ACCOUNTS
of chemical research

Introducing Bioorthogonal Functionalities into Proteins in Living Cells

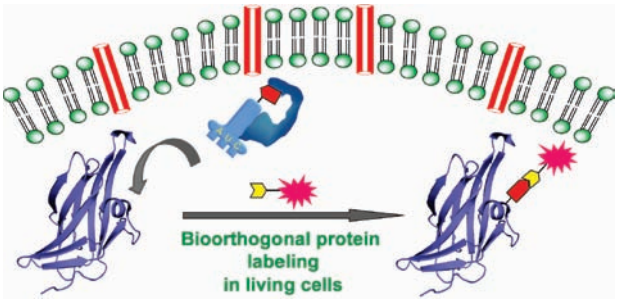
ZIYANG HAO,[†] SENLIAN HONG,[†,‡] XING CHEN,[*,†] AND PENG R. CHEN[*,†,§]

[†]*Beijing National Laboratory for Molecular Sciences and Department of Chemical Biology, College of Chemistry and Molecular Engineering,* [‡]*School of Life Sciences, Peking University, Beijing 100871, China, and* [§]*Laboratory of Chemical Genomics, School of Chemical Biology and Biotechnology, Shenzhen Graduate School of Peking University, Shenzhen 518055, China*

RECEIVED ON MARCH 1, 2011

CONSPECTUS

Proteins are the workhorses of the cell, playing crucial roles in virtually every biological process. The revolutionary ability to visualize and monitor proteins in living systems, which is largely the result of the development of green fluorescence protein (GFP) and its derivatives, has dramatically expanded our understanding of protein dynamics and function. Still, GFPs are ill suited in many circumstances; one major drawback is their relatively large size, which can significantly perturb the functions of the native proteins to which they are fused.



To bridge this gap, scientists working at the chemistry–biology interface have developed methods to install bioorthogonal functional groups into proteins in living cells. The bioorthogonal group is, by definition, a non-native and nonperturbing chemical group. But more importantly, the installed bioorthogonal handle is able to react with a probe bearing a complementary functionality in a highly selective fashion and with the cell operating in its physiological state. Although extensive efforts have been directed toward the development of bioorthogonal chemical reactions, introducing chemical functionalities into proteins in living systems remains an ongoing challenge. In this Account, we survey recent progress in this area, focusing on a genetic code expansion approach.

In nature, a cell uses posttranslational modifications to append the necessary functional groups into proteins that are beyond those contained in the canonical 20 amino acids. Taking lessons from nature, scientists have chosen or engineered certain enzymes to modify target proteins with chemical handles. Alternatively, one can use the cell's translational machinery to genetically encode bioorthogonal functionalities, typically in the form of unnatural amino acids (UAAs), into proteins; this can be done in a residue-specific or a site-specific manner. For studying protein dynamics and function in living cells, site-specific modification by means of genetic code expansion is usually favored.

A variety of UAAs bearing bioorthogonal groups as well as other functionalities have been genetically encoded into proteins of interest. Although this approach is well established in bacteria, tagging proteins in mammalian cells is challenging. A facile pyrrolysine-based system, which might potentially become the "one-stop shop" for protein modification in both prokaryotic and eukaryotic cells, has recently emerged. This technology can effectively introduce a series of bioorthogonal handles into proteins in mammalian cells for subsequent chemical conjugation with small-molecule probes. Moreover, the method may provide more precise protein labeling than GFP tagging. These advancements build the foundation for studying more complex cellular processes, such as the dynamics of important receptors on living mammalian cell surfaces.

Introduction

Proteins are the most abundant biomolecule within cells and participate in essentially all processes. Dissecting these processes in many cases requires the ability to label proteins in the context of living cells.[1] The genetically encoded green fluorescent protein (GFP) and its variants are undeniably the most powerful tool for protein labeling and visualization and have been widely utilized for studying

Published on the Web 06/02/2011 www.pubs.acs.org/accounts
10.1021/ar200067r

protein expression, trafficking, and localization in living systems.[2] However, the GFP fusion technique also has some limitations. The relatively large size of GFP tags can cause significant structural perturbation and thus influence the expression, localization, or function of the proteins of interest. Second, the fusion of GFP is largely confined to the N- or C-terminus of the target proteins. Furthermore, GFP tags can only be visualized by fluorescence methods and are not directly applicable for other imaging techniques. These limitations have therefore promoted the recent expansion of the arsenal of bioorthogonal chemical reactions used to label proteins in their native settings.[3−13]

Bioorthogonal protein labeling involves the incorporation of a unique chemical group into the protein of interest, followed by a bioorthogonal reaction to covalently attach a biophysical probe bearing a complementary functional group (Figure 1).[11,12] A bioorthogonal reaction typically requires the two participating components (bioorthogonal reaction pair) to be mutually reactive while remaining inert to the surrounding molecules under the physiological environment. So far, only a limited number of chemical reactions have been developed to meet the requisite requirements, exemplified by the azide−phosphine Staudinger−Bertozzi ligation and the different versions of azide−alkyne cycloaddition "click" chemistry that will be discussed in details below.[14−19] These reactions have been widely utilized for labeling proteins, as well as nucleic acids, glycans, and lipids in the context of living cells and whole organisms.[1]

One of the key challenges in using this chemical strategy for protein labeling is to selectively incorporate the bioorthogonal chemical groups into the target proteins, which allows the subsequent conjugation with probes bearing the complementary functionalities. These bioorthogonal groups have to be non-native and nonperturbing in order to satisfy the aforementioned stringent "bioorthogonal" requirements. A variety of strategies have therefore been developed in recent years toward installation of these chemical handles into proteins of interest, typically in the form of the substrates of protein modification enzymes or unnatural amino acids (UAAs), which are the main focus of this Account.[3−5,7,10,13,20−23]

Enzymatic and Residue-Specific Incorporation

Enzymatic Conjugation or Conversion. In nature, additional chemistries beyond the functional groups contained in the canonical 20 amino acids are often required to carry out a protein's physiological functions. A cell uses enzyme cofactors and posttranslational modification (PTM) machineries to fulfill this requirement at the posttranslational stage.[5,22−24] Mimicking PTM, enzymatic tools have been developed to conjugate bioorthogonal groups onto proteins of interest.[5,23] The target protein is fused or inserted with a peptide tag to which an enzyme can ligate its substrate with high specificity. For protein labeling, a substrate analogue containing a bioorthogonal group is usually used, and the enzyme is chosen or engineered so that it can tolerate the substrate modification. For example, Ting and co-workers inserted a 22-amino-acid sequence that can be efficiently recognized by *Escherichia. coli* lipoic acid ligase, LplA, into target proteins. The ligase can introduce an azidoalkanoic acid in place of lipoic acid, and the ligated azide was then selectively derivatized with fluorescent probes by a bioorthogonal reaction.[25] Many other enzymes and their substrate mimics have been developed in a similar fashion, which was well reviewed previously.[5,23] It is worthwhile to mention that this strategy is versatile in that some enzymes can be engineered to directly conjugate the whole bulky labeling probes.[5,23,26]

Alternatively, a natural amino acid side chain might be directly converted by an enzymatic reaction to carry a bioorthogonal chemical group. Bertozzi and co-workers recently utilized formylglycine-generating enzyme (FGE) to convert the cysteine residue in a 13- or 6-residue consensus sequence to aldehyde containing formylglycine in bacteria and mammalian cells.[7,13] This approach eliminates the need for enzyme substrate engineering so that it might be advantageous for certain applications. Looking for additional enzymes capable of directly converting a natural amino acid side chain to an unnatural moiety will be an exciting avenue to explore in the future.

Residue-Specific Incorporation. Instead of installing the chemical groups on the residues in a protein by enzymatic modifications (conjugation or conversion), an alternative strategy involves using the cell's translational machinery to genetically encode bioorthogonal functionalities in the form of UAAs into proteins. Cell-free translation systems were first used to chemically charge UAAs to the corresponding tRNAs, which were subsequently delivered into living cells by microinjection or transfection to facilitate the *in vivo* study of protein structure and functions.[27,28] However, this method is limited by the fragile nature and the stoichiometric usage of the aminocyl-tRNAs synthesized, as well as the disruptive delivery methods in some cases. Another approach takes advantage of bacterial strains that are auxotrophic for one of the common 20 amino acids to globally replace that amino

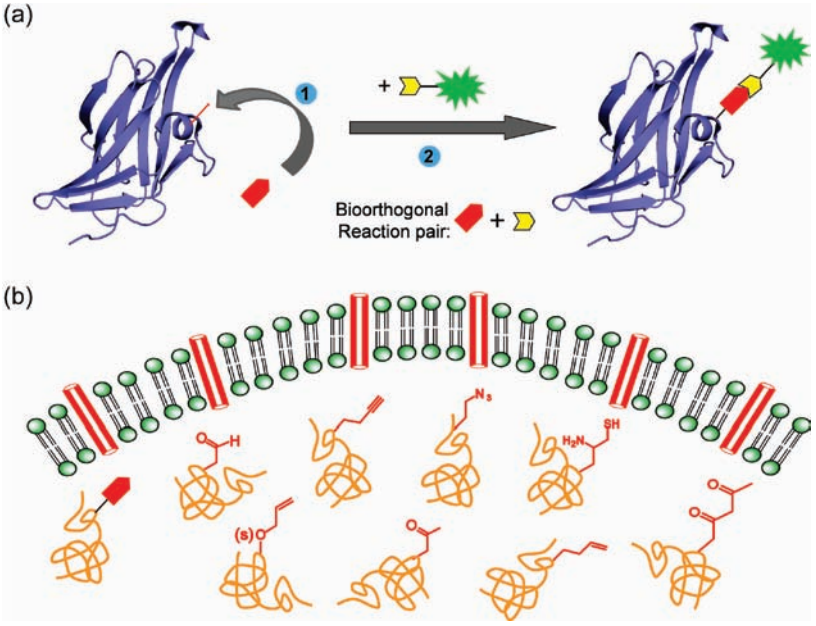


FIGURE 1. The bioorthogonal protein labeling strategy. (a) A bioorthogonal chemical group (red pentagon) is first introduced into a target protein. In a second step, the installed group is chemically conjugated with an exogenously delivered probe bearing a complementary functional group (yellow pentagon). Both groups must be non-native and nonperturbing. (b) Some examples of bioorthogonal functional groups that can be introduced into proteins include azide, alkyne, alkene, *O*-allyl, *S*-allyl, aldehyde, ketone, diketone, and N-terminal cysteine.

acid with its UAA analogue.[29] By this means, all sites of that residue in newly synthesized proteins can be installed with bioorthogonal groups including azides, alkynes, and ketones. Subsequently, the installed functional groups are reacted with imaging probes or enrichment tags linked to the complementary bioorthogonal groups[4] (also see Tirrell's review in this issue). Although this residue-specific incorporation has been very useful for protein labeling, such a "global tagging" method may interfere with the protein's structure and function. Under many circumstances, a more precise strategy, such as the site-specific introduction of bioorthogonal functionalities into proteins, is desired.

Site-Specific Incorporation

Ideally, one would like an UAA (for example, an amino acid carrying a bioorthogonal group on its side chain) to be genetically incorporated into a specific site of the protein of interest with high fidelity and efficiency. This can be accomplished by means of an orthogonal aminoacyl-tRNA synthetase (aaRS)–tRNA pair that incorporates the UAA in response to a nonsense or a unique four-base codon in the gene of interest (Figure 2).[20,30] The exogenously introduced aaRS–tRNA pair does not need to interfere with the endogenous aaRS–tRNA pairs. Moreover, this aaRS can selectively recognize an UAA and aminoacylate it onto the cognate tRNA. The "UAA-charged" tRNA then enters the ribosome and recognizes the corresponding amber codon on mRNA to transfer the UAA to the growing polypeptide chain. Directed evolution of the specificity of the aaRSs in both *E. coli* and yeast has generated nearly 70 UAAs with novel physical, chemical, or biological properties.[20] Among this list is a panel of UAAs bearing bioorthogonal chemical handles such as azide, alkyne, ketone, and alkene (Figure 1b). These UAAs with novel properties have found a wide range of biological applications, as reviewed previously.[8,20,21,31] However, it should be noted that many previously evolved aaRSs in *E. coli* were not compatible with mammalian cells, which prevented a big portion of these ~70 UAAs to be incorporated into mammalian proteins.[20,32] Therefore, it is attractive to develop a general system through which the aaRS–tRNA pairs evolved in *E. coli* can be directly used to encode UAAs in mammalian cells. To this end, an improved methodology, namely, a pyrrolysine-based facile system allowing the concurrent genetic code expansion in both prokaryotic and eukaryotic cells, has emerged and is highlighted below.

A Pyrrolysine-Based Facile System. Pyrrolysine (Pyl, **1**, Figure 3a) is the 22nd naturally occurring amino acid used by certain methanogenic archaea and a Gram-positive bacterium.[33,34] It is genetically encoded by an amber codon (UAG) and its cognate $tRNA^{Pyl}_{CUA}$ in response to the pyrrolysyl-tRNA synthetase (PylRS). The most widely used PylRSs are two close homologues from archaeal species *Methanosarcina*

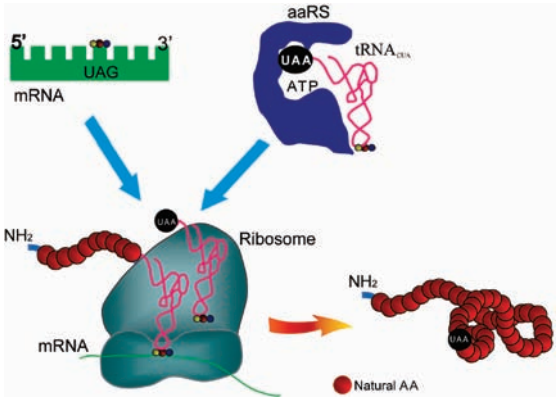


FIGURE 2. Site-specific incorporation of unnatural amino acids (UAAs) into proteins by genetic code expansion in living cells.

mazei (MmPylRS) and *Methanosarcina barkeri* (MbPylRS). Whereas PylRS could be directly used by *E. coli* translation machinery to encode Pyl and its close analogues such as N^{ε}-cyclopentyloxycarbonyl-L-lysine (Cyc, **2**, Figure 3b),[35] the $tRNA^{Pyl}_{CUA}$ was shown not to be recognized by endogenous aaRSs in *E. coli* or mammalian cells as a result of its unique structure features.[36,37] Meanwhile, the crystal structures of MmPylRS in complex with Pyl or Cyc had been solved by the Steitz group.[38] These accomplishments together laid the ground for manipulation of PylRS active sites to recognize more structurally diversified Pyl analogues. Considerable efforts have been made to explore this system for site-specific introduction of unnatural functionalities into proteins in prokaryotic as well as eukaryotic species. The Chin group evolved a MbPylRS mutant to incorporate acetyl lysine into target proteins in *E. coli,* demonstrating that the specificity of the PylRS can be altered by directed evolution.[39] The Yokoyama group utilized the wild-type MmPylRS and an engineered MmPylRS variant to encode Boc-lysine and *Z*-lysine ((N^{ε}-benzyloxycarbonyl)-L-lysine) in mammalian cells, respectively.[40] A more general strategy on utilizing the PylRS–$tRNA^{Pyl}_{CUA}$ pair for genetic code expansion in both prokaryotic and eukaryotic cells, termed a facile system, was first demonstrated by the Schultz group.[32] In their initial work done by Chen, P. R., et al., a mutant MmPylRS was evolved in *E. coli* to encode a photocaged lysine (**3**). More importantly, the evolved PylRS–$tRNA^{Pyl}_{CUA}$ pair can be directly transferred into mammalian cells to incorporate **3** into the targeted protein carrying an in-frame amber codon. Most recently, the Chin group further extended the utility of the PylRS–$tRNA^{Pyl}_{CUA}$ pair to expand the genetic code of yeast.[41] Along this process, the collection of UAAs encoded by PylRS or its mutants has also been significantly expanded (**3**–**14**, Figure 3b), mainly by either rational design or directed

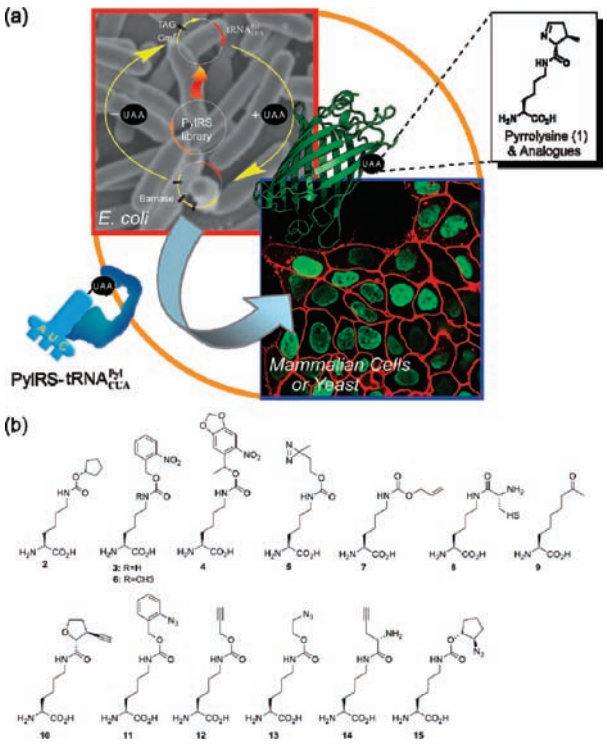


FIGURE 3. A pyrrolysine (**1**)-based facile system for genetic code expansion of *E. coli*, yeast, and mammalian cells. (a) A PylRS mutant was first evolved in *E. coli* that selectively aminoacylates $tRNA^{Pyl}_{CUA}$ with a pyrrolysine analogue. Transfer of this orthogonal PylRS–$tRNA^{Pyl}_{CUA}$ pair into mammalian cells or yeast made it possible to selectively incorporate this UAA into proteins (exemplified as GFP) in these organisms. (b) Structures of Cyc (**2**) as well as genetically encoded Pyl analogues bearing photoactive groups (**3**–**6**) or bioorthogonal groups (**7**–**15**). Besides bioorthogonal handles, UAAs carrying photoactive groups also have the potential to be used to manipulate protein's activities inside the living cells.

evolution approaches.[32,39,41–47] Notably, a panel of Pyl analogues bearing bioorthogonal handles such as ketone, azide, alkyne, alkene, or allyl groups (**7**–**14**) have been developed, which, in conjunction with the bioorthogonal reactions, provides a powerful tool for the exploration of protein's functional role in the native cellular environment. Based on these important developments, it is reasonable to envision that more PylRS mutants with new specificities will be obtained in *E. coli* by directed evolution, and the resulting PylRS mutants can be shuttled into mammalian cells and yeast for diverse applications. Therefore, the PylRS–$tRNA^{Pyl}_{CUA}$ pair can potentially become the "one-stop shop" for people who wish to introduce novel functionalities into proteins in *E. coli*, yeast, and mammalian cells (Figure 3a).

Azido and Alkynyl Pyrrolysine Analogous. In the table of the available bioorthogonal chemical handles, azide and alkynes have proved to be the most versatile ones for protein labeling, since they are small, abiotic, and stable

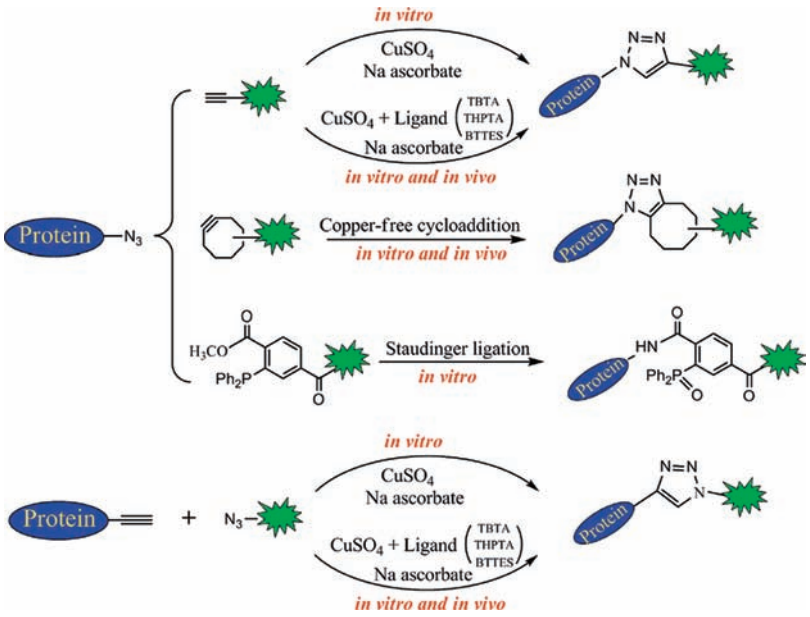


FIGURE 4. Azide and alkyne participate in bioorthogonal reactions on proteins.

in living systems. Azides can react with phosphines[14] or terminal or strain-activated alkynes,[15,19] whereas terminal alkynes can be conjugated with azides (Figure 4). Although the Cu(I)-catalyzed azide−alkyne cycloaddition (CuAAC) has exceptionally broad biological applications, the cytotoxicity of Cu(I) has been observed in both *E. coli* and mammalian cells,[48,49] which limits its use in living conditions. To address this problem, the Bertozzi group developed a copper-free click chemistry, in which the alkyne was activated by ring strain without the need for the copper catalyst.[50] Since the initial development of the strain-promoted cycloaddition, several variations of this reaction have been reported by the Bertozzi group as well as other groups.[19,51] An alternative strategy for curbing the cytotoxicity of CuAAC is to exploit Cu(I) ligands that can accelerate the reaction and reduce the copper toxicity. The TBTA ligand developed by Fokin et al. was shown to remarkably accelerate the CuAAC reaction.[16] Finn et al. developed the THPTA ligand, which was recently optimized for rapid labeling of surface glycans on living mammalian cells using CuAAC.[17] Notably, Wu and co-workers recently screened a library of 14 TBTA analogues and discovered a new ligand, BTTES, which not only showed an enhanced ability of accelerating CuAAC compared with both TBTA and THPTA but also exhibited no apparent toxicity when applied to living cells and even whole organisms.[18] With the progress toward a more biocompatible CuAAC, the terminal alkyne group shall find broader applications as a bioorthogonal handle for protein labeling in living conditions.

The development of azido- and alkynyl-Pyl analogues is an important step toward protein labeling using the aforementioned chemistries in living cells. The first alknyl-Pyl derivative (**10**) developed by Chan et al. is a direct cyclic Pyl mimic but requires a 16-step long synthesis with an overall yield of 17%.[43] The first azido-Pyl analogue (**11**) was developed by Yukoyama et al. as an aromatic derivative, whose synthesis was also demanding.[42] Chin et al. subsequently illustrated that wt-PylRS could recognize the aliphatic azido- and alkyne-Pyl analogues (**12**, **13**), which could be synthesized much more easily (in two steps with 70−80% yield).[52] Continuous efforts from the Chan group have also resulted in the second generation, aliphatic substrates of wt-PylRS carrying an alkynyl group (**14**).[47] Our efforts were to develop an azido-Pyl analogue that reserves the close cyclic mimic, which might be sterically favored in terms of recognition efficiency but does not need challenging synthesis. More importantly, we sought to exploit the azido-Pyl analogue for protein labeling in living cells. To fulfill this task, we developed a highly efficient cyclic azido-Pyl analogue (**15**, Figure 3b) as detailed in the next section.

A Highly Efficient Cyclic Azido Pyrrolysine Analogue. To develop more effective Pyl analogues, which should facilitate applications in living cells, we combined both the rational design and directed evolution approaches. Whereas directed evolution is an extremely powerful tool to select the most effective PylRS mutant for the synthesized Pyl analogue from a large randomized library, the proper design of

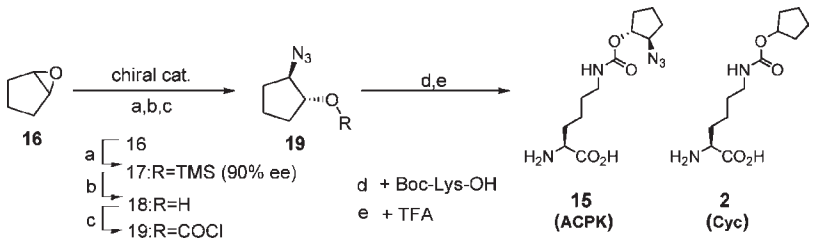


FIGURE 5. Synthesis scheme for ACPK (**15**). The structure of Cyc is shown for comparison.

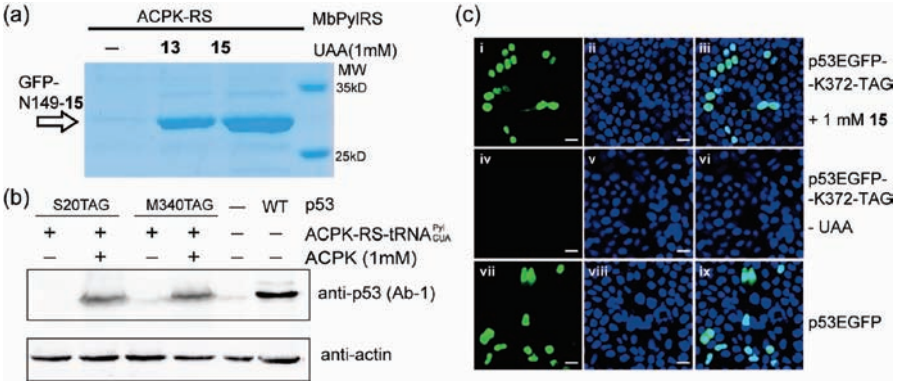


FIGURE 6. Genetic incorporation of ACPK (**15**) into proteins in *E. coli* and mammalian cells. (a) SDS–PAGE analysis demonstrating the incorporation efficiency of **13** and **15** into GFP in *E. coli*. (b) Immunoblotting analysis of ACPK incorporated p53 in H1299 cells (p53-null). (c) Confocal microscopy on the expression and localization of p53EGFP-K372-**15** in HEK293T cells harboring ACPK-RS–$tRNA^{Pyl}_{CUA}$ pair. All scale bars are 20 μm. Reproduced by permission of The Royal Society of Chemistry (RSC).

the Pyl derivatives is also crucial, because it can in turn facilitate the directed evolution process. Such a sequential process takes advantage of both approaches to optimize the final incorporation efficiency, which has been illustrated in our laboratory by the design, synthesis, and evolution of an azide-bearing Pyl mimic N^{ε}-(((1*R*,2*R*)-2-azidocyclopentyloxy)-carbonyl)-L-lysine (ACPK, **15**) exhibiting improved incorporation efficiency than the current azido-Pyl mimics.[53]

A cost-effective and easily scalable synthetic route was developed for ACPK (Figure 5).[53] The design of ACPK was inspired by the fact that various carbamate-based Pyl analogues, the cyclic Pyl mimic Cyc (**2**), in particular, had been effectively recognized by PylRS or its variants (Figure 4). This carbamate linkage could therefore serve as a general connection between N^{ε}-lysine and diverse alcohol-containing molecules including cyclic alcohols. We synthesized *trans*-1,2-azido alcohol (**18**) from an inexpensive five-membered ring-fused meso epoxide (**16**), which was subsequently coupled with Boc-Lys-OH to give the final product **15** with 70% overall yield (in five steps).

Directed evolution allowed us to identify a PylRS mutant, designated as ACPK-RS, that could produce a 2-fold excess of GFP containing **15** at residue 149 position compare with the wt-PylRS-produced GFP carrying **15** at the same amber mutation site (Figure 6a). The calculated yield of full-length GFP-N149-**15** is ~10 mg/L. The evolved ACPK-RS–$tRNA^{Pyl}_{CUA}$ pair from *E. coli* was then readily transferred into mammalian cells, and the tumor suppressor p53 fused with EGFP at C-terminus (p53EGFP) was used as a model protein to confirm the site-specific incorporation of **15** (Figure 6b). Similar to wild-type p53EGFP, the p53EGFP protein carrying ACPK at Lys372 (pEGFP-K372-15) was located within the cell nucleus, confirming that the introduction of ACPK into p53 did not disrupt its subcellular localization (Figure 6c).

Applications of Protein Bioorthogonal Labeling

Elucidating the protein dynamics and function in living cells is a main focus in the field of cell biology. Chemical labeling of proteins with bioorthogonal handles can vastly propel the progress. Since the pioneering work of incorporating a keto-containing UAA into *E. coli*-expressed proteins,[54] the combination of genetic encoded chemical handles and bioorthogonal reactions has proved to be a powerful tool. Largely, mammalian proteins carrying chemical handles have been expressed and purified in *E. coli* and subsequently tagged and studied *in vitro*.

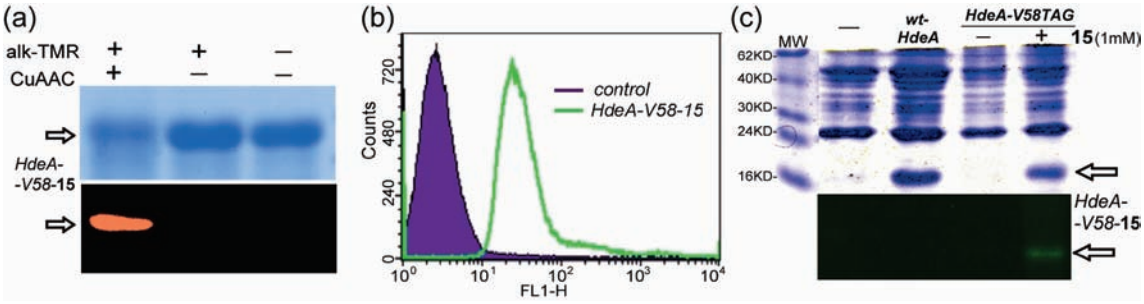


FIGURE 7. Click labeling of HdeA-V58-**15** in vitro and in living bacteria: (a) SDS–PAGE of fluorescent labeled HdeA-V58-**15** by CuAAC and alk-TMR; (b) flow cytometric analysis on biocompatible CuAAC-mediated fluorescent labeling of living *E. coli* cells expressing HdeA-V58-**15** or wt HdeA as a control; (c) SDS–PAGE analysis on lysate of *E. coli* cells expressing wt HdeA or HdeA-V58-**15** or without HdeA after being reacted with alk-coumarin by biocompatible CuAAC in living bacteria. Reproduced by permission of The Royal Society of Chemistry (RSC).

The availability of the recently developed pyrrolysine-based facile system has set the stage for such studies in living mammalian cells.

Labeling of Purified Proteins. Using advanced optical techniques, proteins tagged with fluorophores, particularly in a site-specific manner, have been studied in solution and shed light on the structural and dynamic information. For example, Chan and co-workers incorporated **10** at the Thr34 position of calmodulin (CaM), a calcium binding protein. After purification, CaM functionalized with alkyne was conjugated with azidocoumarin by CuAAC.[43] An acceptor fluorophore Fluor Alexa 488 was then attached to an engineered cysteine at position 134, which enabled the Förster resonance energy transfer (FRET) between these two fluorescent dyes. Upon addition of Ca(II) and the regulatory peptide M13, changes in FRET efficiency were observed, indicating a conformational change on CaM. We site-specifically introduced **15** into the heme recognition domain (NEAT) of a bacterial heme-transfer protein IsdA, which allowed us to conjugate DBCO-Fluor 488 (a variation of cyclooctyne-conjugated dye) near the heme binding pocket by the "copper free" click chemistry. A reversible fluorescence quenching effect was observed, caused by the direct energy transfer between heme and Fluor 488.[53] Since heme is utilized by mammalian hosts to tightly control the essential iron element as an effective defense mechanism, this site-specific "click" labeling tool can be used to study heme transfer among heme-acquisition proteins in Gram-positive pathogens.

Simultaneously incorporating two distinct UAAs into a single protein was first demonstrated by Schultz and co-workers by incorporating L-homoglutamine and *O*-methyl-L-tyrosine in response to a quadruplet codon and an amber codon, respectively.[55] Two bioorthogonal handles can also be installed on a single protein in this way. Recently, Chin and co-workers evolved an orthogonal ribosome that can decode a series of quadruplet codons.[56] Together with two mutually orthogonal aaRS–tRNA pairs (MjTyrRS-$tRNA_{CUA}$ and MbPylRS-$tRNA_{CUA}$), they demonstrated the encoding of an azide-bearing UAA and an alkyne-bearing UAA in response to two new codons created on the same orthogonal mRNA.[56] By mutating the PylRS–tRNA pair to suppress the ochre UAA codon, Liu et al. also showed that the combination of this engineered PylRS–$tRNA_{UAA}$ pair with MjTyrRS–$tRNA_{CUA}$ pair was able to genetically incorporate two distinct UAAs into a single protein in *E. coli*.[57] However, the efficiency for simultaneous incorporation of two UAAs into the same protein still waits for further improvements before it can be applied for versatile bioorthogonal protein labeling with diverse probes.

Protein Labeling in Living Cells. To understand protein function and dynamics in a physiologically relevant context requires a means to label proteins in living cells. Whereas the *in vitro* protein labeling strategy has been widely explored, expanding this methodology for protein labeling in living cells is still at its infancy. Using genetic code expansion and bioorthogonal groups for labeling of proteins in living conditions was first demonstrated on the *E. coli* surface. The acetylphenylalanine was site-specifically incorporated into an outer-membrane protein LamB followed by conjugation with fluorescein hydrazide derivatives for imaging LamB on the living *E. coli* cell surface.[58] "Click" labeling was also applied to proteins metabolically incorporated with azide or alkyne containing methionine analogues, which allowed the imaging of the newly synthesized proteins.[4] Since CuACC is not compatible with living mammalian cells, Tirrell and Bertozzi and co-workers employed the strain-promoted cycloaddition to conjugate cellular proteins carrying metabolically introduced azide handles with membrane-permeable coumarin-cyclooctynes inside the living cells.[9] Elevated

background was observed due to the reactivity of cyclooctynes with cysteine residues, which raised concerns about using this method for live-cell protein labeling.

Alternatively, we have explored the feasibility of using the recently developed BTTES-assisted CuAAC for protein labeling in living conditions.[18] As mentioned above, BTTES not only accelerates CuAAC significantly, but also it makes the reaction adaptable for living mammalian cells. We first incorporated ACPK into HdeA, a periplasmic protein in *E. coli*, to demonstrate the protein "click" labeling using ACPK as a functional handle.[53] APCK was introduced at the V58 position on HdeA followed by CuAAC with an alkyne-modified tetramethylrhodamine fluorophore *in vitro* (Figure 7a). Next, we treated the *E. coli* cells expressing HdeA-V58-**15** with an alkyne-functionalized coumarin by BTTES-assisted CuAAC. Flow cytometric analysis showed that the fluorescently labeled cells have median fluorescence intensity more than 10-fold greater than the wt *E. coli* control (Figure 7b), and no apparent toxicity was observed. The labeling was further confirmed by the SDS–PAGE analysis on cell lysates showing a single fluorescent band corresponding to the fluorescently labeled HdeA-V58-**15** (Figure 7c). Since a similar strategy has been used to label azide-bearing cell surface glycans, one in principle should be able to use ACPK in conjunction with the biocompatible click chemistry to label mammalian cell surface proteins (also refer to the next section for further discussion). It should be noted that the bioorthogonal labeling methodology has mostly been applied to label proteins on the surface of *E. coli* and mammalian cells or in the periplasmic space of *E. coli*.[53] Developing bioorthogonal reactions that can be performed inside the living cells would be of great interest in the future.

In Combination with Other Types of Biomolecules Carrying Bioorthogonal Handles. To further expand the application of protein click labeling in living cells, we aim to combine the azide or alkyne incorporated proteins with other type of biomolecules carrying complementary functional groups, which will allow the bioorthogonal conjugation between these different types of biomolecules in living cells. Our initial effort was to use ACPK as a bioorthogonal handle for chemically attaching proteins onto live-cell surfaces by conjugation with alkyne-functionalized surface glycans.[53] HEK 293T cells were incubated with 50 μM peracetylated *N*-(4-pentynoyl)mannosamine (Ac4ManNAl) to metabolically incorporate the corresponding alkynyl sialic acid (SiaNAl) into their cell-surface glycoconjugates. The resulting cells bearing alkynyl groups on the surface were reacted with 15 μM ACPK-incorporated EGFP by BTTES-assisted CuAAC. Confocal fluorescence microscopy showed robust GFP fluorescence on cell surface (Figure 8). Given the generic nature of this strategy, other proteins can also be specifically attached onto the mammalian cell surface. Further exploration of this strategy is underway in our laboratories.

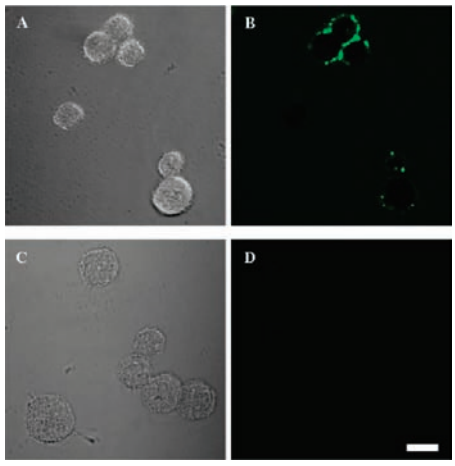


FIGURE 8. Imaging of chemical attachment of EGFP–azide on living HEK 293T cell surface. 293T cells were incubated with (A, B) or without (C, D) 50 μM Ac4ManNAl for 3 days, followed by reaction with 15 μM EGFP–azide by Cu(I)-catalyzed azide–alkyne cycloaddition (CuAAC) using BTTES ligand. Scale bar = 20 μm. Reproduced by permission of The Royal Society of Chemistry (RSC).

More Precise Labeling of Proteins. One of the major advantages for the site-specific protein labeling compared with the genetic encoded fluorescent proteins is the small size of labeling probes that can be incorporated at virtually any desired site on the protein. This feature is extremely valuable where the bulky GFP induces significant perturbations to the target protein's structure and function. Many bacterial toxins are such examples that their function or infectious ability will be significantly affected when fused with GFP tags. During the pathogen's invasion into the host cells, the pathogenic bacteria often attach themselves onto the cell surface followed by delivery of various bacterial toxins into the host cells.[59] Various strategies have therefore been evolved by bacteria to enable the successful entry of toxic effectors into host cells, which often require these proteins to be partially unfolded during the translocation followed by refolding inside the host cells. Unlike many toxins, such as anthrax lethal factor (LF), that naturally undergo this unfolding/refolding cycle, fluorescent proteins can hardly be sustained in such a process and also diminished the LF's cellular entry when fused.[60] We incorporated ACPK into LF at a nonactive site followed by conjugation

with alkyne−Fluor 488 through CuAAC. The resulting LF tagged with fluorophore exhibited similar activities to the wt LF when measured by both cellular-entry assay and the MEK-cleavage assay in living cells (P.R.C., unpublished results). This labeling strategy might permit imaging of the LF entry process in living cells, which is an ongoing project in our laboratory.

Another case where GFP fusion is not applicable is for studying the secretion process of bacterial type III secretion (T3S) effectors, largely due to the perturbed secretion efficiency by GFP tags.[61] In our laboratory, we are applying bioorthogonal chemistry for more precise labeling of these toxins such as OspF, a *Shigella flexneri* T3S effector targeting the host MAPK pathway during pathogenesis.[62] Recently, we introduced ACPK at a site that is well exposed on the surface of OspF. Secretion assay demonstrated that this ACPK-incorporated OspF could be secreted into the extracellular space with a similar efficiency to the wt-OspF. By contrast, OspF fused with GFP at the C-terminus disrupted its secretion (N-terminal GFP fusion will block the secretion signal peptide located at OspF N-terminus). Furthermore, OspF can be directly expressed in mammalian cells and incorporated with the azide handle using the pyrrolysine-based facile system (P.R.C., unpublished results). Using the azide to fluorescently label OspF for studying the secretion process is underway in our group, which might help further understand how *Shigella* utilizes this effector to manipulate host transcriptional responses.

Future Outlook

Introduction of bioorthogonal functionalities into proteins with high efficiency and specificity should open up the possibility of manipulating and monitoring protein function and dynamics in living cells. The pyrrolysine-based facile system has already demonstrated its versatility for encoding UAAs bearing chemical handles in both prokaryotic and eukaryotic cells. For example, mammalian proteins have been incorporated with azides in living cells, as mentioned above. We envision that an exciting future direction is to study the dynamics such as the conformational changes of important cell surface receptors on living cell surface. Given the promising technology developments by now, we believe it should not be a long way to go.

BIOGRAPHICAL INFORMATION

Ziyang Hao received his B.S. in Chemistry from China Agricultural University in 2008 and is currently a graduate student in Professor Peng R. Chen's group at Peking University.

Senlian Hong received his B.S. in Chemistry from Sichuan University in 2007, and he is currently a graduate student in Professor Xing Chen's group at Peking University.

Xing Chen is a Professor of Chemical Biology at Peking University. He graduated from Tsinghua University with B.S. in Chemistry in 2002 and received his Ph.D from UC Berkeley in 2007 (advisors Professor Carolyn Bertozzi and Professor Alex Zettl). From 2007 to 2010, he worked as a LSRF postdoctoral fellow in Professor Timothy Springer's group at Immune Disease Institute and Harvard Medical School.

Peng R. Chen graduated with B.S. in Chemistry from Peking University in 2002 and received his Ph.D under the guidance of Professor Chuan He at the University of Chicago in 2007. He worked as a Postdoctoral Associate in Professor Peter Schultz's group at the Scripps Research Institute between 2007 and 2009. He has been a Professor of Chemical Biology at Peking University since the summer of 2009.

This work was supported by the National Key Basic Research Foundation of China (Grant 2010CB912300), National Natural Science Foundation of China (Grants 9101300 and 20932006), and Peking University. We thank Prof. Peter Schultz for the aaRS/tRNA plasmids. We thank Prof. Carolyn Bertozzi and Prof. Peng Wu for useful reagents.

FOOTNOTES

*To whom correspondence should be addressed. E-mail addresses: pengchen@pku.edu.cn; xingchen@pku.edu.cn.

REFERENCES

1 Prescher, J. A.; Bertozzi, C. R. Chemistry in living systems. *Nat. Chem. Biol.* **2005**, *1*, 13–21.
2 Tsien, R. Y. Constructing and exploiting the fluorescent protein paintbox (Nobel Lecture). *Angew. Chem., Int. Ed.* **2009**, *48*, 5612–5626.
3 Chen, I.; Howarth, M.; Lin, W.; Ting, A. Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods.* **2005**, *2*, 99–104.
4 Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. Residue-specific incorporation of non-canonical amino acids into proteins: Recent developments and applications. *Curr. Opin. Chem. Biol.* **2010**, *14*, 774–780.
5 Sunbul, M.; Yin, J. Site specific protein labeling by enzymatic posttranslational modification. *Org. Biomol. Chem.* **2009**, *7*, 3361–3371.
6 Yoshihara, H. A.; Mahrus, S.; Wells, J. A. Tags for labeling protein N-termini with subtiligase for proteomics. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6000–6003.
7 Wu, P.; Shui, W.; Carlson, B. L.; Hu, N.; Rabuka, D.; Lee, J.; Bertozzi, C. R. 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.
8 Wu, X.; Schultz, P. G. Synthesis at the interface of chemistry and biology. *J. Am. Chem. Soc.* **2009**, *131*, 12497–12515.
9 Beatty, K. E.; Fisk, J. D.; Smart, B. P.; Lu, Y. Y.; Szychowski, J.; Hangauer, M. J.; Baskin, J. M.; Bertozzi, C. R.; Tirrell, D. A. Live-cell imaging of cellular proteins by a strain-promoted azide-alkyne cycloaddition. *ChemBioChem* **2010**, *11*, 2092–2095.
10 Lin, Y. A.; Chalker, J. M.; Davis, B. G. Olefin metathesis for site-selective protein modification. *ChemBioChem* **2009**, *10*, 959–969.
11 Lim, R.; Lin, Q. Bioorthogonal chemistry: Recent progress and future directions. *Chem. Commun.* **2010**, *46*, 1589–1600.
12 Sletten, E. M.; Bertozzi, C. R. Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chem., Int. Ed.* **2009**, *48*, 6974–6998.
13 Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R. Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* **2007**, *3*, 321–322.
14 Saxon, E.; Bertozzi, C. R. Cell surface engineering by a modified Staudinger reaction. *Science* **2000**, *287*, 2007–2010.

15 Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem., Int. Ed.* **2002**, *41*, 2596–2599.

16 Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V. Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org. Lett.* **2004**, *6*, 2853–2855.

17 Hong, V.; Steinmetz, N. F.; Manchester, M.; Finn, M. G. Labeling live cells by copper-catalyzed alkyne−azide click chemistry. *Bioconjugate Chem.* **2010**, *21*, 1912–1916.

18 Soriano Del Amo, D.; Wang, W.; Jiang, H.; Besanceney, C.; Yan, A. C.; Levy, M.; Liu, Y.; Marlow, F. L.; Wu, P. Biocompatible copper(I) catalysts for in vivo imaging of glycans. *J. Am. Chem. Soc.* **2010**, *132*, 16893–16899.

19 Jewett, J. C.; Bertozzi, C. R. Cu-free click cycloaddition reactions in chemical biology. *Chem. Soc. Rev.* **2010**, *39*, 1272–1279.

20 Liu, C. C.; Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **2010**, *79*, 413–444.

21 Wang, Q.; Parrish, A. R.; Wang, L. Expanding the genetic code for biological studies. *Chem. Biol.* **2009**, *16*, 323–336.

22 Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J., Jr. Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chem., Int. Ed.* **2005**, *44*, 7342–7372.

23 Foley, T. L.; Burkart, M. D. Site-specific protein modification: Advances and applications. *Curr. Opin. Chem. Biol.* **2007**, *11*, 12–19.

24 Lin, M. Z.; Wang, L. Selective labeling of proteins with chemical probes in living cells. *Physiology (Bethesda)* **2008**, *23*, 131–141.

25 Fernandez-Suarez, M.; Baruah, H.; Martinez-Hernandez, L.; Xie, K. T.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y. Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat. Biotechnol.* **2007**, *25*, 1483–1487.

26 Uttamapinant, C.; White, K. A.; Baruah, H.; Thompson, S.; Fernandez-Suarez, M.; Puthenveetil, S.; Ting, A. Y. A fluorophore ligase for site-specific protein labeling inside living cells. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 10914–10919.

27 Rodriguez, E. A.; Lester, H. A.; Dougherty, D. A. In vivo incorporation of multiple unnatural amino acids through nonsense and frameshift suppression. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8650–8655.

28 Kohrer, C.; Yoo, J. H.; Bennett, M.; Schaack, J.; RajBhandary, U. L. A possible approach to two different unnatural site-specific insertion of amino acids into proteins in mammalian cells via nonsense suppression. *Chem. Biol.* **2003**, *10*, 1095–1102.

29 Link, A. J.; Mock, M. L.; Tirrell, D. A. Non-canonical amino acids in protein engineering. *Curr. Opin. Biotechnol.* **2003**, *14*, 603–609.

30 Wang, L.; Schultz, P. G. Expanding the genetic code. *Angew. Chem., Int. Ed.* **2004**, *44*, 34–66.

31 Chin, J. W. Modular approaches to expanding the functions of living matter. *Nat. Chem. Biol.* **2006**, *2*, 304–311.

32 Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Geierstanger, B. H.; Schultz, P. G. A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chem., Int. Ed.* **2009**, *48*, 4052–4055.

33 Srinivasan, G.; James, C. M.; Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: Charging of a UAG-decoding specialized tRNA. *Science* **2002**, *296*, 1459–1462.

34 Hao, B.; Gong, W.; Ferguson, T. K.; James, C. M.; Krzycki, J. A.; Chan, M. K. A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **2002**, *296*, 1462–1466.

35 Polycarpo, C. R.; Herring, S.; Berube, A.; Wood, J. L.; Soll, D.; Ambrogelly, A. Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* **2006**, *580*, 6695–6700.

36 Nozawa, K.; O'Donoghue, P.; Gundllapalli, S.; Araiso, Y.; Ishitani, R.; Umehara, T.; Soll, D.; Nureki, O. Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. *Nature* **2009**, *457*, 1163–1167.

37 Polycarpo, C.; Ambrogelly, A.; Berube, A.; Winbush, S. M.; McCloskey, J. A.; Crain, P. F.; Wood, J. L.; Soll, D. An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12450–12454.

38 Kavran, J. M.; Gundliapalli, S.; O'Donoghue, P.; Englert, M.; Soell, D.; Steitz, T. A. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11268–11273.

39 Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. Genetically encoding N-epsilon-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **2008**, *4*, 232–234.

40 Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818–822.

41 Hancock, S. M.; Uprety, R.; Deiters, A.; Chin, J. W. Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *J. Am. Chem. Soc.* **2010**, *132*, 14819–14824.

42 Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N-epsilon-(o-Azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol.* **2008**, *15*, 1187–1197.

43 Fekner, T.; Li, X.; Lee, M. M.; Chan, M. K. A pyrrolysine analogue for protein click chemistry. *Angew. Chem., Int. Ed.* **2009**, *48*, 1633–1635.

44 Li, X.; Fekner, T.; Ottesen, J. J.; Chan, M. K. A pyrrolysine analogue for site-specific protein ubiquitination. *Angew. Chem., Int. Ed.* **2009**, *48*, 9184–9187.

45 Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Wang, Z. Y.; Russell, D. H.; Liu, W. S. Genetic incorporation of an aliphatic keto-containing amino acid into proteins for their site-specific modifications. *Bioorg. Med. Chem.* **2010**, *20*, 878–880.

46 Fekner, T.; Li, X.; Chan, M. K. Pyrrolysine analogs for translational incorporation into proteins. *Eur. J. Org. Chem.* **2010**, 4171–4179.

47 Li, X.; Fekner, T.; Chan, M. K. N6-(2-(R)-propargylglycyl)lysine as a clickable pyrrolysine mimic. *Chem.—Asian J.* **2010**, *5*, 1765–1769.

48 Link, A. J.; Tirrell, D. A. Cell surface labeling of Escherichia coli via copper(I)-catalyzed [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **2003**, *125*, 11164–11165.

49 Link, A. J.; Vink, M. K.; Tirrell, D. A. Presentation and detection of azide functionality in bacterial cell surface proteins. *J. Am. Chem. Soc.* **2004**, *126*, 10598–10602.

50 Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.

51 Ning, X.; Guo, J.; Wolfert, M. A.; Boons, G. J. Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast huisgen cycloadditions. *Angew. Chem., Int. Ed.* **2008**, *47*, 2253–2255.

52 Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W. Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNA(CUA) pair and click chemistry. *J. Am. Chem. Soc.* **2009**, *131*, 8720–8721.

53 Hao, Z.; Song, Y.; Lin, S.; Yang, M.; Liang, Y.; Wang, J.; Chen, P. R. A readily synthesized cyclic pyrrolysine analogue for site-specific protein "click" labeling. *Chem. Commun.* **2011**, *47* (15), 4502–4504.

54 Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G. Addition of the keto functional group to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 56–61.

55 Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G. An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7566–7571.

56 Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **2010**, *464*, 441–444.

57 Wan, W.; Huang, Y.; Wang, Z.; Russell, W. K.; Pai, P. J.; Russell, D. H.; Liu, W. R. A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. *Angew. Chem., Int. Ed.* **2010**, *49*, 3211–3214.

58 Zhang, Z.; Smith, B. A.; Wang, L.; Brock, A.; Cho, C.; Schultz, P. G. A new strategy for the site-specific modification of proteins in vivo. *Biochemistry* **2003**, *42*, 6735–6746.

59 Enninga, J.; Sansonetti, P.; Tournebize, R. Roundtrip explorations of bacterial infection: From single cells to the entire host and back. *Trends Microbiol.* **2007**, *15*, 483–490.

60 Zornetta, I.; Brandi, L.; Janowiak, B.; Dal Molin, F.; Tonello, F.; Collier, R. J.; Montecucco, C. Imaging the cell entry of the anthrax oedema and lethal toxins with fluorescent protein chimeras. *Cell. Microbiol.* **2010**, *12*, 1435–1445.

61 Enninga, J.; Mounier, J.; Sansonetti, P.; Tran Van Nhieu, G. Secretion of type III effectors into host cells in real time. *Nat. Methods* **2005**, *2*, 959–965.

62 Li, H.; Xu, H.; Zhou, Y.; Zhang, J.; Long, C.; Li, S.; Chen, S.; Zhou, J. M.; Shao, F. The phosphothreonine lyase activity of a bacterial type III effector family. *Science* **2007**, *315*, 1000–1003.