A Biocompatible Condensation Reaction for Labeling of Terminal Cysteines on Proteins

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

Molecular Imaging Program at Stanford, Departments of Radiology and Chemistry, Stanford University, 1201 Welch Road, Stanford, California 94305-5484 (USA)

Abstract

This communication describes a novel protein labeling method that uses a single amino-acid tag — N-terminal cysteine residue — and small-molecule probes carrying the cyanobenzothiazole unit for specific labeling of proteins in vitro and at the surface of live cells. This simple ligation reaction proceeds with a high degree of specificity in physiological conditions, and should offer an important alternative to currently available protein labeling methods.

Graphical abstract
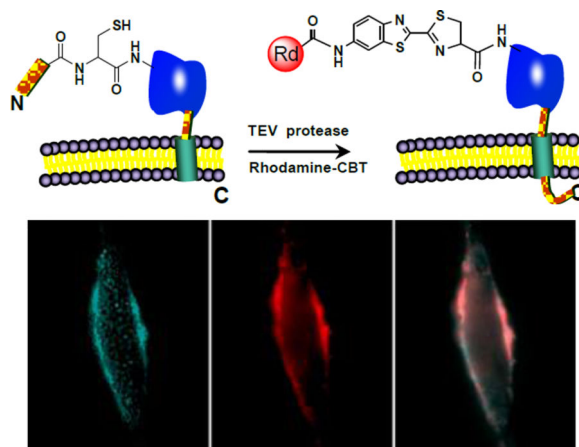
Keywords

Protein Labeling; Condensation; Terminal Cysteine; Chemical ligation; Live-cell Imaging

Site-specific labeling of proteins with molecular tags enables direct visualization of protein dynamics, localization and interactions in single living cells and is a powerful tool for studying structure and function of proteins.[1] Proteins of interest can be labeled by genetic fusions to fluorescent proteins, or chemical reactions with fluorescent dyes. Chemical labeling often employs a receptor protein, for example, a mutant of human O^6-alkylguanine-

[*]Fax: (+1) 650-736-7925, ; Email: jrao@stanford.edu, Homepage: http://raolab.stanford.edu

DNA transferase,[2a] and *E. coli* dihydrofolate reductase,[2b] that binds to or reacts with its 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 the azides and triphenylphosphines,[5] the Huisgen cycloaddition or "Click chemistry" between the azides and alkynes,[6] the reaction between aldehydes (or ketones) and aminooxy containing reagents (or hydrazides).[7] Herein, we describe a water-compatible condensation reaction for labeling terminal cysteine residues on proteins in vitro and at the cell surface.

N-terminal cysteine has been frequently used in protein engineering for site-specific labeling and modification.[8] Thioesters are commonly used in a ligation reaction with terminal cysteines, which proceeds through thioester and S- to N-acyl exchanges.[9] This native chemical ligation reaction has been successfully applied to protein semi-synthesis and labeling.[10]

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

We first investigated whether CBT could react with other functional groups than free cysteine. Homocysteine possessing 1,3-aminothiol group was able to generate a stable condensation product and form a six-member ring. As expected, replacing the thiol group by hydroxyl, for example, β-aminoalcohol and serine, rendered no detectable products under the similar conditions.

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

We evaluated whether other aromatic cyano compounds than CBT could similarly react and cyclize with free cysteine. Both benzonitirile 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 the L-aminoluciferin, as determined by HPLC analysis.

The second order rate constant was determined to be 9.19 $M^{-1} \cdot S^{-1}$ (Supporting Information Figure S4), which is significantly larger than the reported value of a biocompatible click reaction (7.6×10^{-2} $M^{-1} \cdot S^{-1}$).[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 were synthesized. HPLC analyses indicated that they all were able to conjugate with amino-CBT in the pH 7.4 phosphate buffers at room temperature with more than 90% yield within 30 minutes, and the correct identity of each product was confirmed by Mass Spectroscopy (Supporting Information Table S1). For peptides with the cysteine residue in the middle of the sequence, no ligation product was detected on HPLC, suggesting that CBT can specifically label the N-terminal cysteine in peptides.

We next tested this method to label free cysteine at the N-terminus of the bioluminescent protein *Renilla* luciferase. Both proteolytic processing and spontaneous hydrolysis of intein fusion protein have been reported to generate N-terminal cysteine 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). In addition, a sequence of six-histidine tag was added in front of the TEV protease substrate to facilitate the purification. The fusion protein was expressed and purified on the Ni^{2+}/NTA column. TEV protease was then added to cleave the substrate and elution afforded the product N-terminal-Cys luciferase (Cys-rLuc).

When Cys-rLuc was incubated with amino-CBT or FITC-CBT at room temperature for 2 h, a fluorescent ligation product at an expected size was clearly observed on the gel (lane 4 and 5 in Figure 1b). In comparison, three control proteins without an N-terminal cysteine—bovine serum albumin (BSA), lysozyme, and unmodified luciferase showed no labeling products on the gel even all contained cysteine residues in their sequences (lane 1, 2 and 3 in Figure 1b). This result demonstrates that the ligation takes place specifically with the N-terminal cysteine on the protein.

A biotinylated CBT probe was prepared for labeling Cys-rLuc (Scheme 1). Similar reaction afforded the biotinylated Cys-rLuc, which gave a measured molecular weight of 37,330 Dalton (the calculated MW is 37,336 Dalton) (Supporting Information Figure S5). The biotinylated Cys-rLuc was able to bind streptavidin and form the complexes, as revealed by the gel electrophoresis analysis (Supporting Information Figure S6).

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 thiol ester intermediate at the junction between the N-terminus of GyrA and the C-terminus of rLuc. Adding thiol nucleophiles such as L-cysteine or dicysteine (two carboxylate groups are linked by an ethyldiamine) resulted in the cleavage of the thioester, generating a free GyrA and rLuc with the nucleophile conjugated at its C-terminus. For L-cysteine, the product rLuc-Cys only contained a free thiol group, but for

dicysteine, the product rLuc-diCys also contained a free cysteine at the C-terminus available for the CBT labeling. As shown in Figure 2b, both L-cysteine and dicysteine efficiently cleaved the thioester and gave expected products (lane 1 and 2). When both reaction solutions were incubated with a CBT probe (peptide-FITC-CBT; Supporting Information Figure S7), free GyrA formed a fluorescent adduct detectable on the gel because it contained an N-terminal cysteine (lane 3 and 4); on the other hand, rLuc-diCys (lane 4) gave the fluorescent staining but rLuc-Cys did not (lane 3). The uncleaved fusion protein left in the reaction did not give any fluorescent staining on the gel either, suggesting that the thiol group of the cysteine at the reactive site of the fused GyrA protein didn't react with CBT. This result has further shown that free terminal cysteine on a protein can be specifically labeled by the CBT reaction.

The labeling kinetics of GyrA was measured by gel analysis of the reaction mixture with different incubation time (Figure 3a). Consistent with the fast kinetics observed with free cysteine, the labeling of GyrA by amino-CBT was nearly 50% complete in the first 10 mins and 75% complete in 150 mins (Figure 3b). In the example of labeling free cysteine at the C-terminus of rLuc-diCys, it was nearly 100% complete in 1 hour when four equivalents of CBT probe was used (Supporting Information Figure S8).

We were interested in whether this condensation reaction could be applied to specifically label proteins in vivo. Cyan fluorescent protein (CFP) was used as a model target and expressed on the cell surface. The TEV protease was used to generate the N-terminal cysteine residue via 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 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. The TEV protease treatment will remove the murine Ig κ-chain leader sequence and the substrate fragment containing ENLYFQ sequence, and expose the free N-terminus cysteine for 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 at the size of the CFP fusion protein (34 kD) in the presence of TEV protease but no detected band in the absence of TEV protease (Figure 4b). Furthermore, no other fluorescent bands were observed on the gel, confirming 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 min at 37 °C. Subsequent washing removed the unreacted probe before imaging. Live-cell fluorescence imaging at the CFP channel confirmed the membrane expression of the fusion protein (Figure 4c). A clear membrane staining by the probe was observed under the rhodamine channel, which overlaid nearly perfectly with the CFP pattern (Figure 4c). Typically, more than half of the cells were labeled by the rhodamine-CBT probe using this protocol. A control staining without the TEV protease in the labeling solution gave little detectable fluorescence signals (Figure 4d).

Cells transfected with nontagged CFP were not labeled either. Similar labeling was successfully carried out with other probes on other cell lines such as COS-7 (Supporting Information Figure S10). These results demonstrate that the condensation reaction can take place in the context of live cells and specifically label N-terminal cysteine 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 β-mercaptoamine over free thiol compounds. Excess thiol compounds are often added to facilitate the thioester-based ligation but it is unnecessary for the CBT condensation reaction. Furthermore, in vivo labeling proceeds much faster than reported examples of the thioester-based labeling which required 24 hours whereas only 30 mins in our study.[10a,c]

While this labeling method is limited to the terminal cysteine of a target protein, the small tag size – with just one amino acid should offer an important advantage. The simple live-cell labeling procedure and a short labeling time are also attractive. Free cysteine at the C-terminus may be labeled as well though its generation requires extra chemical modifications, which is 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 may be ligated to another fragment containing a CBT moiety chemically introduced at its C-terminus.

In summary, we describe here a condensation reaction for labeling of terminal cysteine residues on proteins in vitro and on cell surface. This simple condensation reaction is compatible with physiological conditions and proceeds with a high degree of specificity and efficiency. It represents a useful addition to the existing protein labeling tools for labeling terminal cysteines.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. a) Tsien RY. Annu. Rev. Biochem. 1998; 67:509. [PubMed: 9759496] b) Marks KM, Nolan GP. Nat. Methods. 2006; 3:591. [PubMed: 16862131] c) Johnsson N, Johnsson K. ACS Chem. Biol. 2007; 2:31. [PubMed: 17243781] d) Dragulescu-Andrasi A, Rao J. Chem Bio Chem. 2007; 8:1099.e) Fernandez-Suarez M, Ting AY. Nat. Rev. Mol. Cell Biol. 2008; 9:929. [PubMed: 19002208]

2. a) Keppler A, Gendreizig S, Gronemeyer T, Pick H, Vogel H, Johnsson K. Nat. Biotechnol. 2003; 21:86. [PubMed: 12469133] b) Miller LW, Sable J, Goelet P, Sheetz MP, Cornish VW. Angew. Chem. 2004; 116:1704. *Angew. Chem. Int. Ed.***2004**, *43*, 1672. c) Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, Wood MG, Learish R, Ohane RF, Urh M, Simpson D, Mendez J, Zimmerman K, Otto P, Vidugiris G, Zhu J, Darzins A, Klaubert DH, Bulleit RF, Wood KV. ACS Chem. Biol. 2008; 3:373. [PubMed: 18533659] d) Mizukami S, Watanabe S, Hori Y, Kikuchi K. J. Am. Chem. Soc. 2009; 131:5016. [PubMed: 19296682]

3. a) Griffin BA, Adams SR, Tsien RY. Science. 1998; 281:269. [PubMed: 9657724] b) Wang T, Yan P, Squier TC, Mayer MU. Chem Bio chem. 2007; 8:1937.c) Hauser CT, Tsien RY. Proc. Natl. Acad. Sci. USA. 2007; 104:3693. [PubMed: 17360414] d) Guignet EG, Hovius R, Vogel H. Nat. Biotechnol. 2004; 22:440. [PubMed: 15034592] e) Yano Y, Yano A, Oishi S, Sugimoto Y, Tsujimoto G, Fujii N, Matsuzaki K. ACS Chem. Biol. 2008; 3:341. [PubMed: 18533657] f) Ojida A, Honda K, Shinmi D, Kiyonaka S, Mori Y, Hamachi I. J. Am. Chem. Soc. 2006; 128:10452. [PubMed: 16895410] g) Halo TL, Appelbaum J, Hobert EM, Balkin DM, Schepartz A. J. Am. Chem. Soc. 2009; 131:438. [PubMed: 19105691]

4. a) Chen I, Howarth M, Lin W, Ting AY. Nat. Methods. 2005; 2:99. [PubMed: 15782206] b) Vivero-Pol L, George N, Krumm H, Johnsson K, Johnsson N. J. Am. Chem. Soc. 2005; 127:12770. [PubMed: 16159249] c) Yin J, Liu F, Li X, Walsh CT. J. Am. Chem. Soc. 2004; 126:7754. [PubMed: 15212504] d) Clarke KM, Mercer AC, La Clair JJ, Burkart MD. J. Am. Chem. Soc. 2005; 127:11234. [PubMed: 16089439] e) Tanaaka T, Yamamoto T, Tsukiji S, Nagamune T. Chem Bio Chem. 2008; 9:802.f) Zhou Z, Cironi P, Lin AJ, Xu Y, Hrvatin S, Golan DE, Silver PA, Walsh CT, Yin J. ACS Chem. Biol. 2007; 2:337. [PubMed: 17465518] g) Gauchet C, Labadie GR, Poulter CD. J. Am. Chem. Soc. 2006; 128:9274. [PubMed: 16848430] h) Duckworth BP, Xu J, Taton TA, Guo A, Distefano MD. Bioconjugate Chem. 2006; 17:967.

5. Saxon E, Bertozzi CR. Science. 2000; 287:2007. [PubMed: 10720325]

6. a) Kolb HC, Finn MG, Sharpless KB. Angew. Chem. 2001; 113:2056. *Angew. Chem. Intl. Ed. Eng.* **2001**, *40*, 2004. b) Link AJ, Tirrell DA. J. Am. Chem. Soc. 2003; 125:11164. [PubMed: 16220915] c) Agard NJ, Prescher JA, Bertozzi CR. J. Am. Chem. Soc. 2004; 126:15046. [PubMed: 15547999]

7. Mahal LK, Yarema KJ, Bertozzi CR. Science. 1997; 276:1125. [PubMed: 9173543]

8. a) Dawson PE, Kent SB. Annu. Rev. Biochem. 2000; 69:923. [PubMed: 10966479] b) Dawson PE, Muir TW, Clark-Lewis I, Kent SBH. Science. 1994; 266:776. [PubMed: 7973629] c) Xiao J, Burn A, Tolbert TJ. Bioconjugate Chem. 2008; 19:1113.d) Gentle IE, De Souza DP, Baca M. Bioconjugate Chem. 2004; 15:658.e) Blanco-Canosa JB, Dawson PE. Angew. Chem. 2008; 120:6957. *Angew. Chem. Int. Ed.***2008**, *47*, 6851. f) Erlanson DA, Chytil M, Verdine GL. Chem. Biol. 1996; 3:981. [PubMed: 9000009]

9. a) Muir TW. Annu. Rev. Biochem. 2003; 72:249. [PubMed: 12626339] b) Yeo DSY, Srinivasan R, Chen GYJ, Yao SQ. Chem. Eur. J. 2004; 10:4664. [PubMed: 15372685] c) Muralidharan V, Muir TW. Nat. Methods. 2006; 3:429. [PubMed: 16721376]

10. a) Yeo DSY, Srinivasan R, Uttamchandani M, Chen GYJ, Zhu Q, Yao SQ. Chem. Comm. 2003:2870. [PubMed: 14680216] b) Busch GK, Tate EW, Gaffney PRJ, Rosivatz E, Woscholski R, Leatherbarrow RJ. Chem. Comm. 2008:3369. [PubMed: 18633492] c) Chattopadhaya S, Srinivasan R, Yeo DSY, Chen GYJ, Yao SQ. Bioorg. Med. Chem. 2009; 17:981. [PubMed: 18261914]

11. White EH, McCapra F, Field GF, McElroy WD. J. Am. Chem. Soc. 1961; 83:2402.

12. Baskin JM, Prescher JA, Laughlin ST, Agard NJ, Chang PV, Miller IA, Lo A, Codelli JA, Bertozzi CR. Proc. Natl. Acad. Sci. U.S.A. 2007; 104:16793. [PubMed: 17942682]

13. a) Telenti A, Southworth M, Alcaide F, Daugelat S, Jacobs WR Jr, Perler FB. J. Bacteriol. 1997; 179:6378. [PubMed: 9335286] b) Xia Z, Xing Y, So M-K, Koh AL, Sinclair R, Rao J. Anal. Chem. 2008; 80:8649. [PubMed: 18922019]
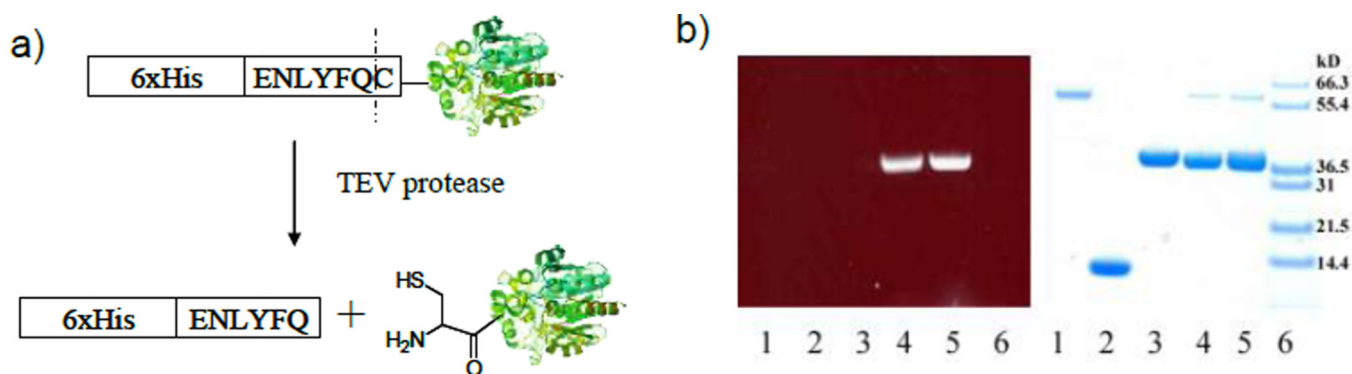
Figure 1.
In vitro labeling of N-terminal cysteine on proteins with CBT probes. a) Generation of N-terminal cysteine via protease processing. b) Fluorescence (left) and white light (right) images of a gel loaded with proteins labeled with amino-CBT (1: BSA; 2: lysozyme; 3: rLuc; 4: Cys-rLuc) or FITC-CBT (5: Cys-rLuc) and stained with Coomassie Blue. 6: size marker. All reactions were quenched with free cysteine before gel loading.
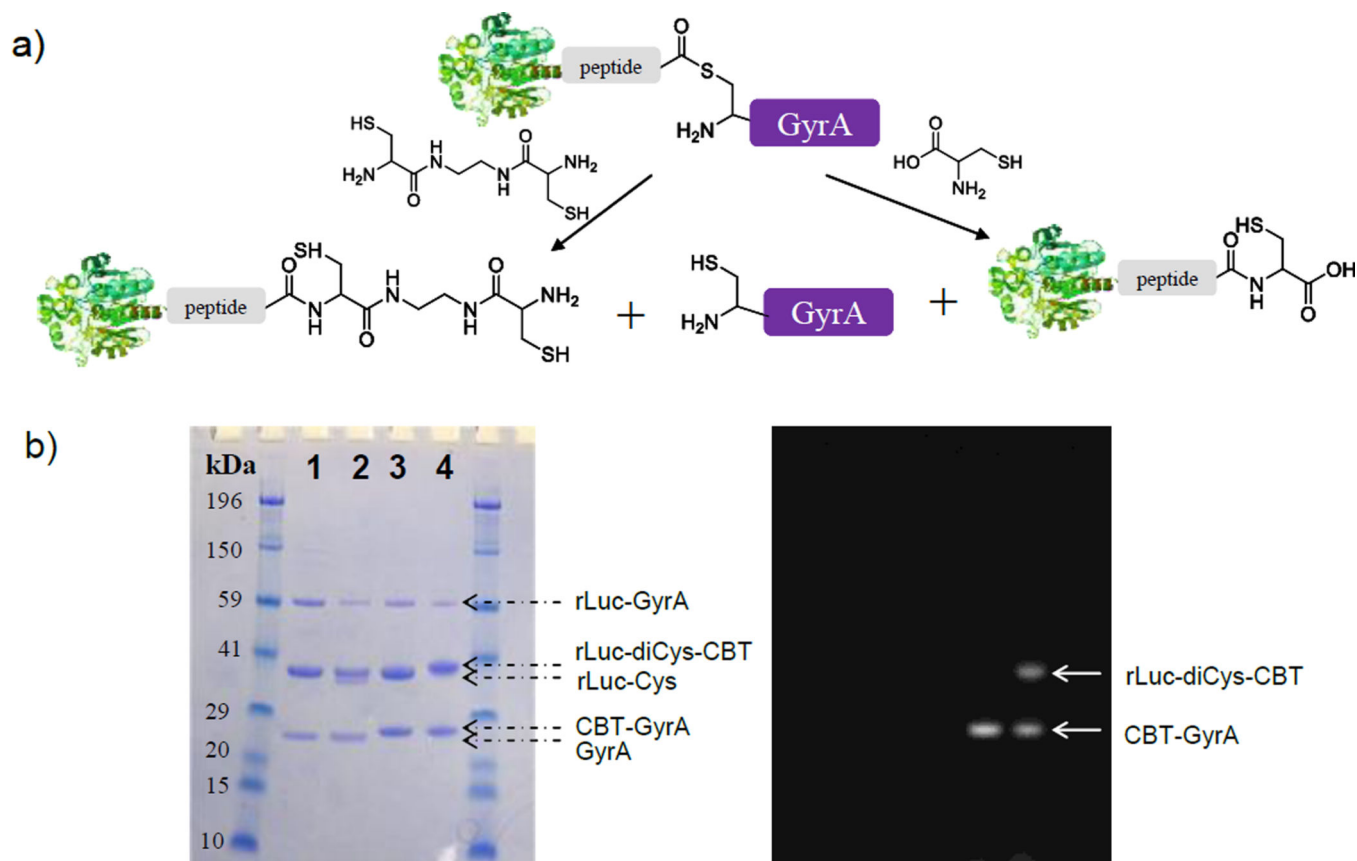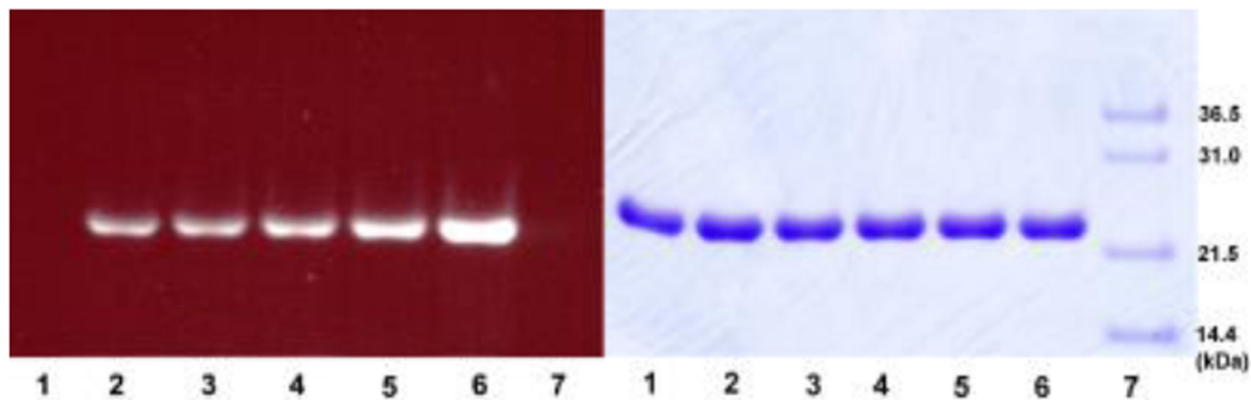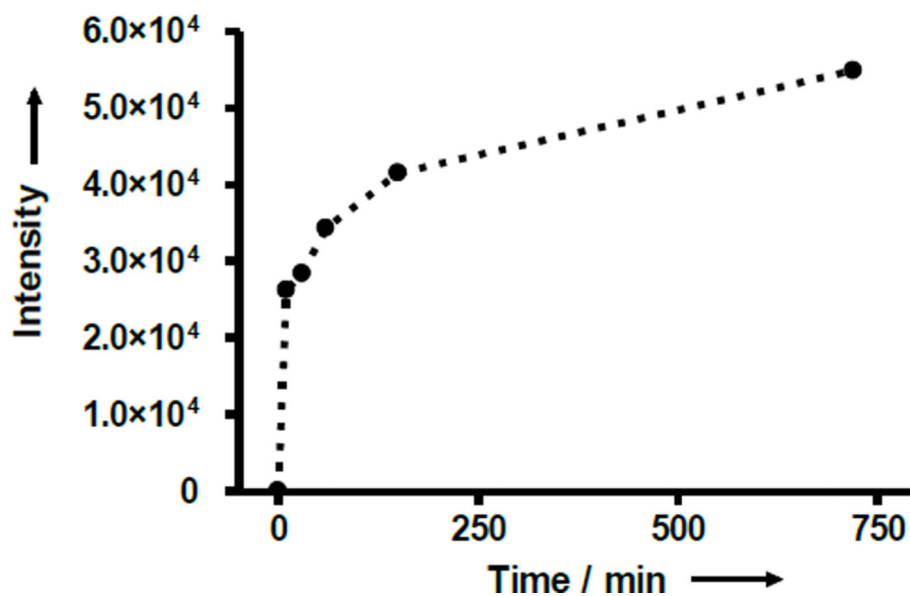
Figure 2.
In vitro labeling of cysteine introduced at the C-terminus of proteins with CBT probes. a) Introduction of free cysteine to the C-terminus of rLuc via intein-mediated splicing; a linker peptide is present between rLuc and GyrA. b) Fluorescence (right) and white light (left) images of a gel loaded with reaction solutions (1: rLuc-Cys; 2: rLuc-diCys; 3: rLuc-Cys + Peptide-FITC-CBT; 4: rLuc-diCys + Peptide-FITC-CBT) and stained with Coomassie Blue. The arrows indicate corresponding proteins.

a)

b)

Figure 3.
Time-dependent labeling of N-terminal cysteine of GyrA. a) Fluorescence (left) and white light (right) images of a gel loaded with reaction solutions containing GyrA and amino-CBT with different reaction time(1: 0; 2: 10 min; 3: 30 min; 4: 1 h; 5: 2.5 h; 6: 12 h) and stained with Coomassie Blue; 7: size marker. The reactions were quenched with cysteine before gel loading. b) Plot of integrated fluorescence band intensity vs. time.
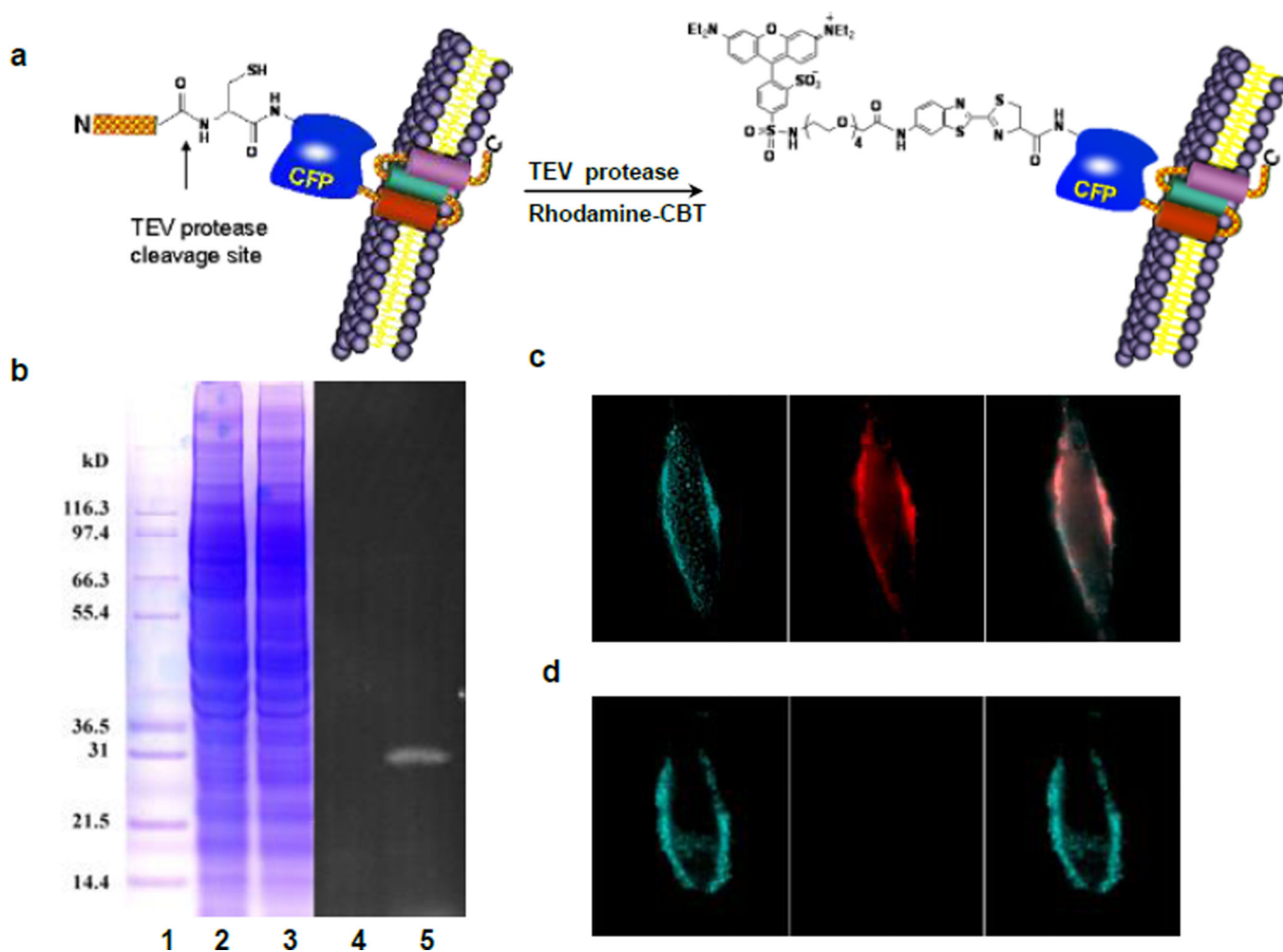
Figure 4.
Labeling of CFP on live HeLa cells with the rhodamine-CBT. a) Schematic of the labeling strategy. b) In vitro labeling of cell lysates with rhodamine-CBT, analyzed by gel electrophoresis. 1: marker; 2: cell lysates without TEV protease; 3: cell lysates with TEV protease; 4 and 5: fluorescence image of Lanes 2 and 3, respectively. c) Labeling of tagged CFP in transfected HeLa cells with TEV protease and rhodamine-CBT. 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.
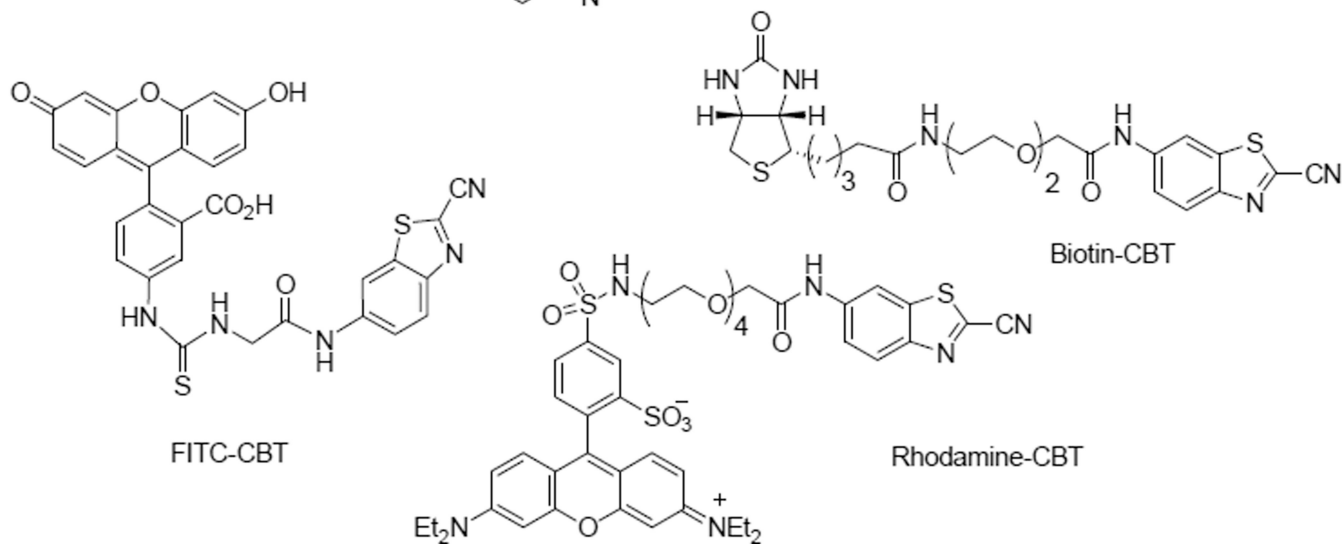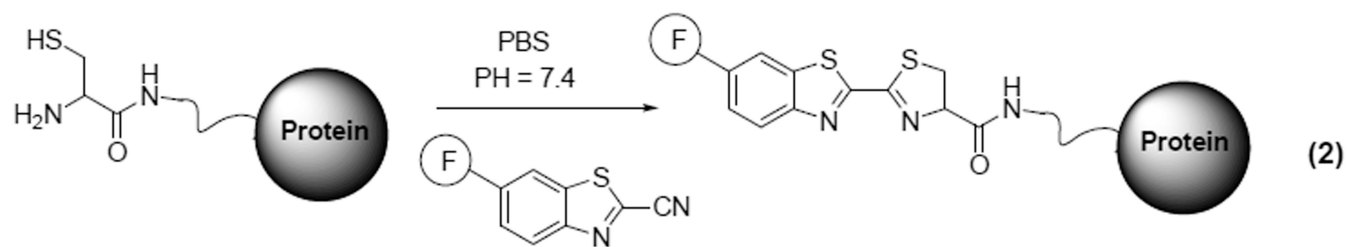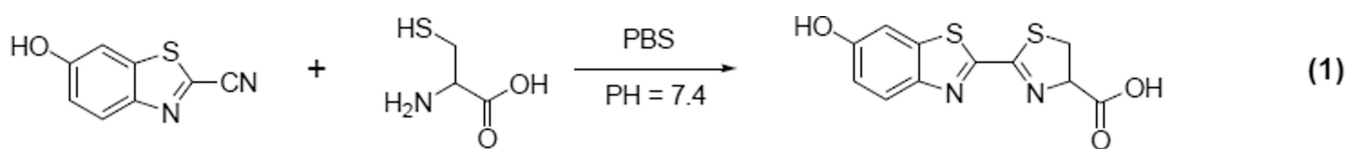
Scheme 1.
Condensation reactions between a free cysteine and CBT in the synthesis of D-luciferin and site-specific protein labeling, and structures of CBT probes used for labeling.