Protein Labeling

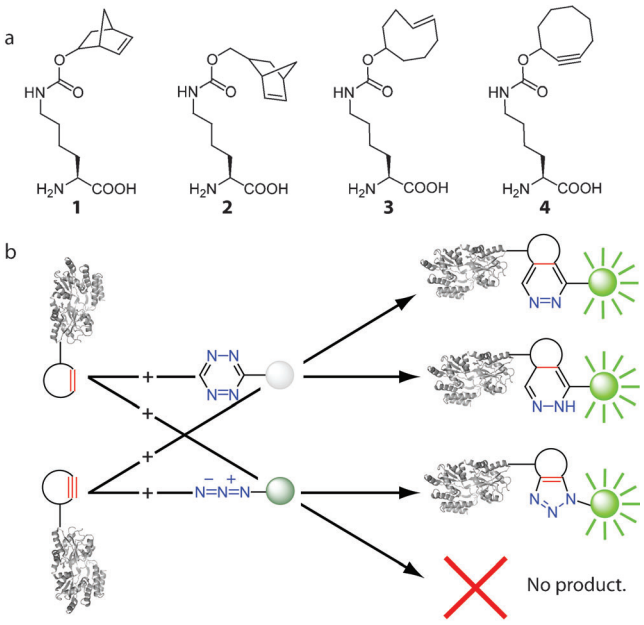
DOI: 10.1002/anie.201108231

Amino Acids for Diels–Alder Reactions in Living Cells**

Tilman Plass, Sigrid Milles, Christine Koehler, Jędrzej Szymański, Rainer Mueller, Manfred Wießler, Carsten Schultz, and Edward A. Lemke**

The endeavour to perform tailored chemical reactions in the challenging environment of the intact cell delves deeply into the biological sciences. Requirements include strict bioorthogonality of the reactants and reactions that occur spontaneously and quickly in an aqueous environment or at the interface of membranes. Commonly used reactions that meet these criteria are Staudinger ligations and various forms of click chemistry.[1] The most prominent among the latter is the Huisgen-type [3+2] cycloaddition between azides and alkynes.[1,2] Through the seminal work of the Bertozzi group, this reaction was stripped of its need for Cu^{I} catalysis by straining the alkyne group, thereby making this chemistry (termed strain-promoted alkyne–azide chemistry, SPAAC) viable in intact cells as well as in living animals.[3] These reactions have been widely used to label molecules on cell surfaces and, in a few cases, inside the cell, for instance to label lipids,[4] nucleotides,[5] or carbohydrates.[6] Another exciting click variant is strain-promoted inverse-electron-demand Diels–Alder cycloaddition (SPIEDAC), which can exhibit accelerated reaction rates by using strained reactants and furthermore is irreversible because of the loss of N_2 (Scheme 1).[7] This chemistry has been used in cells to label small molecules and is magnitudes faster than the classical Huisgen-type cycloadditions.[8]

To date, most biological applications of SPAAC or SPIEDAC do not involve modifications of proteins but instead alter cellular molecules that are not genetically encoded, such as metabolically incorporated sugars.[6] Current tools for site-specific labeling of proteins within the cell use fluorescent protein fusions,[9] self-alkylating protein additions,[10] or high-affinity binding domains.[11] The smallest size



Scheme 1. a) Structures of strained alkene and alkyne UAAs. b) Reaction scheme showing orthogonality and cross-reactivity of SPIEDAC and SPAAC with fluorogenic tetrazine-functionalized dyes (gray sphere) and azide-functionalized dyes (green sphere). Dyes coupled to tetrazine are only fluorescent (green) after successful labeling.

of artificial protein modifications currently available to introduce fluorescent labels are tetracysteine motifs consisting of six amino acids.[12] Ideally the modification unit would be only a single artificial amino acid suitable for specific chemistry in cells. The introduction of such unnatural amino acids (UAAs) is possible by codon reassignment[13] or by suppression of the Amber stop codon.[14] For fluorescent labeling, genetically encoded azides can be used, but azides typically suffer from intracellular reduction.[15] Furthermore, encoding the azide jeopardizes the design of a fluorogenic labeling scheme.[13a,16] Fluorogenicity is of particular relevance for high-contrast imaging and super-resolution techniques, since dyes are turned on only after successful labeling, while nonspecifically attached dyes remain quenched. Rather than encoding azides, a more suitable approach is the use of an amino acid that carries the strained reactant, that is, a cyclooctyne group, thereby leaving the nitrogen-bearing reactants to serve as part of a fluorogenic probe. If suppression of the Amber stop codon is used, a single residue in a specified protein can then be replaced with the strained alkyne. This type of protein labeling by using an artificially introduced cyclooctyne amino acid and fluorogenic azides

[*] T. Plass, S. Milles, C. Koehler, Dr. J. Szymański, Dr. R. Mueller, 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)
E-mail: schultz@embl.de
lemke@embl.de
Prof. Dr. M. Wießler
Biological Chemistry, DKFZ
Im Neuenheimer Feld 280, 69120 Heidelberg (Germany)

[**] We thank P. G. Schultz for a mammalian RS expression plasmid, V. VanDelinder, and A. Ori for helpful discussions. This study was technically supported by the EMBL-ALMF, -PEPCF, and -PCF core facilities. T.P. is a VCI, S.M. a BIF, and J.S. an EMBO fellow. E.A.L. acknowledges funding by the Emmy Noether program and C.S. from Transregio83 of the DFG.

Supporting information (including detailed experimental methods and compound synthesis) for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201108231.

was recently achieved in our labs.[17] Herein we expand strain-promoted click reactions to the direct labeling of proteins by using Diels–Alder reactions between strained dienophiles and tetrazines serving as dienes (Scheme 1). Fluorophores can be made intrinsically fluorogenic by coupling them to the strongly quenching tetrazine.[8b] This quenching is not easily addressed, when the tetrazines are genetically encoded.[18] We synthesized a set of three new dienophilic amino acids and incorporated them into proteins in *E. coli* and in mammalian cells through suppression of the Amber stop codon.

In our previous work, we genetically encoded UAA **4** by means of an engineered tRNA/synthetase pair from *Methanosarcina mazei* that naturally encodes pyrrolysine.[17,19] While the wildtype synthetase pair ($tRNA^{Pyl}$/$PylRS^{WT}$, here termed RS^{WT}) was not efficient in recognizing the unnatural amino acid **4** as a substrate, a rationally designed double mutant (Y306A, Y384F; here termed $tRNA^{Pyl}$/$PylRS^{AF}$, RS^{AF}) efficiently incorporated **4** in response to the Amber codon.[17] We now aimed to explore the possible promiscuity of double mutant RS^{AF} to other unnatural amino acids with similarly bulky side chains. To objectively evaluate our new compounds, we also included unnatural amino acid **4** as a standard for which we reported an expression yield of about 10 mg $GFP^{TAG\rightarrow 4}$ per 1 L of *E. coli* culture (GFP = green fluorescent protein).[17] We first synthesized the norbornene derivatives, compounds **1** and **2**, because they are known dienophiles for SPIEDAC.[7d,20] Unnatural amino acids **1** and **2** were prepared in two steps starting from the commercially available alcohol via a chloroformate intermediate followed by coupling to *tert*-butoxycarbonyl (Boc)-protected lysine and subsequent deprotection using 60% formic acid, which prevents cleavage of the sidechain carbamate. Both compounds were obtained in an overall yield of more than 91%. We then tested the Amber suppression efficiency of the RS^{WT} and RS^{AF} systems by co-expressing a GFP reporter construct with an Amber TAG mutation (GFP^{TAG}). GFP fluorescence occurred only in response to efficient Amber suppression. Full-length GFP protein was purified using a C-terminal purification handle. Figure 1 shows that norbonene UAAs **1** and **2** are efficiently recognized by the RS^{AF} mutant (typical GFP yields from a 1 L *E. coli* culture are 10 mg for **1** and **2**; Table S1 in the Supporting Information lists the confirming results from masspectrometric determination).

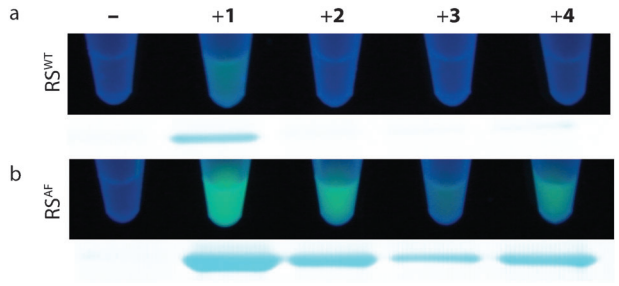


Figure 1. Fluorescence images of microcentrifuge tubes with *E. coli* suspensions expressing $GFP^{TAG\rightarrow UAA}$ in the absence (–) and presence (+) of UAAs **1**–**4** with the corresponding Coomassie-stained SDS PAGE gel after purification of $GFP^{TAG\rightarrow UAA}$ for RS^{WT} (a) and RS^{AF} (b).

Recently it was found that the *trans*-cyclooctene derivatives have much faster reaction kinetics for SPIEDAC than norbornene derivatives.[8a,b] Fast but highly specific reaction kinetics are desirable for labeling studies in vivo. Owing to the close similarity to **4**, we thus aimed to genetically encode the *trans*-cyclooctene UAA **3**. The exploration of the promiscuity of double-mutant synthetase pair RS^{AF} gave typical expression yields of 4 mg $GFP^{TAG\rightarrow 3}$ from 1 L *E. coli* culture. Compound **3** can be synthesized in five steps starting from commercially available 9-oxabicyclo[6.1.0]non-4-ene with an overall yield of 38% (average yield per step above 83%). One key step in the preparation of gram quantities of **3** is the coupling of the *trans*-cyclooctenyl-4-nitrophenyl carbonate to fluorenylmethyloxycarbonyl (Fmoc) lysine in dimethyl sulfoxide and *N*,*N*-diisopropylethylamine, because other standard conditions that were used to generate a carbamate bond were inefficient in yielding pure Fmoc-protected **3**. Subsequent deprotection was carried out using 20% piperidine in *N*,*N*-dimethylformamide. Attempts to synthesize compound **3** by using the well-established Boc synthetic strategy resulted in poorer yields because of less-efficient formation of the carbamate linkage and the lack of suitable Boc-removal conditions that neither catalyze double-bond isomerization nor cleave the sidechain carbamate.

Summarizing our expression results, we find that UAAs **1**, **2**, and **3** are efficiently incorporated by the double-mutant synthetase pair RS^{AF}, while **1** is also accepted with reduced efficiency by the wildtype synthethase pair RS^{WT} (Figure 1; for masspectrometric data see Table S1 in the Supporting Information). We next aimed to further explore the potential of UAAs **1**, **2**, and **3** for in vivo labeling, also in comparison to the reactivity of **4**. *E. coli* cultures expressing $GFP^{TAG\rightarrow UAA}$ were supplemented with 50 μM dye, washed, and then analyzed using SDS PAGE (masspectrometric conformation given in Table S1 in the Supporting Information). Figure S1 in the Supporting Information shows that in line with Scheme 1, UAAs **1**–**4** all reacted with tetrazines but only **4** also with azides. To further test if SPIEDAC and SPAAC reactions were orthogonal, we grew two cultures expressing $MBP^{TAG\rightarrow 3}$ (maltose-binding protein) and $MBP^{TAG\rightarrow 4}$ separately and then mixed them prior to labeling with tetramethylrhodamine (TAMRA) azide, which is predicted not to react with dienophiles (see Scheme 1). Finally the mixed culture was reacted with coumarin tetrazine. Figure 2a shows selectively labeled green- and red-fluorescent *E. coli* (see Figures S1 and S2 in the Supporting Information for additional controls).

To quantitatively analyze the different reactions, *E. coli* lysate was incubated with TAMRA tetrazine or TAMRA azide. The reaction was followed by using a fluorescence spectrometer by exciting GFP (Figure 2). Initially, only green fluorescence was observed. Upon reaction with the TAMRA compounds, orange fluorescence appeared owing to Foerster resonance energy transfer (FRET) from GFP to TAMRA in agreement with the close proximity after covalent coupling of TAMRA to GFP.[17] Figure 2c summarizes the data from all experiments, and Figure S3 and Table S2 in the Supporting Information list the observed reaction kinetics. For the SPIEDAC reaction of UAA **3** with TAMRA tetrazine we found a rate constant of $k = (35000 \pm 3000)\,M^{-1}s^{-1}$, which is

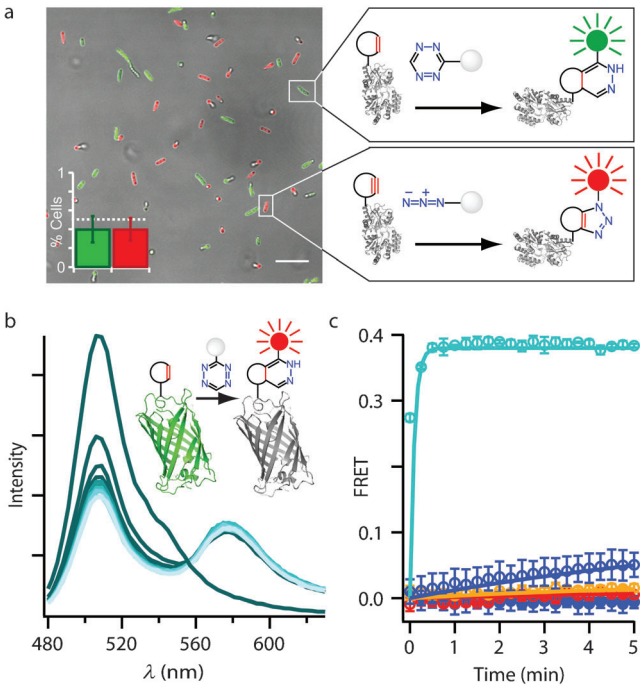


Figure 2. a) Separate *E. coli* cultures expressing MBP$^{TAG\rightarrow3}$ and MBP$^{TAG\rightarrow4}$ were mixed in a 1:1 ratio and reacted first with TAMRA azide (red) and then with coumarin tetrazine (green). The overlay of the two color channels plus the differential interference contrast (DIC) channel visualizes differential labeling, and the inlay shows that 50% of cells were labeled red or green. Scale bar 10 μm. b) GFP$^{TAG\rightarrow3}$ was incubated with TAMRA tetrazine, and the increase of TAMRA fluorescence that is due to FRET (i.e. due to labeling) was followed over time (dark to light blue). c) Summary for reaction of GFP$^{TAG\rightarrow UAA}$: red ○ = **1**, orange ○ = **2**, cyan ○ = **3**, blue ○ = **4** with TAMRA tetrazine and blue ● = **4** with TAMRA azide.

orders of magnitude faster than the next fastest reaction of **4** with TAMRA tetrazine ($k = (400 \pm 200)\,M^{-1}s^{-1}$). It should thus be mentioned that while the reactivity of strained alkynes towards the tetrazine is tuneable by tetrazine substituents,[21] this reactivity is generally much higher than that of the norbornene-bearing amino acids, of which UAA **1** has recently (note added in proof) also been encoded by the Chin lab.[22]

The ultimate goal is that bioorthogonal chemical methods can be used to site-specifically label proteins with small photostable fluorophores in mammalian cells, thereby bypassing the need for bulky fluorescent proteins that typically exhibit less-favorable photophysical properties with respect to photostability and quantum yield. To achieve this goal, one important milestone is to genetically encode unnatural amino acids for the bioorthogonal reactions SPAAC or SPIEDAC in mammalian cells, while another goal is to make small, photostable dyes coupled to tetrazine or azide that are membrane-permeant and to wash out unreacted dye from intact cells. To test if we can achieve the first milestone, we transferred tRNA^{Pyl}/PylRS pairs (RS^{WT} and RS^{AF}) to a mammalian expression plasmid, following a published procedure.[23] The tRNA^{Pyl}/PylRS pair was co-transfected with an expression plasmid of a nuclear localization signal (NLS)-mCherry-GFP^{TAG} fusion protein. In this construct, mCherry expression was visible by orange fluorescence in the nucleus, thereby signifying successful transfection of the plasmid, while green fluorescence reported on successful expression of GFP owing to Amber suppression. By using an automated microscope procedure (see Figures S4–S9 in the Supporting Information), we analyzed the GFP expression depending on the concentration of unnatural amino acids **1**–**4** for RS^{WT} and RS^{AF}. Figure 3 summarizes the data, showing for all compounds a clear dependence of the GFP fluorescence on the UAA concentration; as expected, the fluorescence intensity is low and variable at concentrations below 100 μM. GFP expression typically showed an optimum around 250 μM UAA and then frequently decreased again, which is probably due to mild toxicity of either too efficient Amber suppression or the amino acids themselves. Our systematic concentration screen revealed a way to optimize Amber suppression in mammalian cells, and we are now investigating how Amber suppression in mammalian culture can be further improved. Notably, the selectivity that we observed in *E. coli* is qualitatively conserved in mammalian cells.

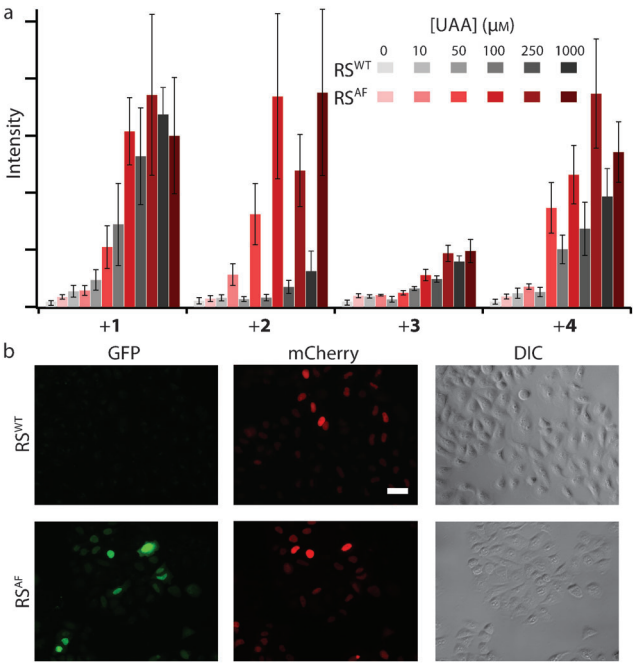


Figure 3. a) UAA-concentration-dependent expression of an NLS-mCherry-GFP^{TAG} construct in HeLa cells expressing either RS^{WT} or RS^{AF}. b) Corresponding representative images (for **2**) in the green, red, and DIC channel. Scale bar 50 μm.

Having now a system at hand that allows for site-specific encoding of UAAs for SPIEDAC and SPAAC into mammalian cells, we aimed to test if these reactions can be used for labeling. We expressed an NLS-MBP-GFP^{TAG} construct with UAAs **1**–**4** and then labeled the culture for only five minutes with 100 nM red fluorescent Cy5 tetrazine. Figure 4 shows that only cells expressing NLS-MBP-GFP$^{TAG\rightarrow3}$ show GFP fluo-

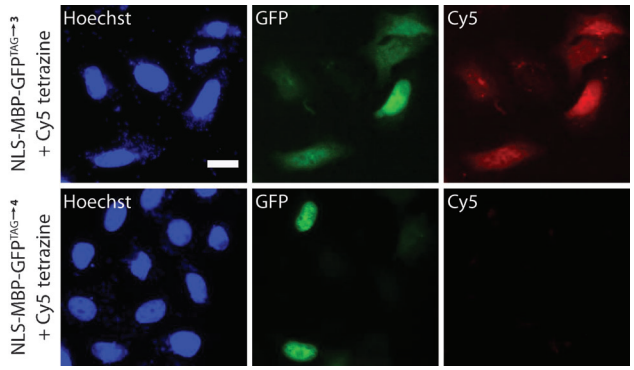


Figure 4. HeLa cells expressing NLS-MBP-GFP$^{TAG\rightarrow3}$ (upper row) and HeLa cells expressing NLS-MBP-GFP$^{TAG\rightarrow4}$ (lower row) were incubated for five minutes with Cy5 tetrazine (100 nM). The corresponding images of nuclear fluorescence from Hoechst staining in blue and GFP in green are shown. Only when UAA **3** is used, also red colocalizing Cy5 fluorescence is observed. Scale bar 20 μm.

rescence colocalized with the red fluorescence of Cy5 (see Figure S10 in the Supporting Information for all other UAAs and details on the labeling procedure). This experiment demonstrates the advantage of using much faster reactions for cell labeling despite the lower expression yield of UAAs **3** versus UAAs **1**, **2**, and **4**. This remarkable result lays a strong foundation for future use of our now faster and more diverse set of bioorthogonal chemical reactions to label and manipulate proteins with single-residue precision in mammalian cell cultures.

In summary, we have genetically encoded UAAs for a biocompatible chemistry in living cells that is orthogonal to the previously described cyclooctyne–azide click chemistry. The basic utility of this new method for labeling of proteins in vivo was demonstrated in *E. coli*. A particular focus of future work will be the development of more hydrophilic UAAs that facilitate easy washout of unincorporated UAA and thus high-contrast imaging. However, we were already able to show the high potential of genetically encoding **3** for fluorescence imaging of specifically labeled proteins in mammalian cell culture. This approach will pave the way for labeling single protein sites with small fluorophores and other analytical or functional labels in physiologically relevant cells in the future.

Received: November 22, 2011
Revised: February 17, 2012
Published online: March 30, 2012

Keywords: amber suppression · click chemistry · Diels–Alder reactions · protein engineering · unnatural amino acids

[1] C. P. Hackenberger, D. Schwarzer, *Angew. Chem.* **2008**, *120*, 10182; *Angew. Chem. Int. Ed.* **2008**, *47*, 10030.
[2] a) H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem.* **2001**, *113*, 2056; *Angew. Chem. Int. Ed.* **2001**, *40*, 2004; b) J. C. van Hest, F. L. van Delft, *ChemBioChem* **2011**, *12*, 1309; c) C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057.
[3] a) S. T. Laughlin, J. M. Baskin, S. L. Amacher, C. R. Bertozzi, *Science* **2008**, *320*, 664; b) P. Kele, X. Li, M. Link, K. Nagy, A. Herner, K. Lorincz, S. Beni, O. S. Wolfbeis, *Org. Biomol. Chem.* **2009**, *7*, 3486.
[4] A. B. Neef, C. Schultz, *Angew. Chem.* **2009**, *121*, 1526; *Angew. Chem. Int. Ed.* **2009**, *48*, 1498.
[5] a) A. Salic, T. J. Mitchison, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2415; b) P. M. Gramlich, C. T. Wirges, A. Manetto, T. Carell, *Angew. Chem.* **2008**, *120*, 8478; *Angew. Chem. Int. Ed.* **2008**, *47*, 8350.
[6] 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.
[7] a) M. L. Blackman, M. Royzen, J. M. Fox, *J. Am. Chem. Soc.* **2008**, *130*, 13518; b) M. Wiessler, W. Waldeck, C. Kliem, R. Pipkorn, K. Braun, *Int. J. Med. Sci.* **2009**, *7*, 19; c) J. Schoch, M. Wiessler, A. Jaschke, *J. Am. Chem. Soc.* **2010**, *132*, 8846; d) J. Sauer, A. Mielert, D. Lang, D. Peter, *Chem. Ber.* **1965**, *98*, 1435.
[8] a) N. K. Devaraj, R. Upadhyay, J. B. Haun, S. A. Hilderbrand, R. Weissleder, *Angew. Chem.* **2009**, *121*, 7147; *Angew. Chem. Int. Ed.* **2009**, *48*, 7013; b) N. K. Devaraj, S. Hilderbrand, R. Upadhyay, R. Mazitschek, R. Weissleder, *Angew. Chem.* **2010**, *122*, 2931; *Angew. Chem. Int. Ed.* **2010**, *49*, 2869; c) R. Rossin, P. R. Verkerk, S. M. van den Bosch, R. C. Vulders, I. Verel, J. Lub, M. S. Robillard, *Angew. Chem.* **2010**, *122*, 3447; *Angew. Chem. Int. Ed.* **2010**, *49*, 3375.
[9] B. N. Giepmans, S. R. Adams, M. H. Ellisman, R. Y. Tsien, *Science* **2006**, *312*, 217.
[10] a) A. Keppler, H. Pick, C. Arrivoli, H. Vogel, K. Johnsson, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9955; b) S. Mizukami, S. Watanabe, Y. Hori, K. Kikuchi, *J. Am. Chem. Soc.* **2009**, *131*, 5016.
[11] R. Wombacher, M. Heidbreder, S. van de Linde, M. P. Sheetz, M. Heilemann, V. W. Cornish, M. Sauer, *Nat. Methods* **2010**, *7*, 717.
[12] a) B. A. Griffin, S. R. Adams, J. Jones, R. Y. Tsien, *Methods Enzymol.* **2000**, *327*, 565; b) A. Rutkowska, C. H. Haering, C. Schultz, *Angew. Chem.* **2011**, *123*, 12867; *Angew. Chem. Int. Ed.* **2011**, *50*, 12655.
[13] a) K. E. Beatty, F. Xie, Q. Wang, D. A. Tirrell, *J. Am. Chem. Soc.* **2005**, *127*, 14150; b) J. K. Montclare, D. A. Tirrell, *Angew. Chem.* **2006**, *118*, 4630; *Angew. Chem. Int. Ed.* **2006**, *45*, 4518.
[14] a) A. Deiters, P. G. Schultz, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1521; b) C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413.
[15] a) J. V. Staros, H. Bayley, D. N. Standring, J. R. Knowles, *Biochem. Biophys. Res. Commun.* **1978**, *80*, 568; b) S. Milles, S. Tyagi, N. Banterle, C. Koehler, V. Vandelinder, T. Plass, A. P. Neal, E. A. Lemke, *J. Am. Chem. Soc.* **2012**, *134*, 5187.
[16] a) Z. Hao, Y. Song, S. Lin, M. Yang, Y. Liang, J. Wang, P. R. Chen, *Chem. Commun.* **2011**, *47*, 4502; b) D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 8720; c) J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 9026; d) T. Fekner, X. Li, M. M. Lee, M. K. Chan, *Angew. Chem.* **2009**, *121*, 1661; *Angew. Chem. Int. Ed.* **2009**, *48*, 1633; e) T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Chem. Biol.* **2008**, *15*, 1187; f) J. T. Ngo, J. A. Champion, A. Mahdavi, I. C. Tanrikulu, K. E. Beatty, R. E. Connor, T. H. Yoo, D. C. Dieterich, E. M. Schuman, D. A. Tirrell, *Nat. Chem. Biol.* **2009**, *5*, 715.
[17] T. Plass, S. Milles, C. Koehler, C. Schultz, E. A. Lemke, *Angew. Chem.* **2011**, *123*, 3964; *Angew. Chem. Int. Ed.* **2011**, *50*, 3878.
[18] J. L. Seitchik, J. C. Peeler, M. T. Taylor, M. L. Blackman, T. W. Rhoads, R. B. Cooley, C. Refakis, J. M. Fox, R. A. Mehl, *J. Am. Chem. Soc.* **2012**, *134*, 2898.

[19] H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232.
[20] N. K. Devaraj, R. Weissleder, S. A. Hilderbrand, *Bioconjugate Chem.* **2008**, *19*, 2297.
[21] M. R. Karver, R. Weissleder, S. A. Hilderbrand, *Angew. Chem.* **2012**, *124*, 944; *Angew. Chem. Int. Ed.* **2012**, *51*, 920.
[22] K. Lang, L. Davis, J. Torres-Kolbus, C. Chou, A. Deiters, J. W. Chin, *Nat. Chem.* **2012**, DOI:10.1038/NCHEM.1250.
[23] P. R. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew. Chem.* **2009**, *121*, 4112; *Angew. Chem. Int. Ed.* **2009**, *48*, 4052.

