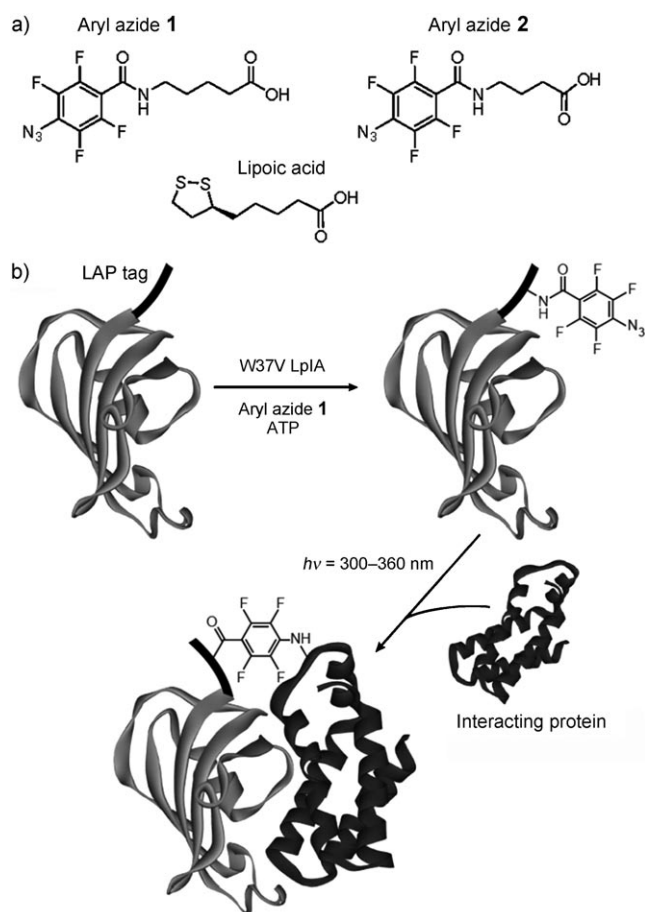


# An Engineered Aryl Azide Ligase for Site-Specific Mapping of Protein–Protein Interactions through Photo-Cross-Linking\*\*

Hemanta Baruah, Sujiet Puthenveetil, Yoon-Aa Choi, Samit Shah, and Alice Y. Ting\*

Protein–protein interactions (PPIs) underlie all signaling networks in cells, and thus, technologies for identifying and characterizing them are extremely important. One powerful approach to PPI detection consists of photo-cross-linking.<sup>[1,2]</sup> A photo-cross-linker probe such as an aryl azide (Figure 1 a) or benzophenone is conjugated to a protein of interest, and irradiation with UV light generates a reactive species that covalently cross-links to any nearby protein (Figure 1 b). Purification of the cross-linked complexes and analysis by mass spectrometry, Western blotting, or other techniques reveal the identities of the interaction partners of the protein of interest.

PPI detection by photo-cross-linking has considerable advantages over other more commonly used PPI detection methods, such as the yeast two-hybrid technique,<sup>[3]</sup> protein complementation assays (based on green fluorescent protein,<sup>[4]</sup> split  $\beta$ -lactamase,<sup>[5]</sup> and split luciferase),<sup>[6,7]</sup> and fluorescence resonance energy transfer (FRET).<sup>[8]</sup> For example, photo-cross-linking can detect interactions with endogenous proteins, whereas other methods detect only interactions with recombinant reporter-fused proteins. The small size of typical photo-cross-linking probes minimizes steric interference, which reduces the frequency of false negatives. Transient interactions can be captured. Finally, photo-cross-linking is compatible with all cell types and all subcellular compartments, in contrast to the yeast two-hybrid method, for example.<sup>[9]</sup> Yet photo-cross-linking is rarely used in living cells for PPI detection, principally because chemical photo-cross-linkers cannot be easily targeted to specific proteins of interest within the cellular context. The best current method for photo-cross-linker introduction is unnatural amino acid mutagenesis.<sup>[10–17]</sup> The non-position-specific form of this technique<sup>[16,17]</sup> introduces high background, reduced signals, and potential interference with protein function. Position-



**Figure 1.** Photoaffinity probes and photo-cross-linking scheme. a) Structures of aryl azide probes 1 and 2 and of lipoic acid. b) Illustration of site-specific aryl azide 1 ligation to a LAP fusion protein, followed by photo-cross-linking to an interacting protein. ATP: adenosine triphosphate, LplA: lipoic acid ligase, LAP: LplA acceptor peptide.

[\*] Dr. H. Baruah, Dr. S. Puthenveetil, Y.-A. Choi, Dr. S. Shah, Prof. A. Y. Ting  
Department of Chemistry, Room 18-496  
Massachusetts Institute of Technology  
Cambridge, MA 02139 (USA)  
Fax: (+1) 617-253-7929  
E-mail: ating@mit.edu

[\*\*] This work was supported by the US National Institutes of Health (grant nos.: R01 GM072670 and P20 GM072029), the Sloan Foundation, and the Dreyfus Foundation. Samsung Corporation provided a Lee Kun Hee scholarship to Y.-A.C. We thank Susan M. Lydon for her assistance with graphics. We thank Kathleen T. Xie and Jackie Chan for help with site-directed mutagenesis experiments and Marta Fernández-Suárez for assistance and advice.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.200802088>.

specific unnatural amino acid mutagenesis<sup>[10–15]</sup> is highly powerful but limited for mammalian cell applications by the prevalence of natural amber codons,<sup>[18]</sup> low suppression efficiencies,<sup>[15,19]</sup> and the generation of prematurely truncated protein products, which can sometimes produce dominant negative effects.<sup>[20]</sup> A simple, robust, high-efficiency, and minimally invasive method for photo-cross-linker targeting in cells is required, to bring the advantages of this approach for PPI detection to cell biology.

As a first step towards this goal, we report herein the design and characterization of an aryl azide ligase. To produce this ligase, we re-engineered the active site of the *Escherichia coli* enzyme lipoic acid ligase (LplA)<sup>[21]</sup> to accept a fluorinated aryl azide photoaffinity probe and catalyze its covalent

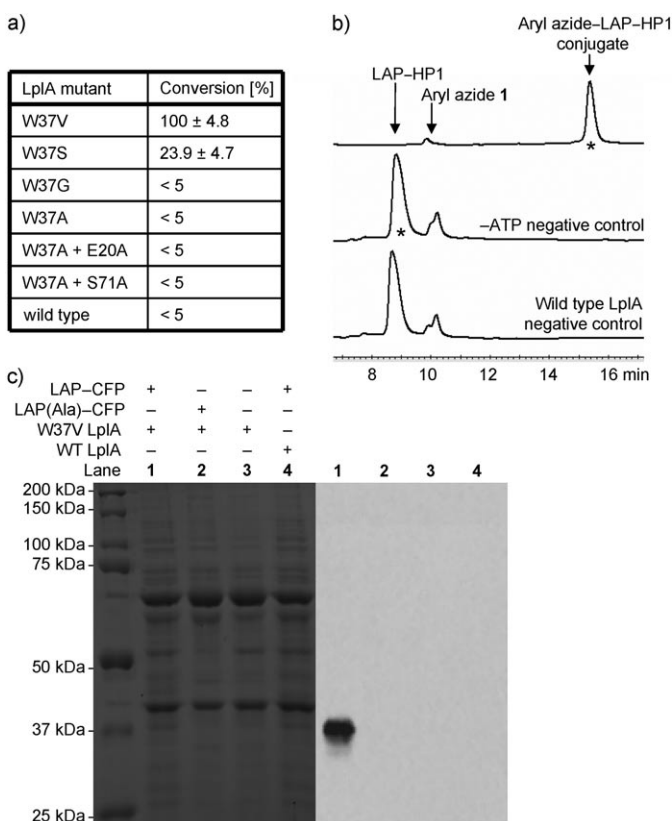
conjugation to a recognition peptide (Figure 1b). The ligation reaction is extremely sequence specific; endogenous mammalian proteins are not recognized and modified. We demonstrate the utility of this method by performing site-specific aryl azide ligation to FK506 binding protein (FKBP) in mammalian cell lysate and then showing rapamycin-dependent photo-cross-linking to the protein's interaction partner, FKBP-rapamycin binding domain (FRB), also within lysate.

*E. coli* LplA is a 38 kDa enzyme that catalyzes the ATP-dependent ligation of the cofactor lipoic acid (Figure 1a) onto a specific lysine side chain of one of its three endogenous protein substrates in *E. coli*<sup>[21,22]</sup> or onto a 17-amino acid unnatural peptide called "LplA acceptor peptide" (LAP).<sup>[23]</sup> We previously harnessed LplA for site-specific fluorophore tagging of LAP fusion proteins on the surface of living mammalian cells<sup>[23]</sup> by capitalizing on 1) LplA's extremely high sequence specificity, which prevents it from recognizing and modifying any endogenous mammalian proteins and 2) the malleability of LplA's active site, which allows it to accept and ligate alkyl azide analogues of lipoic acid. Azide-derivatized cell-surface LAPs were selectively imaged with cyclooctyne-fluorophore conjugates.<sup>[23]</sup>

For site-specific photo-cross-linker introduction, we synthesized two fluorinated aryl azides as candidate substrates for LplA (Figure 1a; syntheses shown in Figure 1 in the Supporting Information). These structures represent a compromise between high photo-cross-linking efficiency,<sup>[1,2]</sup> longer activation wavelength, and small probe size to fit better within the LplA active site. Using an HPLC assay<sup>[23]</sup> with either E2p (a hybrid lipoyl domain derived from one of LplA's natural protein substrates<sup>[22]</sup>) or LAP-HP1 fusion<sup>[23]</sup> as the protein substrate, we tested the ability of wild-type LplA to utilize probes 1 and 2. No product conjugate was detected in the reactions with either probe (data not shown).

Aryl azide probes 1 and 2 are both considerably larger than lipoic acid or the alkyl azide used in our previous study,<sup>[23]</sup> so we attempted to use mutagenesis to increase the volume of the LplA active site. The cocrystal structure of *E. coli* LplA with lipoic acid has been reported,<sup>[24]</sup> but a number of unusual features about this structure led us to examine instead the cocrystal structure of *Thermoplasma acidophilum* LplA with lipoyl-adenosine monophosphate<sup>[25]</sup> (Figure 2 in the Supporting Information). The structure shows nine amino acid side chains (L18, D21, Y39, R72, Y73, T74, H81, L159, and H161) that lie within 7.5 Å of the dithiolane ring of lipoic acid. We aligned the *E. coli* and *T. acidophilum* LplA protein sequences and individually mutated each of the corresponding amino acids in the *E. coli* enzyme (L17, E20, W37, R70, S71, S72, H79, F147, and H149) into alanine in order to decrease their size. HPLC testing of this initial panel of nine LplA mutants showed that W37A, E20A, and S71A mutants could accept probe 1 to a small degree (Figure 3 in the Supporting Information). Under identical reaction conditions, W37A gave more product than either E20A or S71A (Figure 3 in the Supporting Information). No product was detected for probe 2 with any of the mutants, despite the fact that it is smaller than probe 1; this indicates that factors other than size play a role in substrate recognition. Based on these

hits, we then produced a second panel of LplA mutants, shown in Figure 2a. We tested each of these using the same HPLC assay, with LAP-HP1, and found that W37V LplA exhibited the highest ligation activity for probe 1, followed by W37S (Figure 2a). Under identical conditions (saturating probe 1, ATP, and LAP-HP1, 30 min), W37V gave 20-fold more product than did W37A, the best enzyme from the original mutant panel (Figure 2a). It is interesting that W37V LplA is superior to W37A LplA; we surmise that W37V better complements the shape and size of probe 1, whereas W37A may over-enlarge the LplA active site. We selected W37V as our best mutant and used this clone for all subsequent experiments.



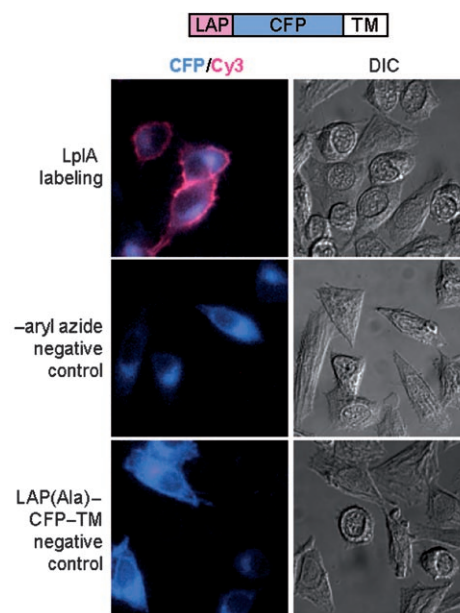
**Figure 2.** Engineering of the aryl azide ligase and characterization of specificity. a) Second-generation LplA mutants screened for aryl azide 1 ligation activity. Wild-type LplA and six mutants were compared under identical reaction conditions (800 μM aryl azide 1, 400 nM enzyme, 30°C for 30 min), and product (aryl azide-LAP-HP1 conjugate) was detected by HPLC. The conversion into product obtained with W37V LplA was normalized to 100% and other conversion values are reported relative to this. All measurements were performed in triplicate. b) HPLC trace showing ligation of aryl azide 1 to LAP-HP1 protein catalyzed by W37V LplA. The results of negative controls with ATP omitted or W37V replaced by wild-type LplA are also shown. The eluents of the two starred peaks were purified and analyzed by mass spectrometry (see Figure 4 in the Supporting Information). c) Aryl azide ligation performed on mammalian cell lysate to test labeling specificity. Lysate was generated from HEK cells expressing LAP-CFP (37 kDa). Lysate was labeled in vitro with LplA and aryl azide 1, then incubated with phosphine-Flag conjugate to derivatize the aryl azide. Blotting with anti-Flag antibody is shown to the right of the Coomassie-stained gel. Negative controls are shown with an alanine mutation in LAP (lane 2), wild-type LplA instead of W37V (lane 4), and untransfected cells (lane 3).

We characterized the W37V-catalyzed ligation of aryl azide **1** in terms of kinetics and specificity. First, we repeated the HPLC assay, with negative controls in which ATP was omitted or W37V was replaced by wild-type LplA. As expected, product was only observed in the presence of ATP and the W37V mutant (Figure 2b). We collected the HPLC-purified product and characterized it by mass spectrometry. Figure 4 in the Supporting Information shows that the mass corresponds to the calculated mass for one molecule of aryl azide **1** plus one LAP–HP1 protein, minus water. Next, we measured the kinetics of ligation. Using HPLC as readout, we determined a rate of catalysis  $k_{\text{cat}} = (0.31 \pm 0.04) \text{ s}^{-1}$ , which surprisingly is even faster than the natural ligation reaction (lipoylation of E2p:  $k_{\text{cat}} = 0.25 \text{ s}^{-1}$ ; [23] Figure 5 in the Supporting Information). Separate measurements showed that the Michaelis constant  $K_{\text{m}}$  was 80  $\mu\text{M}$  or less (data not shown), which is considerably greater than the  $K_{\text{m}}$  value of wild-type lipoic acid for LplA (1.7  $\mu\text{M}$  [21] or 4.5  $\mu\text{M}$  [24]).

We were concerned that remodeling of the LplA active site might change the peptide specificity of the enzyme. Therefore, we characterized the specificity of W37V-catalyzed aryl azide ligation in three ways. First, we prepared a total lysate from HEK cells expressing a LAP fusion with cyan fluorescent protein (LAP–CFP [23]). We labeled the lysate under forcing conditions with aryl azide, ATP, and W37V LplA, and then we derivatized the aryl azide with phosphine–Flag conjugate by using the Staudinger ligation. [26] Western blotting with anti-Flag antibody showed that, even though the LAP–CFP expression level is so low that it cannot be detected above endogenous proteins in the Coomassie-stained gel, only LAP–CFP receives the aryl azide, and the endogenous proteins in the lysate are not modified by W37V LplA (Figure 2c). Negative controls showed that labeling is specific for LAP and dependent on the W37V mutation (Figure 2c).

For our second specificity test, we performed ligation with aryl azide **1** on three non-LAP proteins: FRB–AP, FKBP–AP, [27] and BCCP–87. [28] We characterized the products by mass spectrometry (data not shown). None were found to be modified with aryl azide, despite exposure to reaction conditions that gave complete conversion of LAP–HP1 into the aryl azide conjugate.

Third, we tested peptide specificity by using a live cell imaging assay. A useful feature of the aryl azide probe is that it not only mediates photo-cross-linking but can also serve as a functional-group handle for the introduction of fluorescent probes. Although it is more common to derivatize *alkyl* azides with alkynes through [3+2] cycloaddition [23,29] or with phosphines through the Staudinger ligation, [30–32] we and other groups have observed facile reactions of *aryl* azides with both cyclooctynes [33] and phosphines. [30–32] To make use of this reactivity, we expressed a LAP–CFP–TM (transmembrane) cell-surface construct [23] in HeLa cells and performed extracellular labeling with W37V LplA, aryl azide, and ATP. After washing away excess aryl azide, we derivatized the LAP-ligated azide with cyclooctyne–cyanine 3 (Cy3) conjugate. [23] Imaging showed that only transfected (CFP-positive) cells became labeled with Cy3, while neighboring untransfected cells did not receive Cy3 (Figure 3). Negative controls with aryl azide omitted from the labeling reaction or with LAP–



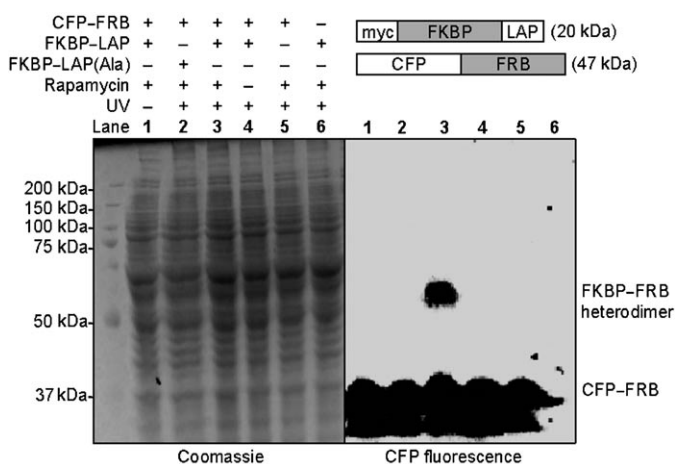
**Figure 3.** Specific fluorophore labeling on living cells through aryl azide **1** ligation. HeLa cells expressing LAP–CFP–TM were labeled with W37V LplA and aryl azide **1** for 40 min and, subsequently, with cyclooctyne–Cy3 conjugate [22] for 20 min, to selectively derivatize the aryl azide. Imaging shows pink Cy3 staining on the membranes of transfected CFP-positive cells. Negative controls are shown with omission of aryl azide (middle row) or with LAP–CFP–TM replaced by an alanine mutant (bottom row). The CFP and Cy3 images are merged. DIC: differential interference contrast.

CFP–TM replaced by an alanine point mutant in the LAP sequence gave no fluorophore labeling (Figure 3). Thus, our three assays collectively demonstrate that aryl azide ligation has the same high sequence specificity as lipoylation catalyzed by wild-type LplA.

To demonstrate the utility of the new aryl azide ligase, we attempted to detect a known PPI within a complex mixture. FKBP is known to interact with FRB in the presence but not absence of the small molecule rapamycin. [34] We prepared an FKBP–LAP construct and expressed it in HEK cells. In separate HEK cells, we expressed a CFP–FRB fusion. Total lysates were prepared from both samples, and they were mixed. We labeled the combined lysates with aryl azide **1**, ATP, and W37V LplA, before photo-cross-linking with 300–360 nm light in the presence or absence of rapamycin. We analyzed the samples by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS PAGE) and detected CFP–FRB by in-gel CFP fluorescence (Figure 4). Only imaging in the presence of rapamycin with UV light applied and the LAP tag intact did we observe a higher-molecular-weight band corresponding to the covalent FKBP–FRB heterodimer. Negative controls with either construct omitted gave no heterodimer band.

In summary, we have engineered an aryl azide ligase that brings us closer to the dream of routine PPI detection inside living cells by photo-cross-linking. Our ligase, generated by active-site engineering of *E. coli* LplA, catalyzes the extremely sequence-specific covalent ligation of a fluorinated aryl azide probe onto LAP fusion proteins, expressed on





**Figure 4.** Aryl azide mediated photo-cross-linking of FKBP and FRB in mammalian cell lysate. Lysate containing FKBP-LAP and CFP-FRB was labeled with W37V LpIA and aryl azide 1. Photo-cross-linking was then performed with 300–360 nm light for 4 min, in the presence (lane 3) or absence (lane 4) of rapamycin. The samples were run on SDS PAGE gels, and cross-linked FKBP-FRB heterodimer was detected by CFP imaging (right-hand gel). Negative controls are shown with no UV irradiation (lane 1), an alanine mutation in FKBP-LAP (lane 2), FKBP-LAP not expressed (lane 5), or CFP-FRB not expressed (lane 6). Coomassie visualization of the same samples (left-hand gel) shows equal loading in all lanes. Note that proteins run differently in the CFP gel because they are not denatured, in order to preserve CFP fluorescence.

living cell surfaces or in total lysate. Our immediate goal is to extend this ligation reaction to the cytosol of living cells.

In addition to photo-cross-linking, the aryl azide ligation reaction can also be used for fluorophore tagging, as previously demonstrated with an alkyl azide analogue of lipoic acid.<sup>[23]</sup> The aryl azide undergoes facile reaction with cyclooctyne-fluorophore conjugates on the surface of living cells.

This study demonstrates the malleability of the LpIA active site. Remarkably, the small-molecule binding site can be made to accommodate a variety of probe structures, without sacrificing the extremely high peptide/protein specificity that makes this enzyme such a useful tool for site-specific labeling applications. We were surprised to find that a single amino acid mutation (W37V) opened up the LpIA active site to a structure quite dissimilar to lipoic acid. Not all enlargements of the lipoate binding pocket had equivalent effects, however; W37A was a much poorer catalyst for aryl azide 1 ligation than W37V. We were also not successful in finding any catalysts for aryl azide 2, despite its smaller size. These lessons will guide us in our continued efforts to engineer LpIA for biotechnological applications.

Received: May 3, 2008

Published online: August 4, 2008

**Keywords:** ligases · photo-cross-linking · protein engineering · protein interactions · site-specific labeling

[1] Y. Tanaka, M. R. Bond, J. J. Kohler, *Mol. Biosyst.* **2008**, *4*, 473–480.

- [2] J. Brunner, *Annu. Rev. Biochem.* **1993**, *62*, 483–514.
- [3] J. R. Parrish, K. D. Gulyas, R. L. Finley, Jr., *Curr. Opin. Biotechnol.* **2006**, *17*, 387–393.
- [4] C. D. Hu, T. K. Kerppola, *Nat. Biotechnol.* **2003**, *21*, 539–545.
- [5] A. Galarneau, M. Primeau, L. E. Trudeau, S. W. Michnick, *Nat. Biotechnol.* **2002**, *20*, 619–622.
- [6] K. E. Luker, M. C. Smith, G. D. Luker, S. T. Gammon, H. Piwnica-Worms, D. Piwnica-Worms, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12288–12293.
- [7] E. Stefan, S. Aquin, N. Berger, C. R. Landry, B. Nyfeler, M. Bouvier, S. W. Michnick, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16916–16921.
- [8] K. Truong, M. Ikura, *Curr. Opin. Struct. Biol.* **2001**, *11*, 573–578.
- [9] W. van Crielinge, R. Beyaert, *Biol. Proced. Online* **1999**, *2*, 1–38.
- [10] 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.
- [11] J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11020–11024.
- [12] E. M. Tippmann, W. Liu, D. Summerer, A. V. Mack, P. G. Schultz, *ChemBioChem* **2007**, *8*, 2210–2214.
- [13] N. Hino, Y. Okazaki, T. Kobayashi, A. Hayashi, K. Sakamoto, S. Yokoyama, *Nat. Methods* **2005**, *2*, 201–206.
- [14] S. Ye, C. Kohrer, T. Huber, M. Kazmi, P. Sachdev, E. C. Yan, A. Bhagat, U. L. RajBhandary, T. P. Sakmar, *J. Biol. Chem.* **2007**, *283*, 1525–1533.
- [15] W. Wang, J. K. Takimoto, G. V. Louie, T. J. Baiga, J. P. Noel, K. F. Lee, P. A. Slesinger, L. Wang, *Nat. Neurosci.* **2007**, *10*, 1063–1072.
- [16] M. Suchanek, A. Radzikowska, C. Thiele, *Nat. Methods* **2005**, *2*, 261–267.
- [17] K. Kirshenbaum, I. S. Carrico, D. A. Tirrell, *ChemBioChem* **2002**, *3*, 235–237.
- [18] W. Liu, A. Brock, S. Chen, S. Chen, P. G. Schultz, *Nat. Methods* **2007**, *4*, 239–244.
- [19] L. Wang, J. Xie, P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249.
- [20] T. Lu, A. Y. Ting, J. Mainland, L. Y. Jan, P. G. Schultz, J. Yang, *Nat. Neurosci.* **2001**, *4*, 239–246.
- [21] D. E. Green, T. W. Morris, J. Green, J. E. Cronan, Jr., J. R. Guest, *Biochem. J.* **1995**, *309*, 853–862.
- [22] S. T. Ali, J. R. Guest, *Biochem. J.* **1990**, *271*, 139–145.
- [23] M. Fernández-Suárez, H. Baruah, L. Martínez-Hernández, K. T. Xie, J. M. Baskin, C. R. Bertozzi, A. Y. Ting, *Nat. Biotechnol.* **2007**, *25*, 1483–1487.
- [24] K. Fujiwara, S. Toma, K. Okamura-Ikeda, Y. Motokawa, A. Nakagawa, H. Taniguchi, *J. Biol. Chem.* **2005**, *280*, 33645–33651.
- [25] D. J. Kim, K. H. Kim, H. H. Lee, S. J. Lee, J. Y. Ha, H. J. Yoon, S. W. Suh, *J. Biol. Chem.* **2005**, *280*, 38081–38089.
- [26] K. L. Klick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 19–24.
- [27] M. Fernandez-Suarez, T. S. Chen, A. Y. Ting, *J. Am. Chem. Soc.* **2008**, *130*, 9251–9253.
- [28] A. Chapman-Smith, D. L. Turner, J. E. Cronan, Jr., T. W. Morris, J. C. Wallace, *Biochem. J.* **1994**, *302*, 881–887.
- [29] M. A. 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.
- [30] F. L. Lin, H. M. Hoyt, H. van Halbeek, R. G. Bergman, C. R. Bertozzi, *J. Am. Chem. Soc.* **2005**, *127*, 2686–2695.
- [31] S. J. Luchansky, S. Goon, C. R. Bertozzi, *ChemBioChem* **2004**, *5*, 371–374.
- [32] J. A. Restituyo, L. R. Comstock, S. G. Petersen, T. Stringfellow, S. R. Rajski, *Org. Lett.* **2003**, *5*, 4357–4360.
- [33] G. Wittig, A. Krebs, *Chem. Ber.* **1961**, *94*, 3260–3275.
- [34] L. A. Banaszynski, C. W. Liu, T. J. Wandless, *J. Am. Chem. Soc.* **2005**, *127*, 4715–4721.