order reaction rate close to 1200 M<sup>-1</sup>s<sup>-1</sup> [104]. Fusion proteins such as EGFR-GFP and jun-mCherry incorporated with **15** in HeLa cells could be efficiently visualized with a TAMRA-tetrazine dye.

Other NAAs that undergo strain-promoted inverse-electron-demanding Diels-Alders cycloaddition with tetrazine-containing molecules include **17-19** in Figure 5. **17** and **18** contain a norbornene moiety that contains a strain-promoted alkene. The genetic incorporation of **17** and **18** has been achieved using mutant PylRS- tRNA<sup>Pyl</sup><sub>CUA</sub> pairs [103,104]. Proteins incorporated with **17** and **18** reacted rapidly with tetrazine-containing dyes. Chin and coworkers showed that mammalian membrane proteins incorporated with **17** could be efficiently labeled with tetrazine-containing dyes [104]. **19** is also genetically encoded using mutant PylRS- tRNA<sup>Pyl</sup><sub>CUA</sub> pairs [103,104]. It has a trans-cyclooctene moiety with a highly strain-promoted alkene group. This strain-promoted alkene group can undergo inverse-electron demanding cycloaddition with a tetrazine dye with a reaction rate close to 35,000 M<sup>-1</sup>s<sup>-1</sup>. As far as we notice, this is the fastest click reaction that has been reported. A fusion protein, NLS-MBP-GFP incorporated with **19** in HeLa cells could be efficiently labeled with a tetrazine-containing Cy5 dye in just about 5 min incubation. This fast labeling process could effectively avoid exposing cells in non-native conditions and increase cell viability. There is another advantage of using inverse-electron-demanding Diels-Alder cycloaddition to label a strain-promoted alkene. The tetrazine moiety itself can efficiently quench fluorescence of a fluorophore that is covalently linked to it. Therefore, before its reaction with a strain-promoted alkene-containing NAA, a tetrazine-containing dye is not fluorescent. However, the tetrazine moiety is lost after reaction and then the fluorophore emits strong fluorescence. Thus, a tetrazine-containing dye is a "turn-on" fluorophore of a strain-promoted alkene. Cleaning the residual dye after reaction is not necessary given its low background.

Like strain-promoted alkene-containing NAAs, a tetrazine-containing NAA could also be genetically encoded. Using an evolved MjTyRS- tRNA<sup>Tyr</sup><sub>CUA</sub> pair, **20** was genetically incorporated into proteins

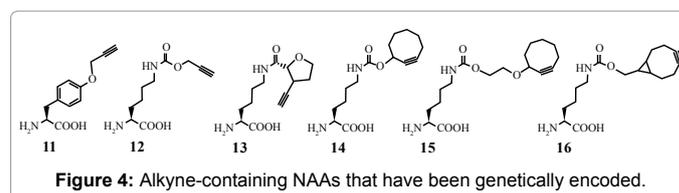


Figure 4: Alkyne-containing NAAs that have been genetically encoded.

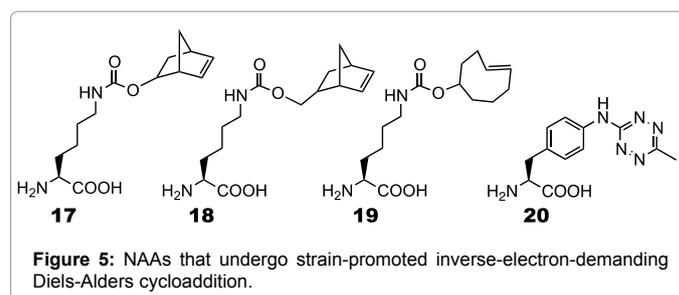
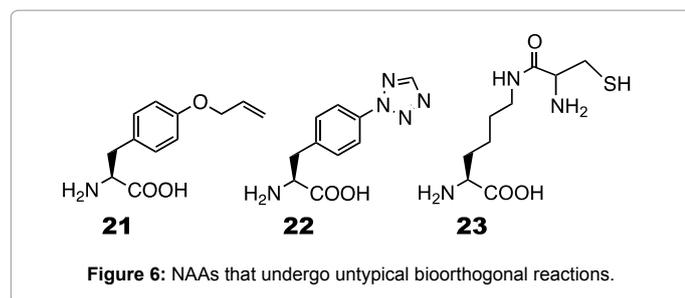


Figure 5: NAAs that undergo strain-promoted inverse-electron-demanding Diels-Alders cycloaddition.



in *E. coli* [107]. Mehl and coworkers showed that **20** reacted with a strained trans-cyclooctene with a rate of  $880 \text{ M}^{-1}\text{s}^{-1}$ . Using this strained trans-cyclooctene in living *E. coli* cells, GFP incorporated with **20** that was not fluorescent due to the fluorescent quenching effect of **20** could be rapidly lighted up.

### Other NAAs

There are two other reaction types that may be considered as click reactions and used in selective protein labeling. Lin and coworkers showed that a tetrazole undergoes photolysis to form a nitrile imine that selectively reacts with an alkene [108]. This is called a photoclick reaction. Using a tetrazole-containing dye, Lin and coworkers demonstrated proteins incorporated with **21** (Figure 6) could be selectively labeled under UV irradiation [109,110]. A tetrazole-containing NAA, **22** was also genetically incorporated into proteins in *E. coli* [109]. Proteins incorporated with **22** underwent photo click reaction with alkene-containing dyes. Cyanobenzothiazole condensation with 1,2-aminothiol is another reaction type that is considered bioorthogonal. Chan, Chin, and coworkers showed that a 1,2-aminothiol-containing NAA **23** could be genetically incorporated into proteins in *E. coli* using either a wild type or evolved evolved PylRS- tRNA<sup>Pyl</sup><sub>CUA</sub> pair [111,112]. The purified protein incorporated with **23** underwent efficient labeling with a cyanobenzothiazole-containing dye.

### A Future Direction

So far, multiple bioorthogonal click type reactions have been developed for selective protein labeling. However, the orthogonal nature of these reactions to each other is not very much explored. Developing two orthogonal click reactions in living cells could be potentially important in selectively labeling one protein with two different dyes for protein folding/unfolding analysis inside living cells and selectively labeling two proteins with two dyes for their interaction analysis. Recently, Liu and Chin Groups independently developed two methods for genetic incorporation of two different NAAs into one protein in living cells [113,114]. These same systems could also be applied to synthesize two different proteins that are incorporated with two different NAAs for their following chemical modifications. In order to undergo selective modifications of two different NAAs in living cells, two orthogonal click reactions are necessary. We think one important future direction of the current NAA-directed fluorescent protein labeling research is to identify and optimize two orthogonal click reactions for rapid labeling of two different NAAs in one cell.

### Acknowledgement

This work was partly supported by Welch Research Grant A-1715 to W. R. Liu.

### References

1. Giepmans BN, Adams SR, Ellisman MH, Tsien RY (2006) The fluorescent toolbox for assessing protein location and function. *Science* 312: 217-224.

2. Zhang J, Campbell RE, Ting AY, Tsien RY (2002) Creating new fluorescent probes for cell biology. *Nat Rev Mol Cell Biol* 3: 906-918.
3. Tsien RY (1998) The green fluorescent protein. *Annu Rev Biochem* 67: 509-544.
4. Lippincott-Schwartz J, Patterson GH (2003) Development and use of fluorescent protein markers in living cells. *Science* 300: 87-91.
5. Chiesa A, Rapizzi E, Tosello V, Pinton P, de Virgilio M, et al. (2001) Recombinant aequorin and green fluorescent protein as valuable tools in the study of cell signalling. *Biochem J* 355: 1-12.
6. Lisenbee CS, Karnik SK, Trelease RN (2003) Overexpression and mislocalization of a tail-anchored GFP redefines the identity of peroxisomal ER. *Traffic* 4: 491-501.
7. Tian GW, Mohanty A, Chary SN, Li S, Paap B, et al. (2004) High-throughput fluorescent tagging of full-length Arabidopsis gene products in planta. *Plant Physiol* 135: 25-38.
8. Andresen M, Schmitz-Salue R, Jakobs S (2004) Short tetracysteine tags to beta-tubulin demonstrate the significance of small labels for live cell imaging. *Mol Biol Cell* 15: 5616-5622.
9. Griffin B A, Adams S R, Tsien R Y (1998) Specific Covalent Labeling of Recombinant Protein Molecules Inside Live Cells *Science* 281: 269-272.
10. Griffin B A, Adams S R, Jones J, Tsien R Y (2000) Fluorescent labeling of recombinant proteins in living cells with FIAsh *Methods Enzymol* 327: 565-578.
11. Adams SR, Campbell RE, Gross LA, Martin BR, Walkup GK, et al. (2002) New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. *J Am Chem Soc* 124: 6063-6076.
12. Machleidt T, Robers M, Hanson GT (2007) Protein labeling with FIAsh and ReAsH. *Methods Mol Biol* 356: 209-220.
13. Hoffmann C, Gaietta G, Bunemann M, Adams SR, Oberdorff-Maass S, et al. (2005) A FIAsh-based FRET approach to determine G protein-coupled receptor activation in living cells. *Nat Methods* 2: 171-176.
14. Dyachok O, Isakov Y, Sagetorp J, Tengholm A (2006) *Nature* 439: 349.
15. Nonaka H, Tsukiji S, Ojida A, Hamachi I (2007) Non-enzymatic covalent protein labeling using a reactive tag. *J Am Chem Soc* 129: 15777-15779.
16. Ojida A, Honda K, Shinmi D, Kiyonaka S, Mori Y, et al. (2006) Oligo-Asp tag/Zn(II) complex probe as a new pair for labeling and fluorescence imaging of proteins. *J Am Chem Soc* 128: 10452-10459.
17. Hauser CT, Tsien RY (2007) A hexahistidine-Zn<sup>2+</sup>-dye label reveals STIM1 surface exposure. *Proc Natl Acad Sci U S A* 104: 3693-3697.
18. Franz KJ, Nitz M, Imperiali B (2003) Lanthanide-Binding Tags as Versatile Protein Coexpression Probes. *ChemBiochem* 4: 265-271.
19. Marks KM, Rosinov M, Nolan GP (2004) In vivo targeting of organic calcium sensors via genetically selected peptides. *Chem Biol* 11: 347-356.
20. Israel DI, Kaufman RJ (1993) Dexamethasone negatively regulates the activity of a chimeric dihydrofolate reductase/glucocorticoid receptor protein. *Proc Natl Acad Sci U S A* 90: 4290-4294.
21. Calloway NT, Choob M, Sanz A, Sheetz MP, Miller LW, et al. (2007) Optimized fluorescent trimethoprim derivatives for in vivo protein labeling. *ChemBiochem* 8: 767-774.
22. Keppler A, Pick H, Arrivoli C, Vogel H, Johnsson K (2004) Labeling of fusion proteins with synthetic fluorophores in live cells. *Proc Natl Acad Sci U S A* 101: 9955-9959.
23. Keppler A, Kindermann M, Gendreizig S, Pick H, Vogel H (2004) Labeling of fusion proteins of O<sup>6</sup>-alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro. *Methods* 32: 437-444.
24. Juillerat A, Gronemeyer T, Keppler A, Gendreizig S, Pick H (2003) Directed Evolution of O<sup>6</sup>-Alkylguanine-DNA Alkyltransferase for Efficient Labeling of Fusion Proteins with Small Molecules In Vivo. *Chem Biol* 10: 313-317.
25. Keppler A, Gendreizig S, Gronemeyer T, Pick H, Vogel H (2003) *Nat Biotechnol* 21: 86.
26. Jansen LE, Black BE, Foltz DR, Cleveland DW (2007) Propagation of centromeric chromatin requires exit from mitosis. *J Cell Biol* 176: 795-805.

27. Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, et al. (2008) HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* 3: 373-382.
28. Lang C, Schulze J, Mendel RR, HÄnsch R (2006) HaloTag: a new versatile reporter gene system in plant cells. *J Exp Bot* 57: 2985-2992.
29. Urh M, Hartzell D, Mendez J, Klaubert D H, Wood K (2008) Methods for Detection of Protein-Protein and Protein-DNA Interactions Using HaloTag™. *Methods Mol Biol* 421: 191-210.
30. George N, Pick H, Vogel H, Johnsson N, Johnsson K (2004) Specific labeling of cell surface proteins with chemically diverse compounds. *J Am Chem Soc* 126: 8896-8897.
31. Yin J, Liu F, Li X, Walsh CT (2004) Labeling proteins with small molecules by site-specific posttranslational modification. *J Am Chem Soc* 126: 7754-7755.
32. Vivero-Pol L, George N, Krumm H, Johnsson K, Johnsson N (2005) Multicolor imaging of cell surface proteins. *J Am Chem Soc* 127: 12770-12771.
33. Prummer M, Meyer BH, Franzini R, Segura JM, George N, et al. (2006) Post-translational covalent labeling reveals heterogeneous mobility of individual G protein-coupled receptors in living cells. *ChemBiochem* 7: 908-911.
34. Meyer BH, Martinez KL, Segura JM, Pascoal P, Hovius R, et al. (2006) Covalent labeling of cell-surface proteins for in-vivo FRET studies. *FEBS Lett* 580: 1654-1658.
35. Meyer BH, Segura JM, Martinez KL, Hovius R, George N, et al. (2006) FRET imaging reveals that functional neurokinin-1 receptors are monomeric and reside in membrane microdomains of live cells. *Proc Natl Acad Sci U S A* 103: 2138-2143.
36. Chen I, Howarth M, Lin W, Ting AY (2005) Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Methods* 2: 99-104.
37. Howarth M, Takao K, Hayashi Y, Ting AY (2005) Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc Natl Acad Sci U S A* 102: 7583-7588.
38. Slavoff SA, Chen I, Choi YA, Ting AY (2008) Expanding the substrate tolerance of biotin ligase through exploration of enzymes from diverse species. *J Am Chem Soc* 130: 1160-1162.
39. Wu P, Shui W, Carlson BL, Hu N, Rabuka D, et al. (2009) 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* 106: 3000-3005.
40. Rush JS, Bertozzi CR (2008) New aldehyde tag sequences identified by screening formylglycine generating enzymes in vitro and in vivo. *J Am Chem Soc* 130: 12240-12241.
41. Carrico IS, Carlson BL, Bertozzi CR (2007) Introducing genetically encoded aldehydes into proteins. *Nat Chem Biol* 3: 321-322.
42. Fernandez-Suarez M, Baruah H, Martinez-Hernandez L, Xie KT, Baskin JM, et al. (2007) Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat Biotechnol* 25: 1483-1487.
43. Baruah H, Puthenveetil S, Choi YA, Shah S, Ting AY (2008) An engineered aryl azide ligase for site-specific mapping of protein-protein interactions through photo-cross-linking. *Angew Chem Int Ed Engl* 47: 7018-7021.
44. Hermanson G T (1996) *Bioconjugate Techniques*. Academic Press: Dan Diego, California.
45. Dawson PE, Muir TW, Clark-Lewis I, Kent SB (1994) Synthesis of proteins by native chemical ligation. *Science* 266: 776-779.
46. Muir TW, Sondhi D, Cole PA (1998) Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A* 95: 6705-6710.
47. Muralidharan V, Muir TW (2006) Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat Methods* 3: 429-438.
48. Wang L, Xie J, Schultz PG (2006) Expanding the genetic code. *Annu Rev Biophys Biomol Struct* 35: 225-249.
49. Wang L, Schultz PG (2004) Expanding the genetic code. *Angew Chem Int Ed Engl* 44: 34-66.
50. Wang L, Schultz P G (2002) *Chem Commun (Camb)*.
51. Wang L, Schultz PG (2001) A general approach for the generation of orthogonal tRNAs. *Chem Biol* 8: 883-890.
52. Wang L, Brock A, Herberich B, Schultz PG (2001) Expanding the genetic code of *Escherichia coli*. *Science* 292: 498-500.
53. Chin JW, Cropp TA, Anderson JC, Mukherji M, Zhang Z, et al. (2003) An expanded eukaryotic genetic code. *Science* 301: 964-967.
54. Liu W, Brock A, Chen S, Schultz P G (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat Methods* 4: 239-244.
55. Furter R (1998) Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in *Escherichia coli*. *Protein Sci* 7: 419-426.
56. Ibba M, Soll D (2002) Genetic code: introducing pyrrolysine. *Curr Biol* 12: 464-466.
57. Srinivasan G, James CM, Krzycki JA (2002) Pyrrolysine encoded by UAG in *Archaea*: charging of a UAG-decoding specialized tRNA. *Science* 296: 1459-1462.
58. Polcarpo C, Ambrogelly A, Berube A, Winbush SM, McCloskey JA, et al. (2004) An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc Natl Acad Sci U S A* 101: 12450-12454.
59. Blight SK, Larue RC, Mahapatra A, Longstaff DG, Chang E, et al. (2004) Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature* 431: 333-335.
60. Longstaff DG, Larue RC, Faust JE, Mahapatra A, Zhang L, et al. (2007) A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proc Natl Acad Sci U S A* 104: 1021-1026.
61. Herring S, Ambrogelly A, Polcarpo C R, Soll D (2007) *Nucleic Acids Res* 35: 1270.
62. Wang L, Zhang Z, Brock A, Schultz PG (2003) Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc Natl Acad Sci U S A* 100: 56-61.
63. Zhang Z, Smith BA, Wang L, Brock A, Cho C, et al. (2003) A new strategy for the site-specific modification of proteins in vivo. *Biochemistry* 42: 6735-6746.
64. Lemke EA, Summerer D, Geierstanger BH, Brittain SM, Schultz PG (2007) Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat Chem Biol* 3: 769-772.
65. Summerer D, Chen S, Wu N, Deiters A, Chin JW, et al. (2006) A genetically encoded fluorescent amino acid. *Proc Natl Acad Sci U S A* 103: 9785-9789.
66. Wang J, Xie J, Schultz PG (2006) A genetically encoded fluorescent amino acid. *J Am Chem Soc* 128: 8738-8739.
67. Lin M Z, Wang L (2008) *Physiology (Bethesda)* 23: 131.
68. Miller LW, Cornish VW (2005) Selective chemical labeling of proteins in living cells. *Curr Opin Chem Biol* 9: 56-61.
69. Marks KM, Nolan GP (2006) Chemical labeling strategies for cell biology. *Nat Methods* 3: 591-596.
70. Dragulescu-Andrasi A, Rao J (2007) Chemical labeling of protein in living cells. *ChemBiochem* 8: 1099-1101.
71. O'Hare HM, Johnsson K, Gautier A (2007) Chemical probes shed light on protein function. *Curr Opin Struct Biol* 17: 488-494.
72. Lacey VK, Parrish AR, Han S, Shen Z, Briggs SP, et al. (2011) A fluorescent reporter of the phosphorylation status of the substrate protein STAT3. *Angew Chem Int Ed Engl* 50: 8692-8696.
73. Mills JH, Lee HS, Liu CC, Wang J, Schultz PG (2009) A Genetically Encoded Direct Sensor of Antibody-Antigen Interactions. *ChemBiochem* 10: 2162-2164.
74. Charbon G, Brustad E, Scott KA, Wang J, Lobner-Olesen A, et al. (2011) Subcellular protein localization by using a genetically encoded fluorescent amino acid. *ChemBiochem* 12: 1818-1821.
75. Charbon G, Brustad E, Scott KA, Wang J, LÅbner-Olesen A, et al. (2011) Subcellular protein localization by using a genetically encoded fluorescent amino acid. *ChemBiochem* 12: 1818-1821.
76. Shen B, Xiang Z, Miller B, Louie G, Wang W, et al. (2011) Genetically encoding unnatural amino acids in neural stem cells and optically reporting voltage-sensitive domain changes in differentiated neurons. *Stem Cells* 29: 1231-1240.
77. Lee HS, Guo J, Lemke EA, Dimla RD, Schultz PG (2009) Genetic incorporation of a small, environmentally sensitive, fluorescent probe into proteins in *Saccharomyces cerevisiae*. *J Am Chem Soc* 131: 12921-12923.

78. Nguyen D P, Lusic H, Neumann H, Kapadnis P B, Deiters A (2009) Genetic Encoding and Labeling of Aliphatic Azides and Alkynes in Recombinant Proteins via a Pyrrolysyl-tRNA Synthetase/tRNA<sub>CUA</sub> Pair and Click Chemistry *J Am Chem Soc* 131: 8720-8721.
79. Sharpless K B, Kolb H C (1999) *Abstr Pap Am Chem S* 217: 95.
80. Eliot AC, Kirsch JF (2004) Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu Rev Biochem* 73: 383-415.
81. van Poelje PD, Snell EE (1990) Pyruvoyl-dependent enzymes. *Annu Rev Biochem* 59: 29-59.
82. Brustad EM, Lemke EA, Schultz PG, Deniz AA (2008) A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. *J Am Chem Soc* 130: 17664-17665.
83. Zeng H, Xie J, Schultz PG (2006) Genetic introduction of a diketone-containing amino acid into proteins. *Bioorg Med Chem Lett* 16: 5356-5359.
84. Huang Y, Wan W, Russell W K, Pai P J, Wang Z (2010) *Med Chem Lett* 20: 878.
85. Neumann H, Peak-Chew SY, Chin JW (2008) Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat Chem Biol* 4: 232-234.
86. Kolb HC, Finn MG, Sharpless KB (2001) Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl* 40: 2004-2021.
87. Chin JW, Santoro SW, Martin AB, King DS, Wang L, et al. (2002) Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J Am Chem Soc* 124: 9026-9027.
88. Deiters A, Cropp TA, Summerer D, Mukherji M, Schultz PG (2004) Site-specific PEGylation of proteins containing unnatural amino acids. *Bioorg Med Chem Lett* 14: 5743-5745.
89. Tian F, Tsao ML, Schultz PG (2004) A phage display system with unnatural amino acids. *J Am Chem Soc* 126: 15962-15963.
90. Wu B, Wang Z, Huang Y, Liu W R (2012) *Chembiochem* 13: 1405.
91. Tsao ML, Tian F, Schultz PG (2005) Selective Staudinger modification of proteins containing p-azidophenylalanine. *Chembiochem* 6: 2147-2149.
92. Besanceney-Webler C, Jiang H, Zheng T, Feng L, Soriano Del Amo D, et al. (2011) Increasing the Efficacy of Bioorthogonal Click Reactions for Bioconjugation: A Comparative Study. *Angew Chem Int Ed Engl*.
93. Hong V, Presolski SI, Ma C, Finn MG (2009) Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew Chem Int Ed Engl* 48: 9879-9883.
94. Jiang, H, Feng L, Soriano del Amo D, Seidel III R D, Marlow F, et al. (2011) Imaging Glycans in Zebrafish Embryos by Metabolic Labeling and Bioorthogonal Click Chemistry. *J. Vis. Exp.* 52: 2686.
95. Saxon E, Bertozzi CR (2000) Cell surface engineering by a modified Staudinger reaction. *Science* 287: 2007-2010.
96. Agard NJ, Prescher JA, Bertozzi CR (2004) A strain-promoted [3+2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J Am Chem Soc* 126: 15046-15047.
97. Hao Z, Song Y, Lin S, Yang M, Liang Y (2011) *Chem Commun (Camb)* 47: 4502.
98. Yang M, Song Y, Zhang M, Lin S, Hao Z, et al. (2012) Converting a Solvatochromic Fluorophore into a Protein-Based pH Indicator for Extreme Acidity. *Angew Chem Int Ed Engl* .
99. Deiters A, Schultz PG (2005) In vivo incorporation of an alkyne into proteins in *Escherichia coli*. *Bioorg Med Chem Lett* 15: 1521-1524.
100. Deiters A, Cropp TA, Mukherji M, Chin JW, Anderson JC, et al. (2003) Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*. *J Am Chem Soc* 125: 11782-11783.
101. Wang YS, Fang X, Wallace AL, Wu B, Liu WR (2012) A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. *J Am Chem Soc* 134: 2950-2953.
102. Fekner T, Li X, Lee MM, Chan MK (2009) A pyrrolysine analogue for protein click chemistry. *Angew Chem Int Ed Engl* 48: 1633-1635.
103. Plass T, Milles S, Koehler C, Szymanski J, Mueller R, et al. (2012) Amino acids for Diels-Alder reactions in living cells. *Angew Chem Int Ed Engl* 51: 4166-4170.
104. Lang K, Davis L, Wallace S, Mahesh M, Cox DJ (2012) *J Am Chem Soc* 134: 10317.
105. Devaraj NK, Upadhyay R, Haun JB, Hilderbrand SA, Weissleder R (2009) Fast and sensitive pretargeted labeling of cancer cells through a tetrazine/trans-cyclooctene cycloaddition. *Angew Chem Int Ed Engl* 48: 7013-7016.
106. Plass T, Milles S, Koehler C, Schultz C, Lemke EA (2011) *Angew Chem Int Ed Engl* 50: 3878.
107. Seitchik JL, Peeler JC, Taylor MT, Blackman ML, Rhoads TW, et al. (2012) Genetically encoded tetrazine amino acid directs rapid site-specific in vivo bioorthogonal ligation with trans-cyclooctenes. *J Am Chem Soc* 134: 2898-2901.
108. Song W, Wang Y, Qu J, Madden MM, Lin Q (2008) A photoinducible 1,3-dipolar cycloaddition reaction for rapid, selective modification of tetrazole-containing proteins. *Angew Chem Int Ed Engl* 47: 2832-2835.
109. Wang J, Zhang W, Song W, Wang Y, Yu Z, et al. (2010) A biosynthetic route to photoclick chemistry on proteins. *J Am Chem Soc* 132: 14812-14818.
110. Song W, Wang Y, Qu J, Lin Q (2008) Selective functionalization of a genetically encoded alkene-containing protein via "photoclick chemistry" in bacterial cells. *J Am Chem Soc* 130: 9654-9655.
111. Nguyen DP, Elliott T, Holt M, Muir TW, Chin JW (2011) Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. *J Am Chem Soc* 133: 11418-11421.
112. Li X, Fekner T, Ottesen JJ, Chan MK (2009) A pyrrolysine analogue for site-specific protein ubiquitination. *Angew Chem Int Ed Engl* 48: 9184-9187.
113. Wan W, Huang Y, Wang Z, Russell WK, Pai PJ, et al. (2010) A facile system for genetic incorporation of two different noncanonical amino acids into one protein in *Escherichia coli*. *Angew Chem Int Ed Engl* 49: 3211-3214.
114. Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* 464: 441-444.