

## Genetically Encoded Copper-Free Click Chemistry\*\*

Tilman Plass, Sigrid Milles, Christine Koehler, Carsten Schultz,\* and Edward A. Lemke\*

The ability to visualize biomolecules within living specimen by engineered fluorescence tags has become a major tool in modern biotechnology and cell biology. Encoding fusion proteins with comparatively large fluorescent proteins (FPs) as originally developed by the Chalfie and Tsien groups is currently the most widely applied technique.<sup>[1]</sup> As synthetic dyes typically offer better photophysical properties than FPs, alternative strategies have been developed based on genetically encoding unique tags such as Halo and SNAP tags, which offer high specificity but are still fairly large.<sup>[2]</sup> Small tags like multi-histidine<sup>[3]</sup> or multi-cysteine motifs<sup>[4]</sup> may be used to recognize smaller fluorophores, but within the cellular environment they frequently suffer from poor specificity as their basic recognition element is built from native amino acid side chains. Such drawbacks may be overcome by utilizing bioorthogonal chemistry that relies on coupling exogenous moieties of non-biological origin under mild physiological conditions. A powerful chemistry that fulfils these requirements is the Huisgen type (3+2) cycloaddition between azides and alkynes (a form of click chemistry<sup>[5]</sup>). By utilizing supplementation-based incorporation techniques and click reactions Beatty et al. coupled azide derivatized dyes to *Escherichia coli* expressing proteins bearing linear alkynes.<sup>[6]</sup> However, this azide-alkyne cycloaddition required copper(I) as a catalyst (CuAAC), which strongly reduces biocompatibility (but see Ref. [7]). This limitation has been overcome by Bertozzi and co-workers, who showed that the “click” reaction readily proceeds when utilizing ring-strained alkynes as a substrate<sup>[8]</sup> and since then this strain-promoted azide-alkyne cycloaddition (SPAAC) has found increasing applications in labeling, for example, carbohydrates,<sup>[9]</sup> nucleotides,<sup>[10]</sup> and lipids.<sup>[11]</sup> Further expanding the potential of this approach, Ting and co-workers engineered a lipolic acid ligase which ligates a small genetically encoded recognition peptide to a cyclooctyne-containing substrate. In a second

step the incorporated cyclooctyne moiety then functioned as a specific site for labeling in cells.<sup>[12]</sup>

The direct genetic encoding of fluorescent unnatural amino acids (UAAs) has overcome many drawbacks of previous approaches by offering exquisite specificity, freedom of placement within the target protein, and minimal, if any, structural change. This approach was first achieved by Summerer et al. who evolved a leucyl tRNA/synthetase (tRNA/RS) pair from *E. coli* to genetically encode the UAA dansylalanine into *Saccharomyces cerevisiae*.<sup>[13]</sup> In response to the amber stop codon TAG, dansylalanine was readily incorporated into proteins by the host translational machinery. This approach has since been used to genetically encode several small dyes,<sup>[14]</sup> but owing to the need to evolve new tRNA/RS pairs and potential size limitations imposed by the translational machinery, larger dyes with enhanced photophysical properties have not yet been encoded.

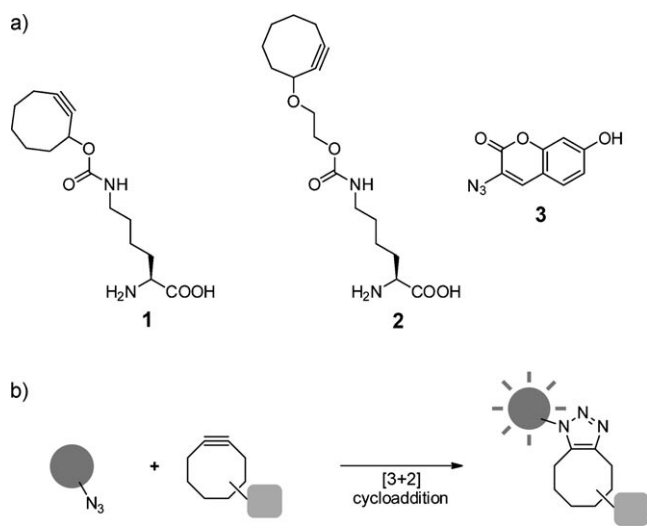
Targeted incorporation of UAAs should in principle make it possible to genetically incorporate strained alkyne and azide functional groups. In fact, a variety of azides have been genetically encoded by the use of, for example, engineered tyrosyl *Methanococcus jannaschii* tRNA<sup>Tyr</sup>/RS, leucyl *E. coli* tRNA<sup>Leu</sup>/RS, and pyrrolysine *Methanosarcina barkeri/mazei* tRNA<sup>pyl</sup>/pyIRS<sup>[15,16]</sup> pairs (Tyr = tyrosine, Leu = leucyl, pyl = pyrrolysyl). However, genetically encoding the functionality necessary for metal-free click ligations, that is, the strained alkyne, has not been achieved to date, owing to the large size of the cyclic side-chain moiety. Direct encoding of strained alkynes offers numerous advantages over encoding the azide functionality. Many commercial compounds, such as fluorescent dyes or fatty acids, are only available as the azide derivatives and thus not compatible with a copper-free click reaction with the available genetically encoded UAAs. Another major advantage of directly encoding strained alkynes is that they can react with fluorogenic azides, which are dyes whose fluorescence is dramatically quenched in the azide form and increases strongly after a successful click reaction (Scheme 1).<sup>[17]</sup> This fluorogenic approach was applied, for example, to specifically label lipids within living mammalian cells with a coumarin azide derivative, which offered high contrast due to a lack of background fluorescence.<sup>[11]</sup> The benefit of fluorogenic labels has also been highlighted in recent work of the Weissleder group who utilized the ability of tetrazines to quench fluorescence and to react with strained dienophiles in a Diels–Alder cycloaddition.<sup>[18]</sup>

To overcome the current limitations of engineering copper-free click chemistry into living cells, we aimed to genetically encode strained alkynes into *E. coli*. We reasoned that the natural amber suppressor pyrrolysine tRNA<sup>pyl</sup>/pyIRS pair from *M. mazei* might be a suitable starting point because the amino acid substrate binding pocket is partially exposed,

[\*] T. Plass, S. Milles, C. Koehler, Priv.-Doz. Dr. C. Schultz, Dr. E. A. Lemke  
Structural and Computational Biology Unit and Cell Biology and Biophysics Unit, EMBL  
Meyerhofstrasse 1, 69117 Heidelberg (Germany)  
Fax: (+49) 6221-397-536  
E-mail: schultz@embl.de  
lemke@embl.de

[\*\*] We are grateful for the help of V. Vandelinder, A. Neal, and C. Besir in the preparation of this work. We thank P. G. Schultz for materials. This study was technically supported by the EMBL-ALMF, -PEPCF, and -PCF core facilities. T.P. is a VCI and S.M. a BIF fellow. E.A.L. acknowledges funding by the Emmy Noether program and C.S. from the Transregio83 of the DFG.

Supporting information for this article (Detailed experimental methods and compound synthesis is available on the WWW under <http://dx.doi.org/10.1002/anie.201008178>).



**Scheme 1.** a) Structures of UAAs **1** and **2**, and fluorogenic dye **3**.  
 b) Reaction scheme for copper-free click reaction with fluorogenic azides.

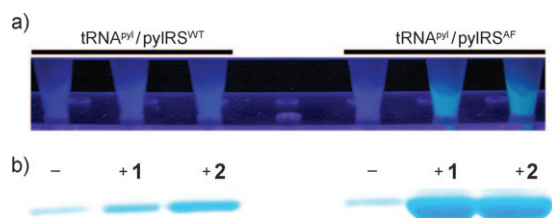
offering space to harbor a larger side chain. Furthermore, the carbamate bond at the lysine side chain was recently shown to be an important discriminator for successful incorporation of UAAs.<sup>[16,19,20]</sup> Accordingly, we synthesized lysine derivative **1** in six steps, starting from a commercially available cycloheptene precursor as outlined in Scheme S1 of the Supporting information. In the last step, the *tert*-butoxycarbonyl (Boc) protected **1** required particularly mild deprotection with 70% formic acid, as other standard conditions resulted in cleavage of the side-chain cyclooctyne moiety. The average yield per step was above 80% and allowed convenient synthesis of gram quantities. We then tested the amber suppression efficiency of the wildtype (WT) system tRNA<sup>pyl</sup>/pyIRS<sup>WT</sup> by co-expressing a green fluorescent protein (GFP) reporter construct. This construct contains an amber TAG mutation at the amino acid position 39 (GFP<sup>TAG</sup>). Hence, full-length and fluorescent GFP is only obtained if the amber stop codon is suppressed. Compared to a control, very little protein expression was observed in the presence of 1 mM **1** as shown by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) analysis and the fluorescence of the *E. coli* culture (Figure 1 and Figure S1 in the Supporting Information). We then speculated that the proximity of the large side chain to the carbamate bond inhibited efficient

substrate recognition by the synthetase. We therefore prepared amino acid **2**, featuring a longer linker between the carbamate and the cyclooctyne group. As evident from Figure 1, a minor improvement of expression yield could be detected. To further improve the acceptance of the UAA, we expanded the substrate binding pocket of pyIRS<sup>WT</sup> by introducing a Y306A mutation into the synthetase and another mutation Y384F, which was previously found by Yanagisawa et al. to support recognition of bulkier side chains.<sup>[20]</sup> As summarized in Figure 1, GFP<sup>TAG</sup> was efficiently expressed only in the presence of either **1** or **2** for this mutant tRNA<sup>pyl</sup>/pyIRS<sup>AF</sup> pair and mass spectrometry confirmed the site-specific incorporation of the UAA into the protein (Supporting Information, Table S1). As scaling up the synthesis of **1** was easier than of **2**, we focused on compound **1** for further experiments and determined that typically more than 10 mg GFP<sup>TAG-1</sup> was obtained from a 1 L culture in the presence of 1 mM **1**, demonstrating high suppression efficiency and fidelity of our tRNA<sup>pyl</sup>/pyIRS<sup>AF</sup> construct.

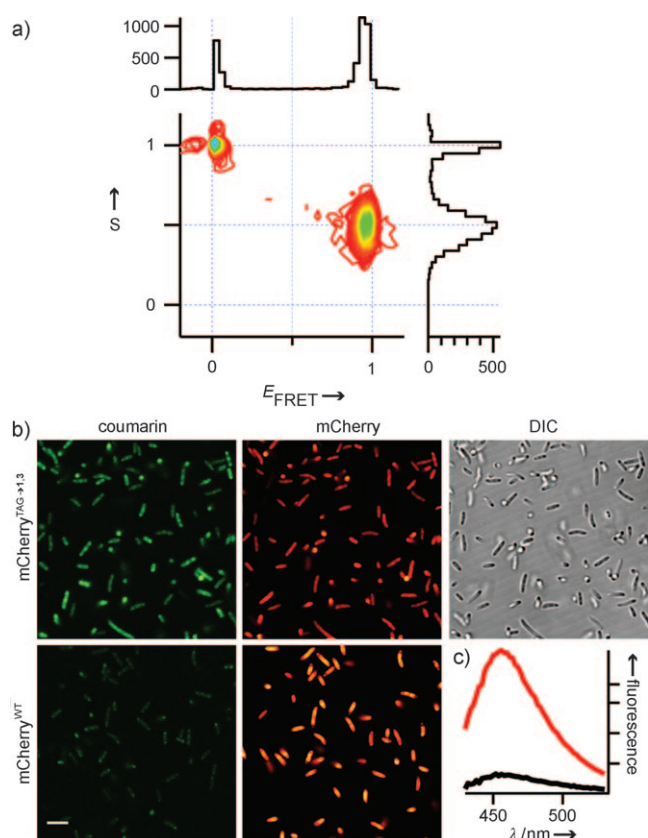
To investigate the utility of **1**, we aimed to perform single-molecule (sm) studies of labeled proteins. In particular, single-molecule observation of fluorescence resonance energy transfer (FRET) is a widely applied tool to study protein structure and dynamics. smFRET typically requires covalent, high-yielding, site-specific labeling of proteins with photostable dyes because of the low concentrations used for single-molecule studies and the high demand for overcoming the fundamental noise limit. To this end, we purified GFP<sup>TAG-1</sup> and ligated the commercially available azide derivative of Atto647N under mild conditions. Owing to the spectral properties of this fluorescent species, the natural GFP chromophore served as a donor (D) while Atto647N served as the acceptor dye (A). We then observed freely diffusing GFP<sup>TAG-1-Atto647N</sup> at single-molecule resolution using a confocal detection geometry (Supporting information, Figure S2).<sup>[21]</sup> Emission bursts stemming from individual GFP<sup>TAG-1-Atto647N</sup> molecules were analyzed based on their labeling stoichiometry (S) and for the occurrence of energy transfer ( $E_{\text{FRET}}$ ). As shown in Figure 2a, we clearly observed a species that contains both D and A molecules ( $S=0.5$ ,  $E_{\text{FRET}}=1$ ). This observed high FRET efficiency is in good agreement with the crystal structure of GFP,<sup>[22]</sup> indicating that the dye attached at position 39 is within 30 Å of the GFP chromophore.

As mentioned above, the cyclooctyne moiety is ideally suited to achieve site-specific labeling within the living cell. To verify the feasibility of such experiments with our approach, we expressed the red fluorescent protein mCherry<sup>TAG</sup> in *E. coli* in the presence of 1 mM **1** and tRNA<sup>pyl</sup>/pyIRS<sup>AF</sup>. Cells were then incubated in a 50 μM solution of the fluorogenic coumarin azide **3**.<sup>[17]</sup> Figure 2 shows that only cells expressing mCherry<sup>TAG-1,3</sup> exhibit a strong fluorescence when excited at  $\lambda=405$  nm. The fluorogenic in vivo labeling was also confirmed by fluorescence spectrometry (Figure 2c) and by analyzing the fluorescence of cell lysate using SDS PAGE (see Supporting Information Figure S3–S5, also for additional imaging experiments with other proteins).

In summary, we have genetically encoded one of the most potent functional groups for in vivo chemistry into *E. coli* and



**Figure 1.** a) Fluorescent images of *E. coli* suspensions in microcentrifuge tubes expressing GFP<sup>TAG</sup> in the absence (–) and presence (+) of **1** and **2**. b) Corresponding Coomassie stained SDS PAGE gel after purification of GFP<sup>TAG</sup> by a C-terminal affinity handle.



**Figure 2.** a) 2D histogram (S versus  $E_{\text{FRET}}$ ) of smFRET data of GFP<sup>TAG-1,Atto647N</sup> molecules. b) Cells incubated with **3** expressing tRNA<sup>pyl</sup>/pyIRS<sup>AF</sup> with mCherry<sup>TAG-1,3</sup> (plus differential interference contrast (DIC) image) and mCherry<sup>WT</sup>, imaged for fluorescence of coumarin and mCherry. c) Emission scan for coumarin of *E. coli* suspensions used in (b), mCherry<sup>TAG-1,3</sup> = red, mCherry<sup>WT</sup> = black. Scale bar = 5  $\mu\text{m}$ .

demonstrated its basic utility for in vivo labeling as well as high-resolution single-molecule measurements. SPAAC chemistry is now available to site-specifically and non-invasively modify proteins in living cells. As the tRNA<sup>pyl</sup>/pyIRS<sup>AF</sup> showed no obvious dependence on linker length (**1** versus **2**), it is conceivable that slightly altered derivatives, such as mono- and difluorinated cyclooctynes, and possibly bicyclonones,<sup>[23]</sup> could be directly used in this system. Other enhanced cyclooctynes, such as dibenzocyclooctynes,<sup>[24]</sup> could pose substantial challenges to the synthetase and/or the host translational machinery owing to their larger size. As pylRS from *M. mazei* is orthogonal in a variety of eukaryotic organisms,<sup>[19,25]</sup> we are now evaluating the transfer of this system to mammalian cells, where the technique would not only greatly expand our abilities to track proteins in living specimen but also to introduce other type of functional groups, such as cross-linkers or spin-labels for NMR spectroscopy and magnetic resonance imaging (MRI) in living specimens.

Received: December 25, 2010  
 Revised: February 3, 2011  
 Published online: March 23, 2011

**Keywords:** amino acids · click chemistry · gene expression · proteins · strained molecules

- [1] B. N. Giepmans, S. R. Adams, M. H. Ellisman, R. Y. Tsien, *Science* **2006**, *312*, 217–224.
- [2] a) A. Keppler, H. Pick, C. Arrivoli, H. Vogel, K. Johnsson, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9955–9959; b) G. V. Los et al., *ACS Chem. Biol.* **2008**, *3*, 373–382.
- [3] V. Roullier, S. Clarke, C. You, F. Pinaud, G. G. Gouzer, D. Schaible, V. Marchi-Artzner, J. Piehler, M. Dahan, *Nano Lett.* **2009**, *9*, 1228–1234.
- [4] B. A. Griffin, S. R. Adams, J. Jones, R. Y. Tsien, *Methods Enzymol.* **2000**, *327*, 565–578.
- [5] a) H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem.* **2001**, *113*, 2056–2075; *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021; b) C. P. Hackenberger, D. Schwarzer, *Angew. Chem.* **2008**, *120*, 10182–10228; *Angew. Chem. Int. Ed.* **2008**, *47*, 10030–10074; c) J. C. Jewett, C. R. Bertozzi, *Chem. Soc. Rev.* **2010**, *39*, 1272–1279; d) M. F. Debets, C. W. van der Doelen, F. P. Rutjes, F. L. van Delft, *ChemBioChem* **2010**, *11*, 1168–1184.
- [6] K. E. Beatty, F. Xie, Q. Wang, D. A. Tirrell, *J. Am. Chem. Soc.* **2005**, *127*, 14150–14151.
- [7] V. Hong, N. F. Steinmetz, M. Manchester, M. G. Finn, *Bioconjugate Chem.* **2010**, *21*, 1912–1916.
- [8] N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.
- [9] P. V. Chang, J. A. Prescher, E. M. Sletten, J. M. Baskin, I. A. Miller, N. J. Agard, A. Lo, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1821–1826.
- [10] I. Singh, F. Heaney, *Chem. Commun.* **2010**, DOI: 10.1039/c0cc03985c.
- [11] A. B. Neef, C. Schultz, *Angew. Chem.* **2009**, *121*, 1526–1529; *Angew. Chem. Int. Ed.* **2009**, *48*, 1498–1500.
- [12] C. Uttamapinant, K. A. White, H. Baruah, S. Thompson, M. Fernandez-Suarez, S. Puthenveetil, A. Y. Ting, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 10914–10919.
- [13] D. Summerer, S. Chen, N. Wu, A. Deiters, J. W. Chin, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 9785–9789.
- [14] a) J. Wang, J. Xie, P. G. Schultz, *J. Am. Chem. Soc.* **2006**, *128*, 8738–8739; b) H. S. Lee, J. Guo, E. A. Lemke, R. D. Dimla, P. G. Schultz, *J. Am. Chem. Soc.* **2009**, *131*, 12921–12923.
- [15] a) J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 9026–9027; b) J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang, P. G. Schultz, *Science* **2003**, *301*, 964–967.
- [16] D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 8720–8721.
- [17] K. Sivakumar, F. Xie, B. M. Cash, S. Long, H. N. Barnhill, Q. Wang, *Org. Lett.* **2004**, *6*, 4603–4606.
- [18] a) N. K. Devaraj, R. Upadhyay, J. B. Haun, S. A. Hilderbrand, R. Weissleder, *Angew. Chem.* **2009**, *121*, 7147–7150; *Angew. Chem. Int. Ed.* **2009**, *48*, 7013–7016; b) N. K. Devaraj, S. Hilderbrand, R. Upadhyay, R. Mazitschek, R. Weissleder, *Angew. Chem.* **2010**, *122*, 2931–2934; *Angew. Chem. Int. Ed.* **2010**, *49*, 2869–2872.
- [19] T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818–822.
- [20] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Chem. Biol.* **2008**, *15*, 1187–1197.
- [21] E. A. Lemke, Y. Gambin, V. Vandelinder, E. M. Brustad, H. W. Liu, P. G. Schultz, A. Groisman, A. A. Deniz, *J. Am. Chem. Soc.* **2009**, *131*, 13610–13612.
- [22] M. Ormö, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien, S. J. Remington, *Science* **1996**, *273*, 1392–1395.

- [23] a) J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16793–16797; b) J. Dommerholt, S. Schmidt, R. Temming, L. J. A. Hendriks, F. P. J. T. Rutjes, J. C. M. van Hest, D. J. Lefeber, P. Friedl, F. L. van Delft, *Angew. Chem.* **2010**, *122*, 9612–9615; *Angew. Chem. Int. Ed.* **2010**, *49*, 9422–9425.
- [24] X. Ning, J. Guo, M. A. Wolfert, G. J. Boons, *Angew. Chem.* **2008**, *120*, 2285–2287; *Angew. Chem. Int. Ed.* **2008**, *47*, 2253–2255.
- [25] P. R. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew. Chem.* **2009**, *121*, 4112–4115; *Angew. Chem. Int. Ed.* **2009**, *48*, 4052–4055.
-