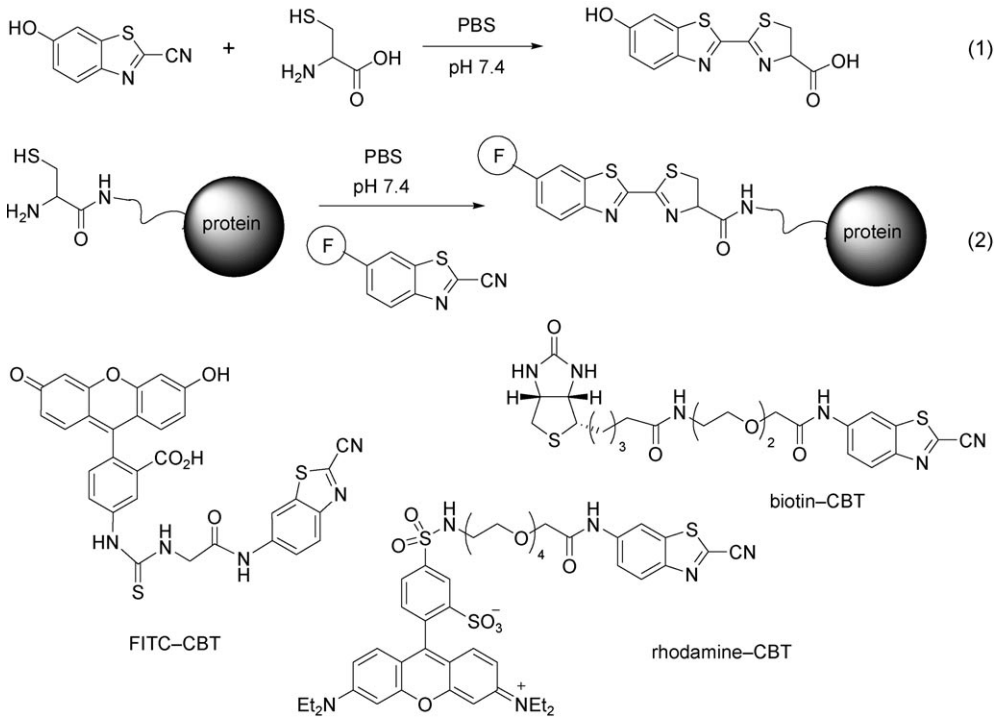
Protein Labeling

DOI: 10.1002/anie.200903627

A Biocompatible Condensation Reaction for the Labeling of Terminal Cysteine Residues on Proteins**

*Hongjun Ren, Fei Xiao, Ke Zhan, Young-Pil Kim, Hexin Xie, Zuyong Xia, and Jianghong Rao**

The site-specific labeling of proteins with molecular tags enables the direct visualization of protein dynamics, localization, and interactions in single living cells and is a powerful tool for studying the structure and function of proteins.[1] Proteins of interest can be labeled by genetic fusion to fluorescent proteins or chemical reactions with fluorescent dyes. For chemical labeling, a receptor protein is often used, for example, a mutant of human O^6-alkylguanine-DNA transferase[2a] or *Escherichia coli* dihydrofolate reductase,[2b] which binds to or reacts with a fluorescently tagged ligand.[2] Alternatively, smaller tags, such as short peptides, can be labeled by selective binding to fluorogenic dyes[3] or by enzyme-catalyzed ligation to fluorescent probes.[4] Water-compatible chemical reactions can also be applied to protein labeling, such as the Staudinger reaction between azides and triphenylphosphane,[5] the Huisgen cycloaddition or "click reaction" between azides and alkynes,[6] or reactions between aldehydes (or ketones) and aminooxy-containing reagents (or hydrazides).[7] Herein, we describe a water-compatible condensation reaction for the labeling of terminal cysteine residues on proteins in vitro and at the cell surface.



Scheme 1. Condensation reactions between cysteine/N-terminal cysteine residues and CBT derivatives in the synthesis of D-luciferin and site-specific protein labeling, and structures of the CBT probes used for labeling. PBS = phosphate-buffered saline.

N-Terminal cysteine residues have frequently been used in protein engineering for site-specific labeling and modification.[8] Thioesters are commonly used in a ligation reaction with terminal cysteine residues that proceeds through transthioesterification and S-to-N acyl exchange.[9] This native chemical ligation reaction has been applied successfully to protein semisynthesis and labeling.[10]

Our method to label the terminal cysteine residue of a protein is based on the condensation of 2-cyanobenzothiazole (CBT) and D-cysteine. This reaction, which is used as the last step of the synthesis of D-luciferin, a common substrate for firefly luciferase (reaction 1 in Scheme 1),[11] can proceed smoothly in aqueous solutions. We hypothesized that CBT could react with the terminal cysteine residue of a protein. If a fluorophore was conjugated to the CBT motif, this reaction should enable the ligation of a fluorescent label specifically to the terminal cysteine residue of the protein (reaction 2 in Scheme 1).

We first investigated whether CBT could react with functional groups other than free cysteine. With homocysteine, which contains a 1,3-aminothiol group, a stable condensation product with a six-membered ring was gener-

[*] Dr. H. Ren, F. Xiao, K. Zhan, Y.-P. Kim, H. Xie, Z. Xia, Prof. J. Rao
Molecular Imaging Program at Stanford
Departments of Radiology and Chemistry, Stanford University
1201 Welch Road, Stanford, California 94305-5484 (USA)
Fax: (+1) 650-736-7925
E-mail: jrao@stanford.edu
Homepage: http://raolab.stanford.edu

[**] This research was supported by a grant from NIGMS (R01GM086196-01). We thank Prof. Matthew Bogyo at Stanford for access to the mass spectrometry facility.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200903627.

ated. As expected, when the thiol group was replaced with a hydroxy group, for example, in β-amino alcohols and serine, no detectable products were formed under similar conditions.

When glutathione or β-mercaptoethanol was mixed with 2-cyano-6-aminobenzothiazole (amino-CBT) at a molar ratio of 2:1, a new peak was detected by HPLC analysis after 30 minutes besides that of the remaining amino-CBT (50%). However, both of these peaks disappeared after the further addition of free L-cysteine, and only the condensation product L-aminoluciferin was observed. This result suggests that CBT can react reversibly with free thiol groups. The reaction is selective for 1,2-aminothiol (or 1,3-aminothiol) substrates over simple thiol groups. Indeed, when amino-CBT was mixed with cysteine and glutathione (or 2-mercaptoethanol) at a molar ratio of 1:5:5, only the condensation product L-aminoluciferin was observed by HPLC analysis (see Figure S2 in the Supporting Information).

We evaluated whether aromatic cyano compounds other than CBT could similarly react and cyclize with free cysteine. Both benzonitrile and picolinonitrile failed to produce detectable products under the same conditions. A mixture of picolinonitrile and amino-CBT (1:1) with free L-cysteine only afforded L-aminoluciferin, as determined by HPLC analysis. The second-order rate constant for this reaction was determined to be $9.19\,M^{-1}s^{-1}$ (see Figure S4 in the Supporting Information), which is significantly larger than the value reported for a biocompatible click reaction ($7.6\times10^{-2}\,M^{-1}s^{-1}$).[12]

Before applying the reaction to protein labeling, we tested the labeling method with cysteine-containing peptides. Several peptides with an N-terminal L-cysteine residue were synthesized. HPLC analysis indicated that they all underwent conjugation with amino-CBT in a phosphate buffer at pH 7.4 at room temperature in more than 90% yield within 30 minutes; the identity of each product was confirmed by mass spectrometry (see Table S1 in the Supporting Information). For peptides with the cysteine residue in the middle of the sequence, no ligation product was detected by HPLC analysis, which suggests that CBT specifically labels N-terminal cysteine residues of peptides.

We next tested this method to label a cysteine residue at the N terminus of the bioluminescent protein *Renilla* luciferase. Both proteolytic processing and the spontaneous hydrolysis of intein fusion protein have been reported to generate an N-terminal cysteine residue on a protein.[8c,10c] We fused the peptide substrate of tobacco etch virus (TEV) protease (ENLYFQ↓C; arrow indicates the cleavage site) to the N terminus of *Renilla* luciferase (Figure 1a). Furthermore, a six-histidine tag sequence was added in front of the TEV protease substrate to facilitate purification. The fusion protein was expressed and purified on a Ni^{2+}/nitrilotriacetic acid (NTA) column. TEV protease was then added to cleave the substrate, and elution afforded the product N-terminal-cysteine luciferase (Cys-rLuc).

When Cys-rLuc was incubated with amino-CBT or FITC–CBT (FITC = fluorescein isothiocyanate) at room temperature for two hours, a fluorescent ligation product of the expected size was observed clearly on the gel (Figure 1b, lanes 4 and 5). In comparison, no labeling products were

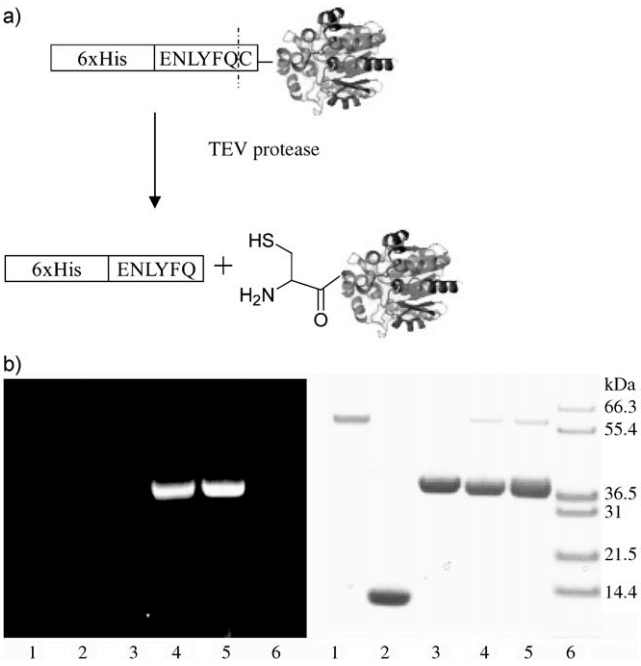


Figure 1. In vitro labeling of N-terminal cysteine residues on proteins with CBT probes. a) Generation of an N-terminal cysteine residue through protease processing. b) Fluorescence (left) and white-light (right) images of a gel loaded with proteins (10 μM) labeled with amino-CBT (50 μM; lane 1: BSA; lane 2: lysozyme; lane 3: rLuc; lane 4: Cys-rLuc) or FITC–CBT (lane 5: Cys-rLuc) and stained with Coomassie Blue (lane 6: size markers). All reactions were quenched with free cysteine before gel loading.

observed on the gel for three control proteins without an N-terminal cysteine residue—bovine serum albumin (BSA), lysozyme, and unmodified luciferase—even though all contained cysteine residues in their sequences (Figure 1b, lanes 1–3). This result demonstrates that the ligation takes place specifically with N-terminal cysteine residues.

A biotinylated CBT probe was prepared for labeling Cys-rLuc (Scheme 1). A similar reaction afforded biotinylated Cys-rLuc with a measured molecular weight of 37330 Da (the calculated MW is 37336 Da; see Figure S5 in the Supporting Information). The biotinylated Cys-rLuc was able to bind streptavidin to form the expected complex, as revealed by gel electrophoresis (see Figure S6 in the Supporting Information).

A free cysteine residue may be introduced at the C terminus of rLuc by the intein-mediated cleavage reaction for CBT labeling (Figure 2a). A recombinant protein containing rLuc and *Mex* GyrA intein (*Mycobacterium xenopi* gyrase A intein, a 198 aa natural mini intein[13] which lacks a central intein endonuclease domain) was expressed and purified. The GyrA intein catalyzed the formation of the thioester intermediate between the N terminus of GyrA and the C terminus of rLuc. The addition of thiol nucleophiles, such as L-cysteine or dicysteine (two carboxylate groups are linked by an ethyldiamine moiety), resulted in the cleavage of the thioester to generate a free GyrA species and rLuc with the nucleophile conjugated at its C terminus. With L-cysteine, the product rLuc-Cys only contained a free thiol group, but

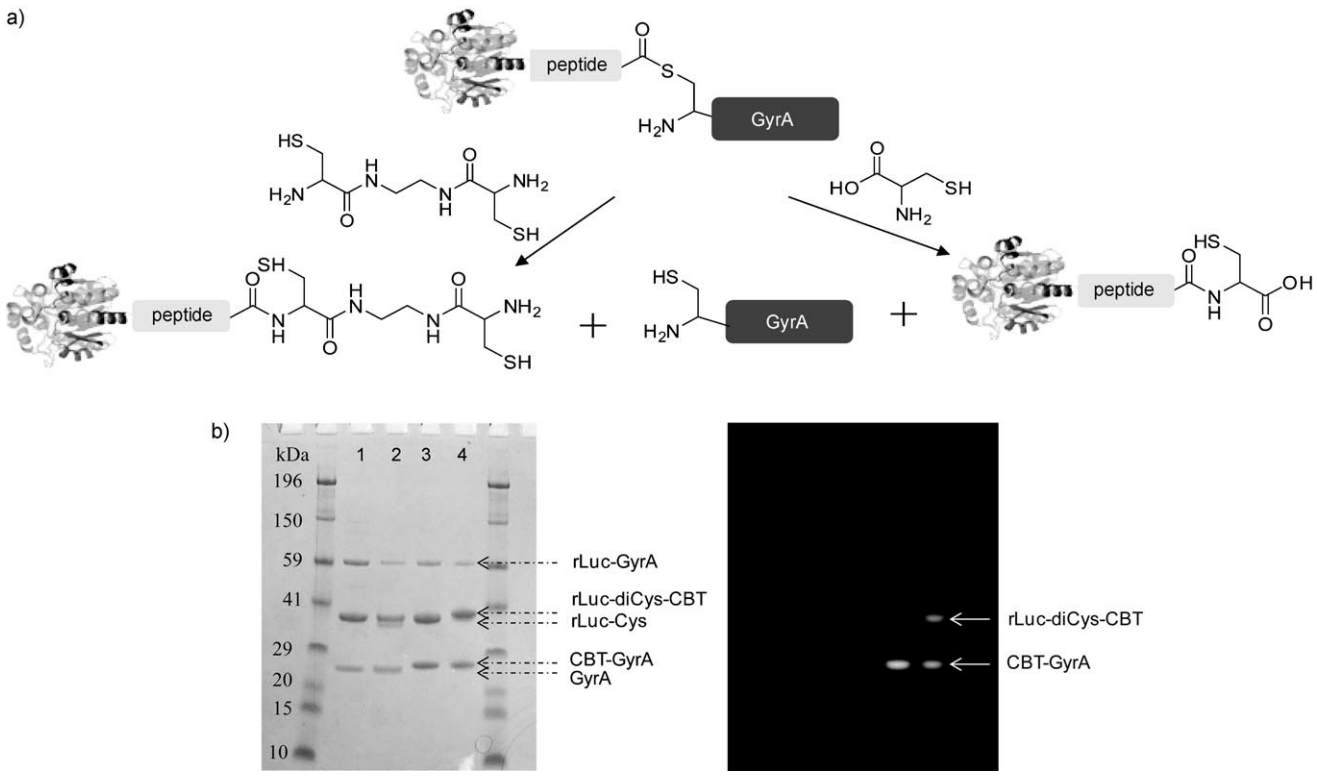


Figure 2. In vitro labeling of cysteine residues introduced at the C terminus of proteins with CBT probes. a) Introduction of free cysteine residues at the C terminus of rLuc through intein-mediated splicing; a linker peptide is present between rLuc and GyrA. b) Fluorescence (right) and white-light (left) images of a gel loaded with reaction solutions (lane 1: rLuc-Cys; lane 2: rLuc-diCys; lane 3: rLuc-Cys + peptide–FITC–CBT; lane 4: rLuc-diCys + peptide–FITC–CBT) and stained with Coomassie Blue.

with dicysteine, the product rLuc-diCys also contained a free cysteine residue at the C terminus available for CBT labeling. Both L-cysteine and dicysteine cleaved the thioester efficiently and gave the expected products (Figure 2b, lanes 1 and 2). When the two reaction solutions, containing 5 μM protein, were incubated with a CBT probe (peptide–FITC–CBT at 50 μM; see Figure S7 in the Supporting Information), free GyrA formed a fluorescent adduct detectable on the gel because it contained an N-terminal cysteine residue (Figure 2b, lanes 3 and 4); on the other hand, fluorescent staining was observed for rLuc-diCys (lane 4) but not for rLuc-Cys (lane 3). No fluorescent staining on the gel was observed for the uncleaved fusion protein that remained in the reaction mixture either, which suggested that the thiol group of the cysteine residue at the reactive site of the fused GyrA protein did not react with CBT. This result again showed that a terminal cysteine residue on a protein can be labeled specifically by the CBT reaction.

The labeling kinetics of GyrA was measured by gel analysis of the reaction mixture with different incubation times (Figure 3a). Consistent with the fast kinetics observed with free cysteine, the labeling of GyrA by amino-CBT was nearly 50% complete after the first 10 minutes and 75% complete after 150 minutes (Figure 3b). The labeling of free cysteine at the C terminus of rLuc-diCys was nearly complete (conversion almost 100%) after one hour when 4 equivalents of the CBT probe were used (see Figure S8 in the Supporting Information).

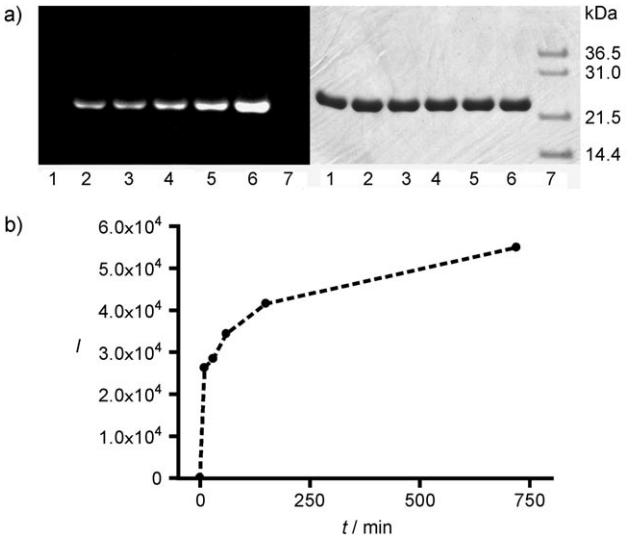


Figure 3. Time-dependent labeling of the N-terminal cysteine residue of GyrA. a) Fluorescence (left) and white-light (right) images of a gel loaded with reaction solutions containing GyrA (7.1 μM) and amino-CBT (57 μM) at different reaction times (lane 1: 0 min; lane 2: 10 min; lane 3: 30 min; lane 4: 1 h; lane 5: 2.5 h; lane 6: 12 h) and stained with Coomassie Blue (lane 7: size markers). The reactions were quenched with cysteine before gel loading. b) Plot of integrated fluorescence-band intensity versus time.

We were interested in whether this condensation reaction could be applied to the specific labeling of proteins in vivo.

Cyan fluorescent protein (CFP) was used as a model target and expressed on the cell surface. TEV protease was used to generate an N-terminal cysteine residue through the proteolytic processing of the TEV protease substrate fused to the N terminus of CFP. To ensure that the fusion protein was targeted on the extracellular side of the cell membrane, the murine immunoglobulin (Ig) κ-chain leader sequence was added immediately before the N terminus of the substrate sequence of TEV protease, and its C terminus contained the platelet derived growth factor receptor (PDGFR) transmembrane domain. TEV protease treatment would remove the murine Ig κ-chain leader sequence and the substrate fragment containing the ENLYFQ sequence, and expose the free N-terminal cysteine residue to labeling with the rhodamine–CBT probe (Figure 4a).

We first verified the labeling of the fusion protein in cell lysates. HeLa cells were transfected with the CFP construct and harvested for lysis. The collected supernatant was incubated with the rhodamine–CBT probe in the presence or absence of TEV protease. Gel analysis revealed a fluorescent band corresponding to the size of the CFP fusion protein (34 kDa) in the presence of TEV protease; however, no band was detected in the absence of TEV protease (Figure 4b). Furthermore, no other fluorescent bands were observed on the gel. This result confirmed the specificity of the labeling.

Live transfected HeLa cells were incubated with the labeling solution containing both the rhodamine–CBT probe and TEV protease for 30 minutes at 37 °C. Subsequent washing removed the unreacted probe before imaging. Live-cell fluorescence imaging in the CFP channel confirmed the membrane expression of the fusion protein (Figure 4c). Clear membrane staining by the probe was observed in the rhodamine channel and was almost perfectly superimposable with that observed in the CFP channel (Figure 4c). Typically, more than half of the cells were labeled by the rhodamine–CBT probe by this procedure. A control staining without TEV protease in the labeling solution gave little detectable fluorescence (Figure 4d). Cells transfected with nontagged CFP were not labeled either. Similar labeling was carried out successfully with other probes on other cell lines, such as COS-7 (see Figure S10 in the Supporting Information). These results demonstrate that the condensation reaction can take place in the context of live cells with the specific labeling of the N-terminal cysteine residue of target proteins on the cell surface.

The condensation reaction is mechanistically different from the thioester-based ligation, and the CBT probes show good stability and selectivity for 1,2-aminothiols over thiol groups without an adjacent amino group. An excess of the thiol is often added to facilitate the thioester-based ligation but is unnecessary for the CBT condensation reaction.

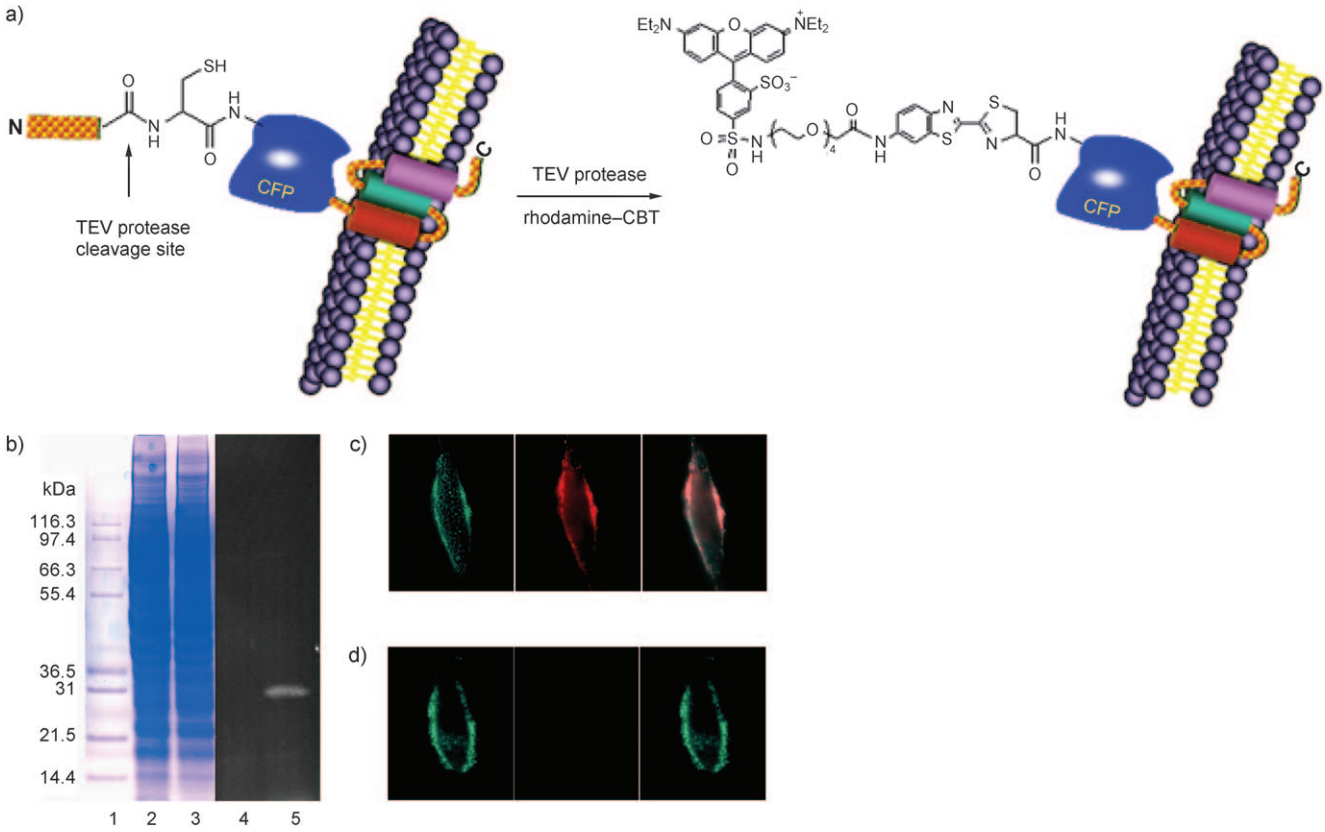


Figure 4. Labeling of CFP on live HeLa cells with rhodamine–CBT. a) Schematic illustration of the labeling strategy. b) Analysis by gel electrophoresis of the in vitro labeling of cell lysates with rhodamine–CBT (lane 1: size markers; lane 2: cell lysates without TEV protease; lane 3: cell lysates with TEV protease ; lanes 4 and 5: fluorescence image of lanes 2 and 3, respectively). c) Labeling of tagged CFP in transfected HeLa cells with TEV protease (10 units) and rhodamine–CBT (2 μM). Left: CFP fluorescence; middle: rhodamine fluorescence; right: an overlay of left and middle images. d) Images of HeLa cells transfected with tagged CFP under the same conditions as in (c) but without TEV protease.

Furthermore, in vivo labeling proceeds much faster than reported examples of thioester-based labeling, which required 24 hours as opposed to only 30 minutes in our study.[10a,c]

Although this labeling method is limited to the terminal cysteine residue of a target protein, the small tag size (with just one amino acid) should offer an important advantage. The simple procedure for live-cell labeling and the short labeling time are also attractive. A free cysteine residue at the C terminus may be labeled as well, although its generation requires extra chemical modifications that are less straightforward than the enzyme-mediated ligation approach.[4e] This reaction may also be applied to protein ligation: a protein fragment with an N-terminal cysteine residue may be ligated to another fragment containing a CBT moiety that has been introduced chemically at its C terminus.

In summary, we have described herein a condensation reaction for the labeling of terminal cysteine residues on proteins in vitro and on cell surfaces. A useful alternative to existing approaches to protein labeling, this simple condensation reaction is compatible with physiological conditions and proceeds with a high degree of specificity and efficiency.

Received: July 3, 2009
Revised: September 24, 2009
Published online: November 18, 2009

Keywords: chemical ligation · condensation · fluorescent probes · live-cell imaging · protein modifications

[1] a) R. Y. Tsien, *Annu. Rev. Biochem.* **1998**, *67*, 509; b) K. M. Marks, G. P. Nolan, *Nat. Methods* **2006**, *3*, 591; c) N. Johnsson, K. Johnsson, *ACS Chem. Biol.* **2007**, *2*, 31; d) A. Dragulescu-Andrasi, J. Rao, *ChemBioChem* **2007**, *8*, 1099; e) M. Fernandez-Suarez, A. Y. Ting, *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 929.

[2] a) A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, *Nat. Biotechnol.* **2003**, *21*, 86; b) L. W. Miller, J. Sable, P. Goelet, M. P. Sheetz, V. W. Cornish, *Angew. Chem.* **2004**, *116*, 1704; *Angew. Chem. Int. Ed.* **2004**, *43*, 1672; c) G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohane, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit, K. V. Wood, *ACS Chem. Biol.* **2008**, *3*, 373; d) S. Mizukami, S. Watanabe, Y. Hori, K. Kikuchi, *J. Am. Chem. Soc.* **2009**, *131*, 5016.

[3] a) B. A. Griffin, S. R. Adams, R. Y. Tsien, *Science* **1998**, *281*, 269; b) T. Wang, P. Yan, T. C. Squier, M. U. Mayer, *ChemBiochem* **2007**, *8*, 1937; c) C. T. Hauser, R. Y. Tsien, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 3693; d) E. G. Guignet, R. Hovius, H. Vogel, *Nat. Biotechnol.* **2004**, *22*, 440; e) Y. Yano, A. Yano, S. Oishi, Y. Sugimoto, G. Tsujimoto, N. Fujii, K. Matsuzaki, *ACS Chem. Biol.* **2008**, *3*, 341; f) A. Ojida, K. Honda, D. Shinmi, S. Kiyonaka, Y. Mori, I. Hamachi, *J. Am. Chem. Soc.* **2006**, *128*, 10452; g) T. L. Halo, J. Appelbaum, E. M. Hobert, D. M. Balkin, A. Schepartz, *J. Am. Chem. Soc.* **2009**, *131*, 438.

[4] a) I. Chen, M. Howarth, W. Lin, A. Y. Ting, *Nat. Methods* **2005**, *2*, 99; b) L. Vivero-Pol, N. George, H. Krumm, K. Johnsson, N. Johnsson, *J. Am. Chem. Soc.* **2005**, *127*, 12770; c) J. Yin, F. Liu, X. Li, C. T. Walsh, *J. Am. Chem. Soc.* **2004**, *126*, 7754; d) K. M. Clarke, A. C. Mercer, J. J. La Clair, M. D. Burkart, *J. Am. Chem. Soc.* **2005**, *127*, 11234; e) T. Tanaka, T. Yamamoto, S. Tsukiji, T. Nagamune, *ChemBioChem* **2008**, *9*, 802; f) Z. Zhou, P. Cironi, A. J. Lin, Y. Xu, S. Hrvatin, D. E. Golan, P. A. Silver, C. T. Walsh, J. Yin, *ACS Chem. Biol.* **2007**, *2*, 337; g) C. Gauchet, G. R. Labadie, C. D. Poulter, *J. Am. Chem. Soc.* **2006**, *128*, 9274; h) B. P. Duckworth, J. Xu, T. A. Taton, A. Guo, M. D. Distefano, *Bioconjugate Chem.* **2006**, *17*, 967.

[5] E. Saxon, C. R. Bertozzi, *Science* **2000**, *287*, 2007.

[6] a) H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem.* **2001**, *113*, 2056; *Angew. Chem. Int. Ed.* **2001**, *40*, 2004; b) A. J. Link, D. A. Tirrell, *J. Am. Chem. Soc.* **2003**, *125*, 11164; c) N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126*, 15046.

[7] L. K. Mahal, K. J. Yarema, C. R. Bertozzi, *Science* **1997**, *276*, 1125.

[8] a) P. E. Dawson, S. B. Kent, *Annu. Rev. Biochem.* **2000**, *69*, 923; b) P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776; c) J. Xiao, A. Burn, T. J. Tolbert, *Bioconjugate Chem.* **2008**, *19*, 1113; d) I. E. Gentle, D. P. De Souza, M. Baca, *Bioconjugate Chem.* **2004**, *15*, 658; e) J. B. Blanco-Canosa, P. E. Dawson, *Angew. Chem.* **2008**, *120*, 6957; *Angew. Chem. Int. Ed.* **2008**, *47*, 6851; f) D. A. Erlanson, M. Chytil, G. L. Verdine, *Chem. Biol.* **1996**, *3*, 981.

[9] a) T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 249; b) D. S. Y. Yeo, R. Srinivasan, G. Y. J. Chen, S. Q. Yao, *Chem. Eur. J.* **2004**, *10*, 4664; c) V. Muralidharan, T. W. Muir, *Nat. Methods* **2006**, *3*, 429.

[10] a) D. S. Y. Yeo, R. Srinivasan, M. Uttamchandani, G. Y. J. Chen, Q. Zhu, S. Q. Yao, *Chem. Commun.* **2003**, 2870; b) G. K. Busch, E. W. Tate, P. R. J. Gaffney, E. Rosivatz, R. Woscholski, R. J. Leatherbarrow, *Chem. Commun.* **2008**, 3369; c) S. Chattopadhaya, R. Srinivasan, D. S. Y. Yeo, G. Y. J. Chen, S. Q. Yao, *Bioorg. Med. Chem.* **2009**, *17*, 981.

[11] E. H. White, F. McCapra, G. F. Field, W. D. McElroy, *J. Am. Chem. Soc.* **1961**, *83*, 2402.

[12] 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.

[13] a) A. Telenti, M. Southworth, F. Alcaide, S. Daugelat, W. R. Jacobs, Jr., F. B. Perler, *J. Bacteriol.* **1997**, *179*, 6378; b) Z. Xia, Y. Xing, M.-K. So, A. L. Koh, R. Sinclair, J. Rao, *Anal. Chem.* **2008**, *80*, 8649.

