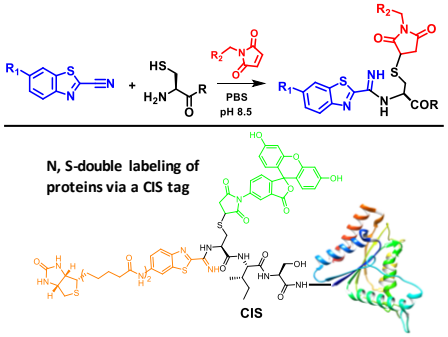
N, S-Double Labeling of N-Terminal Cysteines via an Alternative Conjugation Pathway with 2-Cyanobenzothiazole

Wenjian Wang and Jianmin Gao*

Department of Chemistry, Boston College, Merkert Chemistry Center, 2609 Beacon Street, Chestnut Hill, MA 02467.

*Correspondence should be sent to jianmin.gao@bc.edu



Abstract

Conjugation of 2-cyanobenzothiazole (CBT) with N-terminal cysteines (NCys) typically gives a luciferin product. We herein report an alternative reaction pathway leading to an N-terminal amidine rendering the side chain thiol available for further modification. Examination of peptide sequence dependence of this amidine conjugation reveals a tripeptide tag CIS that allows facile N, S-double labeling of a protein of interest with over 90% yield. This alternative reaction pathway of CBT-NCys condensation presents a significant addition to the toolbox for site-specific protein modifications.

Targeted and site-specific modification of proteins enables facile incorporation of reporters to facilitate studying protein function in complex biological milieu. Protein modification with precision is also highly desirable for protein engineering towards protein-based therapeutics.[1-3] While a number of publications document such protein modifications via the use of unnatural amino acids,[4] it remains a challenge to modify native proteins in a targeted and site-specific manner,[5-9] which is however most desirable as it would avoid the use of sophisticated technologies for unnatural amino acid incorporation. Towards this end, cysteine, as one of the least abundant amino acids, has been frequently used to label purified proteins with site specificity.[5] Again by taking advantage of its low abundance in a proteome, Tsien and colleagues developed a tetra-cysteine motif (Cys-Cys-Pro-Gly-Cys-Cys) to enable targeted and site-specific modification of a protein of interest in cells.[10] More recently, Pentelute and co-workers reported that a "π-clamp" (Phe-Cys-Pro-Phe) tag that can be selectively labelled by a perfluoroaromatic probe.[11] These two pieces of elegant work highlight the potential of using the microenvironment to fine tune cysteine reactivity for site-specific ligations.

In contrast to these peptide tags, a single cysteine residue placed at the N-terminus can serve as a handle for site-specific labeling of a target protein as well. The unique reactivity of N-terminal cysteines (NCys) has given rise to the native and expressed protein ligation technologies (Figure 1a),[12,13] which allow chemical synthesis of proteins for various applications. Furthermore, several chemotypes are found to selectively conjugate with an NCys in complex biological

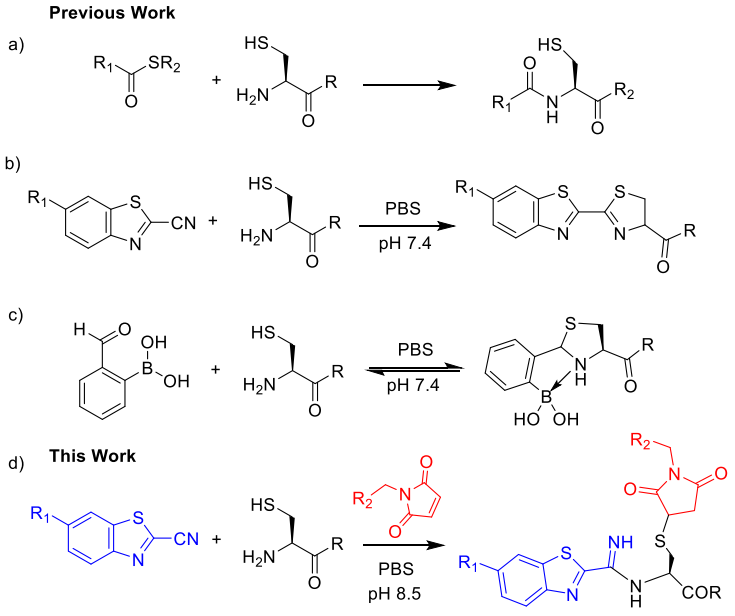


Fig. 1 Chemoselective conjugation reactions of N-terminal cysteines. a) Native chemical ligation; b) CBT-NCys condensation; c) Thiazolidino boronate formation; d) N, S-double labeling of NCys.

media (Figure 1b, c).[14-16] A prominent example is 2-cyanobenzothiazole (CBT), which is known to rapidly conjugate with an NCys under physiologic conditions to give a luciferin product. While this luciferin-producing conjugation has proven to be a powerful tool for protein labeling,[14,17-22] we report herein an alternative reaction pathway (Figure 1d) in which CBT conjugates with an NCys to give an N-terminal amidine product leaving the side chain thiol for secondary labeling. This N, S-double labeling mechanism further expands the utility of CBT as a chemotype for NCys-mediated protein modifications.

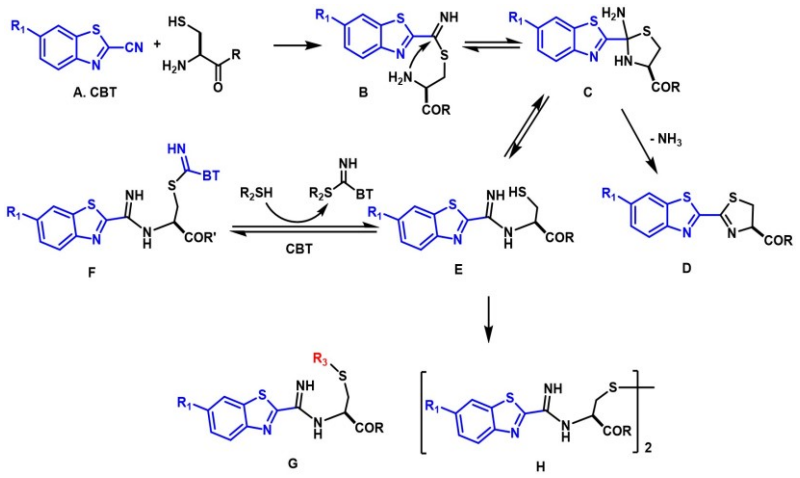


Fig 2. Postulated reaction pathways of CBT-NCys conjugation. The CBT-NCys conjugation is initiated with the thiol addition into the cyano functionality to give a thioimidate **B**, which then reacts intramolecularly with the NCys amine to give the 2-aminothiazolidine **C**. Elimination of **C** can proceeds via three possible routes: elimination of exocyclic amine gives the luciferin product **D**, elimination of the endocyclic amine gives back **C**, and elimination of the thiol gives the amidine product **E**. With a free thiol, **E** could capture another CBT to give a bis-CBT adduct **F**. Alternatively, the free thiol of **E** could be captured by an alkylating reagent to give a stable product **G** or oxidize to give a dimer **H**.

The CBT-NCys condensation initiates with the NCys side chain thiol attacking the CBT cyano group to give a thioimidate **B**, which then cyclizes to afford a 2-aminothiazolidine intermediate **C** (Figure 2). Elimination of the exocyclic amine of **C** gives the luciferin product **D**.[23] While previous reports primarily focused on neutral conditions that favours luciferin formation (*vide infra*), we serendipitously discovered a new molecular species when we examined the conjugation of a short peptide CIY with CBT under varied pH conditions. Specifically, mixing CIY with five equivalents of CBT at pH 8.5 gave 50% of the luciferin product **D** according to LC-MS analysis (Figure 3a, top). An additional product was observed that corresponds to a bis-CBT modified peptide **F** (Figure S1). ^{1}H-NMR characterization of the reaction gave consistent results: two new sets of aromatic peaks were observed in addition to the CBT resonances (Figure

S2). Mechanistically speaking, this bis CBT-labelled product can result from CBT conjugation with the amidine product **E**, which arises from the elimination of the side chain thiol of the 2-aminothiazolidine intermediate **C**. Through this alternative elimination pathway of **C**, the CBT moiety is passed along from the side chain thiol to main chain amine under our experimental conditions. To the best of our knowledge, this is the first documentation of amidine formation in CBT-NCys conjugation, although a similar thioimidate-to-amidine transformation was recently reported between internal cysteine and lysine side chains by Bertozzi and coworkers.[24] The formation of the bis adduct **F** is further supported by treating the conjugate with 1 mM free cysteine, which triggered the disappearance of **F** and increased formation of **D** according to LC-MS analysis (Figure 3a, bottom). This result is consistent with the labile nature of the thioimidate moiety in **F**. It also indicates that the amidine product **E** can revert to **C** and further transform into the luciferin product **D**. The reversibility of these reaction pathways presents a previously underappreciated aspect of the CBT-NCys conjugation chemistry as well.

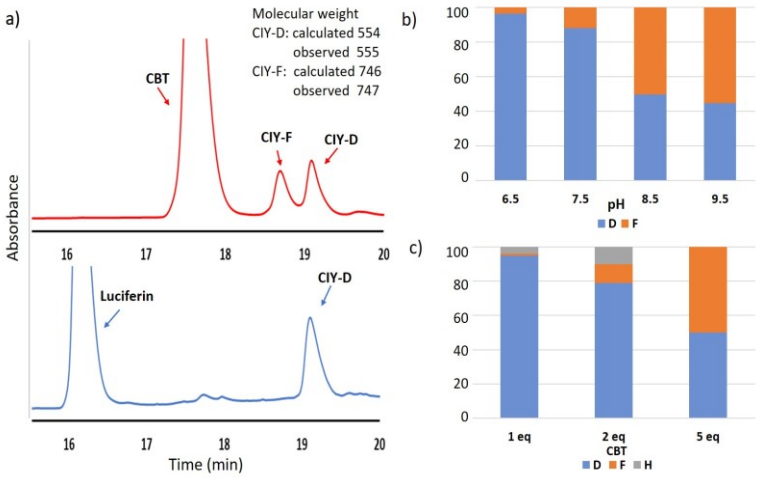


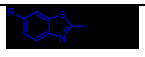
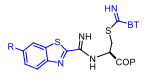
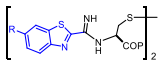
Fig 3. Characterization of the reaction between CIY and CBT. (a) LC-MS analysis of 100 µM CIY treated with 5 eq CBT at pH 8.5 for 2 hr (top) and then with 5 eq cysteine (bottom). (b) pH dependence of CIY-F formation. 100 µM CIY was incubated with 5 eq CBT at pH 6.5, 7.5, 8.5 and 9.5 respectively for 2 hr before LC-MS analysis. (c) Effect of CBT equivalency on CIY-F formation. 100 µM CIY was incubated at pH 8.5 for 2 hr with 1, 2, and 5 eq CBT respectively, and then subjected to LC-MS analysis.

To gain further insight into the luciferin versus non-luciferin reaction pathways, we tested the pH dependence of **F** formation in the CBT-CIY conjugation (Figure 3b). Interestingly, the conjugation gives predominantly the luciferin product at pH 6.5, yet the non-luciferin product increases with pH and reaches 55% at pH 9.5. In addition to pH, the

amount of the CBT used has significant effect on the outcomes of the CBT-CIY conjugation. In contrast to the bis-CBT adduct **F** seen at 50% with 5 equivalent CBT, CBT at 1 equivalent yielded only 3% **F** and 87% of the luciferin product **D**. Consistently, 2 equivalents of CBT gave 15% **F** and 73% **D** (Figure 3c). In addition, a peptide dimer **H** (~10%) was observed because of disulfide formation of **E**. The observed dimerization can be rationalized by the slower **E**-to-**F** conversion caused by the low concentrations of CBT used.

With the deepened understanding of the reaction mechanisms, we further postulated that CBT-NCys reaction outcome might be tunable with peptide sequences downstream of the NCys. To test this hypothesis, we have synthesized a panel of short peptides and analysed their conjugation with CBT (Table 1). The results revealed a significant sequence dependence in terms of reaction outcomes. Specifically, CIY conjugation with 5 equivalents of CBT afforded 50% of the bis-CBT adduct **F**. In contrast, CIA under the same experimental conditions gave exclusively the luciferin product **D**. Additional CIX mutants, in which X consists of a hydrophobic amino acid, elicited 22 to 50% **F** formation with large hydrophobic residues favouring the bis-CBT adduct. While mutating the Ile-to-Ala mutation (CIY vs. CAY) had minimal effect on **F** formation, Ile-to-Pro mutation abolished formation of **F**, giving luciferin as the sole product. Collectively, these results suggest that the hydrophobic interactions between the Tyr side chain and the CBT moiety favour the non-luciferin pathway.

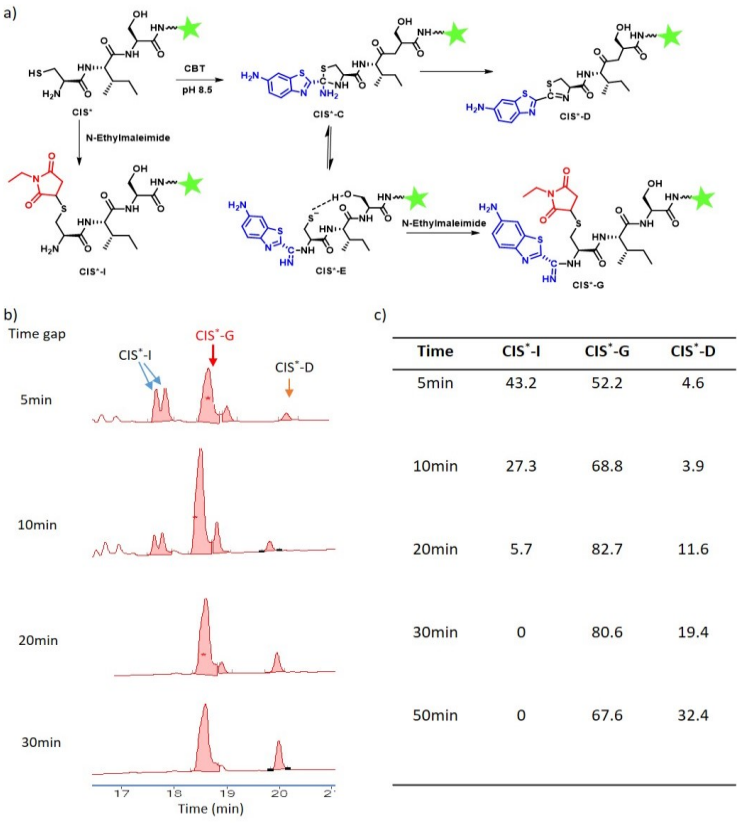
Table 1. Outcomes of CBT conjugation with various peptides.

Peptides	CIA	CIY	CIL	CIW	CIF	CICha	CIT	CAY	CPY	CY	CIS	CIDap	CIK	CVS	CAS	CPS	CS
 D	100	50	78	56	76	51	96	54	100	87	17	100	100	80	100	100	60
 F	-	50	22	43	24	49	4	46	-	13	-	-	-	-	-	-	-
 H	-	-	-	-	-	-	-	-	-	-	83	-	-	20	-	-	40

The reaction was performed in PBS buffer pH 8.5 with 100 uM peptides and 500 uM CBT. **D**, **F** and **H** (%) formation was determined by LC-MS

Surprisingly, we found that CIS (Cys-Ile-Ser), in contrast to the hydrophobic CIX variants, conjugates with 5 equivalent CBT to give only 17% of the luciferin product **D**, with the remaining product (83%) appearing as a disulfide linked dimer **H** (Figure S3, Table 1). This result indicates the CIS peptide strongly disfavours luciferin formation by

promoting thiol elimination of **C** to allow disulfide bond formation. Remarkably, CIS is quite unique in this behaviour as CIK and CI(Dap) both gave the luciferin product **D** exclusively. Mutating CIS to CVS dramatically reduced the percentage of **H** formation to 20%, and further CAS and CPS did not give **H** at all. We postulated that the CIS tripeptide adopts a unique structure that favours hydrogen bonding interaction between the Ser side chain and the thiolate and this stabilization favours thiol elimination over luciferin formation (Figure 4a). The proposed mechanism is supported by the sharp pH dependence of the CIS-CBT conjugation: the disulfide dimer **H** was observed at only 7% (93% luciferin product **D**) at pH 7.5, in contrast to 83% at pH 8.5 (Figure S4). In comparison, a much shallower pH dependence was observed for the CBT conjugation with CIY, which gave the bis adduct CIY-F as the main product (Figure S4).



Time	CIS*-I	CIS*-G	CIS*-D
5min	43.2	52.2	4.6
10min	27.3	68.8	3.9
20min	5.7	82.7	11.6
30min	0	80.6	19.4
50min	0	67.6	32.4

Fig 4. N, S-double labeling of a CIS peptide. (a) Postulated reaction pathways of CBT conjugation with a CIS* peptide leading to varied reaction outcomes. (b) LC-MS characterization of CIS* (100 µM) sequentially treated with CBT (200 µM) and N-ethylmaleimide (100 µM) with varied time gap. CIS*-I appeared as a pair of peaks presumably due to the generation of diastereomers at the newly generated chiral centre on the succinimide moiety of CIS*-I. (c) Percentages of product composition of the reactions shown in (b).

We envisioned that the transiently formed **E**, instead of forming a disulfide or conjugating with another CBT, could be captured by proper electrophiles to afford an N, S-doubly labelled NCys (Figure 4a). To test this hypothesis, we included TCEP in the reaction mixture to avoid disulfide formation, and reduced the stoichiometry of CBT to 2 equivalents to minimize formation of the bis-CBT adduct **F**. To facilitate analysis, we synthesized a fluorophore labelled peptide with CIS on its N-terminus (CIS*, Figure 4a, Figure S5). N-ethylmaleimide was first used as a test reagent to capture the free thiol. Due to the necessity of the NCys thiol for the initiation of the CBT conjugation, N-ethylmaleimide (1 equivalent) was added a later time to allow CBT conjugation with NCys. Not surprisingly, the time gap between reaction initiation and N-ethylmaleimide addition is critical (Figure 4b): adding N-ethylmaleimide too early (5 min) resulted in significant formation of CIS*-**I** (43%), which is no longer able to conjugate with CBT. Adding N-ethylmaleimide too late (50 min) results in increased presence of the luciferin product CIS*-**D**. 20 to 30 min of time gap appeared to be optimal, yielding over 80% of the N,S- double labelled product CIS*-**G**, a small percentage of the luciferin product CIS*-**D**, and little to no CIS*-**I** (Figure 4c). Use of N-ethylmaleimide at 4 equivalents increased the yield of CIS*-G even further to 91% (Figure S6). These results suggest a $t_{1/2}$ of ~5min for the amidine formation. Based on this, we estimate a rate constant of ~20 M^{-1} s^{-1} for the amidine formation, which is on par with the CBT conjugation to give luciferins.[12] Interestingly, the efficiency of CIS* double labeling was only slightly reduced with the addition of 2 equivalents of CIA as an competitor (Figure S7), indicating CBT preferentially conjugates with CIS* over other NCys residues. Finally and importantly, we found that the doubly labelled product CIS*-G displays robust stability under physiologic conditions with no decomposition over 24 hours, even in the presence of 10 mM glutathione (Figure S8).

To demonstrate the utility of this double labeling strategy for protein modification, a CIS tag was introduced at the N-terminus of a model protein, *E. coli* azoreductase (AzoR, Figure 5a). Specifically, a TEV protease recognition epitope (ENLYFQ↓CIS) was fused to the N-terminus of AzoR. TEV protease cleavage under the dialysis conditions yielded AzoR with an N-terminal CIS tag (CIS-AzoR), which was purified by gel filtration and confirmed with ESI-MS (Figure S9). For N, S-double labeling, the protein was treated with biotin-CBT (Scheme S1) at pH 8.7 for 20 min and then fluorescein-5-maleimide was added. LC-MS analysis showed that the major product (>90%) corresponds to the doubly labelled CIS-AzoR with a minimal luciferin product (Figure 5b, Figure S10). In comparison, only the luciferin

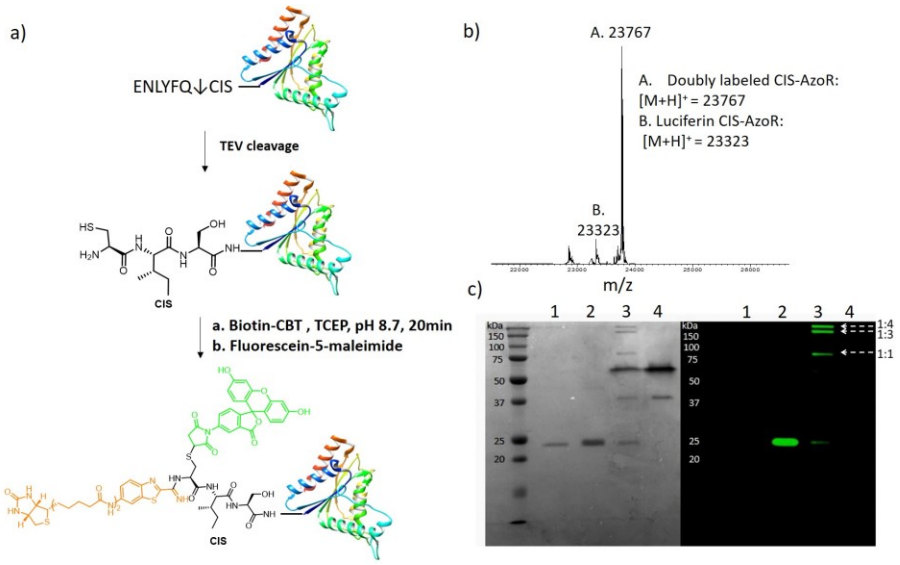


Fig 5. N, S-double labeling of proteins with an N-terminal CIS tag. (a) Schematic illustration of the preparation and labeling of CIS-AzoR. (b) Deconvoluted mass spectrum of the modified CIS-AzoR. Peak A corresponds to the doubly labelled CIS-AzoR with the calculated mass 23766 and observed mass 23767. Peak B corresponds to luciferin CIS-AzoR with the calculated mass 23322 and observed mass 23323 (c) Fluorescence (right) and Coomassie-staining (left) images of a gel loaded with CIS-AzoR after sequential labeling by Biotin-CBT and fluorescein-5-maleimide. Lane 1: CIS-AzoR; lane 2: CIS-AzoR subjected to double labeling conditions; lane 3: CIS-AzoR subjected to double labeling and then mixed with streptavidin; lane 4: streptavidin alone. Lane 3 shows the formation of 1:1, 1:3, and 1:4 complexes between streptavidin and the doubly labelled CIS-AzoR.

product was observed for the negative control Cys-AzoR, which presents a single cysteine instead of the CIS tag on the N-terminus (Figure S11). The high efficiency of N, S-double labeling was further corroborated using a gel shift assay, in which the labelled CIS-AzoR was incubated with streptavidin and then subjected to SDS-PAGE analysis. The fluorescence image of the gel revealed little protein remaining at the expected molecular weight of AzoR. Instead, the protein predominantly (>90%) appeared as complexes with streptavidin (Figure 5c). This result confirms efficient double labeling of CIS-AzoR as only the biotinylated protein can bind streptavidin to form complexes.

To better define the scope of applications of this NCys double labeling strategy, we tested whether the double labeling protocol would have any negative impact on the enzymatic activity of AzoR. The *E. coli* AzoR is an FMN-dependent NADH-azoreductase, which can degrade methyl red.[25] We determined the enzyme activity by monitoring the decrease in absorbance at 430 nm. To our satisfaction, the doubly labelled CIS-AzoR showed non-compromised enzymatic activity (73.2 units/mg) in comparison to the unmodified CIS-AzoR (68.7 units/mg) (Figure S13). These

values are consistent with a prior literature report, where the wild type enzyme activity was estimated to be 63.9 units/mg.[25] These results show that the N, S- double labeling protocol is well tolerated by protein structures, which suggests potentially broad applications of this chemistry.

In summary, this contribution describes a new strategy for N-terminal cysteine (NCys) modification that allows labeling of the main chain amine and the side chain thiol with two different functional groups. The N, S- double labeling is accomplished by re-engineering the reaction landscape of the NCys conjugation chemistry with CBT (caynobenzothiazole). While CBT-NCys conjugation typically gives a luciferin product, we report an alternative reaction pathway that proceeds under mildly basic conditions and leads to rapid formation of an N-terminal amidine, leaving the side chain thiol free for further modification. Exploration of the peptide sequence dependence reveals that a tripeptide tag CIS strongly favours the amidine conjugation pathway, which enables N, S-double labeling of peptides and proteins in excellent yields. The utility of the CBT-CIS conjugation is demonstrated by facile double labeling of azoreductase (AzoR) as a model protein. Remarkably, the mild labeling conditions elicited little reduction of the enzymatic activity of AzoR. We note that this alternative pathway of CBT-NCys conjugation displays similarity to the native chemical ligation chemistry, which also yields a free side chain thiol for further modification. However, the CBT-induced amidine formation proceeds with much faster kinetics than native chemical ligation, which typically gives a rate constant around 0.1 M^{-1} s^{-1} (Reference 26), in contrast to 20 M^{-1} s^{-1} estimated for the amidine formation in this work. Furthermore, the thioesters needed for native chemical ligation are often prone to hydrolysis,[27] while CBT enjoys robust stability under physiologic conditions. These characteristics make the CBT mediated NCys double labeling unique and appealing for site-specific labeling of proteins.

Experimental Section

General Methods

6-Amino-2-cyanobenzothiazole (CBT), N-ethylmaleimide and fluorescein-5-maleimide were purchased from Fischer. ProTEV Plus was purchased from Promega Corporation. Slide-A-Lyzer MINI Dialysis Devices were purchased from Thermo SCIENTIFIC. Illustra NAP-5 columns were obtained from GE Healthcare for desalting and buffer exchange. Q5® Site-Directed Mutagenesis Kit was obtained from New England biolabs (NEB). Primers were ordered from Integrated DNA Technologies (IDT). Rink Amide

MBHA resin and HBTU were purchased from Novabiochem (San Diego, CA). Fmoc-protected amino acids were purchased from Advanced Chemtech (Louisville, KY) or Chem-Impex Int'l Inc (Wood Dale, IL). Peptide synthesis was carried out on a Tribute peptide synthesizer from Protein Technologies. A Nanodrop UV-vis spectrometer was used to measure the concentration of the fluorescein labelled compound. NMR data were collected on VNMRS 500 & 600 MHz NMR spectrometers. LC-MS data were generated by an Agilent 6230 LC TOF mass spectrometer. The gradients used for LC-MS analysis are listed below.

Method I for LC-MS (Agilent 6230 LC TOF): Agilent Extend C18 (1.8 μm, 2.1 × 50 mm) analytical column using water-acetonitrile mobile phase with 0.1% formic acid and a flow rate of 0.2 mL/min. Gradient used: isocratic 5% CH_3CN for 1 min (0-1 min), then gradient from 5% to 95% CH_3CN over 9 min (1-9 min), then gradient from 95% to 5% CH_3CN over 0.01 min (9-9.01 min), then isocratic 5% CH_3CN for 7 min (9.01-16 min).

Method II for LC-MS (Agilent 6230 LC TOF): Aeris^{TM} XB-C8 (3..6 μm, 4.6 × 100 mm) analytical column using water-acetonitrile mobile phase with 0.1% formic acid and a flow rate of 0.2 mL/min. Gradient used: isocratic 5% CH_3CN for 5 min (0-5 min), then gradient from 5% to 95% CH_3CN over 19min (5-24 min), 95% CH_3CN for 3 min (24-27 min), then gradient from 95% to 5% CH_3CN over 1 min (27-28 min), then isocratic 5% CH_3CN for 5 min (28-33 min).

Method III for LC-MS (Agilent 6230 LC TOF): Aeris^{TM} XB-C8 (3..6 μm, 4.6 × 100 mm) analytical column using water-acetonitrile mobile phase with 0.1% formic acid and a flow rate of 0.2 mL/min. Gradient used: isocratic 5% CH_3CN for 3 min (0-3 min), then gradient from 5% to 95% CH_3CN over 15 min (3-18 min), then gradient from 95% to 5% CH_3CN over 2 min (18-20 min), then isocratic 5% CH_3CN for 5 min (20-25 min).

Conjugation of CBT and CXX peptides

Short peptides were synthesized on Rink Amide MBHA resin and purified by HPLC. 1 μL of 50 mM CBT stock solution in DMF was added into 100 μL of 100 μM CXX peptides solution in PBS buffer (final CBT concentration was 500 μM). The reaction mixture was adjusted to a desired pH to initiate the reaction and then checked again at the end of the reaction to ensure steady pH throughout the reaction. Unless noted otherwise, all reactions presented in this contribution were performed at room temperature. The reaction was analyzed by LC-MS (Method I) after 2 hr incubation. To facilitate quantification, CIS-D and CIS-H were purified by HPLC and used as representative compounds to determine the extinction coefficient of the luciferin and amidine conjugates respectively. Percentages of the conjugation products were calculated by the UV absorbance at 280 nm using molar extinction coefficients ($M^{-1}cm^{-1}$) as follows: Tyrosine, 1280; Tryptophan, 5690; Luciferin conjugate, 2964; CBT amidine conjugate, 2221. CIS-D: HRMS (ESI-

TOF) m/z: $[M + H]^+$, Calcd for $C_{20}H_{27}N_6O_4S_2^+$ 479.1530, found 479.1659; CIS-H: HRMS (ESI-TOF) m/z: $[M + H]^+$, Calcd for $C_{40}H_{57}N_{14}O_8S_4^+$ 989.3361, found 989.3536.

Protocol for double labeling of a model CIS peptide CISAERGDap-FAM

Solid phase peptide synthesis was performed on a rink amide resin using Fmoc chemistry. An alloc-protected diaminopropionic acid (Dap) residue was installed at the C-terminus for on-resin coupling of the fluorophore, 5(6)-FAM. The peptide was purified by HPLC. For double labeling, 100 μM peptide solution with 1 mM TCEP in PBS buffer pH 8.5 was prepared. To 100 μL peptide solution, 1 μL of 20 mM CBT stock solution in DMF was added (final concentration was 200μM). After a time gap specified in individual figures, N-ethylmaleimide (50 mM stock solution in DMF) was added to a desired final concentration. The reactions were then incubated for 1 hr before LC-MS analysis (Method II).

Preparation of azoreductase with N-terminal cysteine (Cys-AzoR) and CIS tag (CIS-AzoR)

pET28a-TEV-Cys-AzoR was obtained as a gift from Prof. Hua Lu at Peking University. The protein has an N-terminal GMENLYFQ↓C peptide sequence and a C-terminal His_6 tag. The plasmid was transformed into BL21(DE3) competent *E. coli*. The cells were cultured in 125 mL LB media with kanamycin (100 μg/mL) to OD_{600} = 0.5-0.8 followed by the addition of isopropyl β-D-1-thiogalactopyr-anoside (IPTG, 1 mM). The cells were cultured at 30 °C, 250 rpm for 5 h and harvested by centrifugation (5000 g for 20 min) at 4 °C. The cell pellet was resuspended in 3 mL resuspension buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10 mM imidazole, 10% (vol/vol) glycerol) and lysed by sonication. The cell debris were removed by centrifugation (5000 g for 20 min) at 4 °C. The filtered supernatant was loaded onto Ni-NTA column, washed by 1mL washing buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 20 mM imidazole, 10% (vol/vol) glycerol)) and eluted by 600 μL elution buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 250 mM imidazole, 10% (vol/vol) glycerol). The expression yield of TEV-Cys-AzoR was 30 mg protein per liter of LB media. The molecular weight of the protein was confirmed by LC-MS: calculated 23624, found 23624.

The plasmid of TEV-CIS-AzoR was constructed by Q5® Site-Directed Mutagenesis Kit. The forward primer sequence was: ATTTCTGGTATGAGCAAGGTATTAG. The reverse primer sequence was: ACATTGGAAGTACAGGTTC. The protein was produced in the same way with the yield of 72 mg per liter of LB media. The molecular weight of the protein was confirmed by LC-MS: calculated 23824, found 23824.

The TEV protease cleavage of TEV-Cys-AzoR and TEV-CIS-AzoR was done under the dialysis conditions. In a 1000MWCO dialysis bag, 0.5mg protein and 50U of Pro TEV Plus protease were mixed in 100μL 1XPBS buffer pH 7.4 containing EDTA (0.5 mM), and DTT (1 mM). The dialysis bag was gently stirred in 300mL dialysate (1XPBS buffer pH 7.4, 0.5mM EDTA, and 1mM

DTT) for 3h at room temperature. The reaction was monitored by LC-MS and showed quantitative conversion. The proteins were further purified by gel filtration (illustra NAP-5 columns). The molecular weight of the Cys-AzoR was confirmed by mass-spec: calculated 22640, found 22641. The molecular weight of the CIS-AzoR was confirmed by mass-spec: calculated 22840, found 22840.

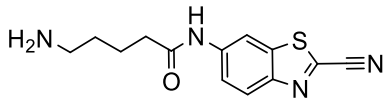
Protocol for double labeling of CIS-AzoR

1 μL of 5 mM Biotin-CBT stock solution in DMF was added into 20 μL of 18 μM CIS-AzoR solution in PBS buffer pH8.7 with 1 mM TCEP. After 20 min, 1 μL of 8 mM N-ethylmaleimide or Fluorescein-5-maleimide was added. The reaction was incubated for 1 hr and then subjected to LC-MS analysis (Method III).

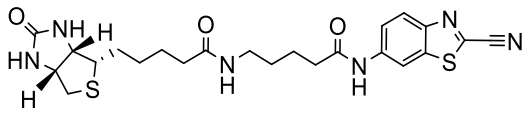
Azoreductase activity assay

The assay was done according to the reference 23 with following modifications. The activity of AzoR was determined in 500μL of 25 mM TrisHCl (pH 7.4) buffer containing 25 μM methyl red, 0.1 mM NADH, 20 μM FMN, and enzyme (0.2μg of CIS-AzoR, 0.1μg of dual labeled CIS-AzoR). The enzyme was added to initiate the reaction. The decrease in absorbance at 430 nm in the first 1 min at room temperature was used to calculate the initial reaction rate (the molar absorption coefficient: 23360 $M^{-1}cm^{-1}$). One unit of methyl red reductase activity was defined as the amount catalyzing the degradation of 1μmol of methyl red/min at room temperature.

Synthesis of Biotin-CBT



5-amino-N-(2-cyanobenzo[d]thiazol-6-yl)pentanamide ***(a)*** 52mg 5-((tert-butoxycarbonyl)amino)pentanoic acid (0.24mmol) was dissolved in 3mL THF with 38μL NMM (0.35mmol).[1] 25μL isobutyl chloroformate (0.19mmol) was added at 0ºC under N_2 and the reaction was stirred for 30min on ice. 20mg 6-amino-CBT was added into the reaction mixture and stirred for 2h on ice and then room temperature overnight. Saturated sodium bicarbonate solution was added to the reaction. The reaction mixture was extracted with ethyl acetate, which was washed by brine, dried by Na_2SO_4 and evaporated. The crude product was dissolved in 2mL 20% TFA in DCM and stirred for 2h at room temperature. TFA and DCM was evaporated. 5-amino-N-(2-cyanobenzo[d]thiazol-6-yl)pentanamide was purified by HPLC. ^{1}H NMR (600 MHz, methanol-d3): 8.55 (s, 1H), 8.01 (d, J = 9.0 Hz, 1H), 7.60 (d, J = 9.0, 1H), 3.22 (s,2H), 2.91 (t, J = 7.2 Hz, 2H), 2.45 (t, J = 6.9 Hz, 2H), 1.77 – 1.63 (m, 4H). $^{13}C\{^1H\}$ NMR (150 MHz, methanol-d3): 175.3, 150.9, 142.1, 139.3, 137.8, 127.2, 123.3, 115.3, 113.7, 41.7, 38.2, 29.3, 24.4. HRMS (ESI-TOF) m/z: $[M + H]^+$ Calcd for $C_{13}H_{15}N_4OS^+$ 275.0961; Found 275.0940.



Biotin-CBT 19mg tert-butyl (5-((2-cyanobenzo[d]thiazol-6-yl)amino)-5-oxopentyl)carbamate (0.07mmol) and 25.8uL DIPEA (0.14mmol) were dissolved in 1mL DMF. 23.6mg Biotin-NHS (0.07mmol) was added and stirred for 2h at room temperature. Biotin-CBT was purified by HPLC. ^{1}H NMR (600 MHz, DMSO-d6): 10.40 (s, 1H), 8.75 (s, 1H), 8.18 (d, J = 9.1 Hz, 1H), 7.79 (t, J = 5.6 Hz, 1H), 7.74 – 7.70 (m, 1H), 6.41 (s, 1H), 6.35 (s, 1H), 4.28 (dd, J = 7.1, 5.7 Hz, 1H), 4.11 (dd, J = 8.4, 3.6 Hz, 1H), 3.07 (dt, J = 13.6, 6.5 Hz, 3H), 2.79 (dd, J = 12.4, 5.1 Hz, 1H), 2.55 (d, J = 12.4 Hz, 1H), 2.39 (t, J = 7.4 Hz, 2H), 2.05 (t, J = 7.4 Hz, 2H), 1.66 – 1.56 (m, 3H), 1.53 – 1.40 (m, 5H), 1.36 – 1.21 (m, 2H). ^{13}C{^{1}H} NMR (150MHz, DMSO-d6) δ 171.6, 171.6, 162.4, 147.2, 139.6, 136.5, 134.5, 124.5, 120.4, 113.3, 110.7, 60.8, 58.9, 55.2, 37.8, 35.8, 35.0, 28.5, 27.9, 27.8, 25.1, 22.2. HRMS (ESI-TOF) m/z: [M + H]$^{+}$ Calcd for $C_{23}H_{29}N_6O_3S_2^{+}$ 501.1737; Found 501.16914.

Associated content

Supporting Information: NMR spectrum of Biotin-CBT, additional characterization data (NMR, LC-MS) for the conjugation reactions of peptides and proteins.

Author information

Corresponding Author *E-mail: jianmin.gao@bc.edu.

ORCID

Wenjian Wang: 0000-0002-1880-3050

Jianmin Gao: 0000-0001-9341-1287

Acknowledgements

The authors thank Professor Hua Lu at Peking University for providing the plasmid of AzoR. The financial support of this project is provided by NSF grant CHE-1904874 to JG.

References

1 Lambert, J. M.; Berkenblit, A. Antibody–drug conjugates for cancer treatment. *Ann. Rev. Med.* **2018**, *69*, 191-207.

2 Akkapeddi, P.; Azizi, S.-A.; Freedy, A. M.; Cal, P. M. S. D.; Gois, P. M. P.; Bernardes, G. J. L. Construction of homogeneous antibody–drug conjugates using site-selective protein chemistry. *Chem. Sci.* **2016**, *7*, 2954-2963.

3 Agarwal, P.; Bertozzi, C. R. Site-specific antibody–drug conjugates: the nexus of bioorthogonal chemistry, protein engineering, and drug development. *Bioconjug. Chem.* **2015**, *26*, 176-192.

4 Kim, C. H.; Axup, J. Y.; Schultz, P. G. Protein conjugation with genetically encoded unnatural amino acids. *Curr. Opin. Chem. Biol.* **2013**, *17*, 412-419.

5 Krall, N.; da Cruz, F. P.; Boutureira, O.; Bernardes, G. J. L. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **2015**, *8*, 103.

6 Matos, M. J.; Oliveira, B. L.; Martínez-Sáez, N.; Guerreiro, A.; Cal, P. M. S. D.; Bertoldo, J.; Maneiro, M.; Perkins, E.; Howard, J.; Deery, M. J.; Chalker, J. M.; Corzana, F.; Jiménez-Osés, G.; Bernardes, G. J. L. Chemo-and regioselective lysine modification on native proteins. *J. Am. Chem. Soc.* **2018**, *140*, 4004-4017.

7 Adusumalli, S. R.; Rawale, D. G.; Singh, U.; Tripathi, P.; Paul, R.; Kalra, N.; Mishra, R. K.; Shukla, S.; Rai, V. Single-site labeling of native proteins enabled by a chemoselective and site-selective chemical technology. *J. Am. Chem. Soc.* **2018**, *140*, 15114-15123.

8 Chen, D.; Disotuar, M. M.; Xiong, X.; Wang, Y.; Chou, D. H-C. Selective N-terminal functionalization of native peptides and proteins. *Chem. Sci.* **2017**, *8*, 2717.

9 Dai, Y.; Weng, J.; Chen, H.; Lin, Q.; George, J.; Wang, J.; Royzen, M.; Zhang, Q. Three-Component Protein Modification Using Mercaptobenzaldehyde Derivatives. *Org. Lett.* **2019**, *21*, 3828.

10 Griffin, B. A.; Adams, S. R.; Tsien, R. Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **1998**, *281*, 269-272.

11 Zhang, C.; Welborn, M.; Zhu, T.; Yang, N. J.; Santos, M. S.; Van Voorhis, T.; Pentelute, B. L. π-Clamp-mediated cysteine conjugation. *Nat. Chem.* **2016**, *8*, 120-8.

12 Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **1994**, *266*, 776-779.

13 Muralidharan, V.; Muir, T. W. Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat. Methods* **2006**, *3*, 429-38.

14 Ren, H.; Xiao, F.; Zhan, K.; Kim, Y. P.; Xie, H.; Xia, Z.; Rao, J. A Biocompatible Condensation Reaction for the Labeling of Terminal Cysteine Residues on Proteins. *Angew. Chem. Int. Ed.* **2009**, *48*, 9658-9662.

15 Bandyopadhyay, A.; Cambray, S.; Gao, J. Fast and selective labeling of N-terminal Cysteines via Boronic acid-accelerated conjugation. *Chem. Sci.* **2016**, *7*, 4589-4593.

16 Faustino, H.; Silva, M. J. S. A.; Veiros, L. F.; Bernardes, G. J. L.; Gois, P. M. P. Iminoboronates are efficient intermediates for selective, rapid and reversible N-terminal cysteine functionalisation. *Chem. Sci.* **2016**, *7*, 5052-5058.

17 Nguyen, D. P.; Elliott, T.; Holt, M.; Muir, T. W.; Chin, J. W. Genetically encoded 1, 2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. *J. Am. Chem. Soc.* **2011**, *133*, 11418-11421.

18 Cui, L.; Rao, J. 2-Cyanobenzothiazole (CBT) condensation for site-specific labeling of proteins at the terminal cysteine residues. *Methods Mol. Biol.* **2015**, *1266*, 81-92.

19 Cheng, Y.; Peng, H.; Chen, W.; Ni, N.; Ke, B.; Dai, C.; Wang, B. Rapid and specific post-synthesis modification of DNA through a biocompatible condensation of 1, 2-aminothiols with 2-cyanobenzothiazole. *Chemistry* **2013**, *19*, 4036-4042.

20 Wang, H. C.; Yu, C. C.; Liang, C. F.; Huang, L. D.; Hwu, J. R.; Lin, C. C. Site-selective protein immobilization through 2-cyanobenzothiazole-cysteine condensation. *Chembiochem* **2014**, *15*, 829-835.

21 Chen, K. T.; Ieritano, C.; Seimbille, Y. Early-Stage Incorporation Strategy for Regioselective Labeling of Peptides using the 2-Cyanobenzothiazole/1, 2-Aminothiol Bioorthogonal Click Reaction. *ChemistryOpen* **2018**, *7*, 256-261.

22 Patil, N. A.; Karas, J. A.; Turner, B. J.; Shabanpoor, F. Thiol-Cyanobenzothiazole Ligation for the Efficient Preparation of Peptide–PNA Conjugates. *Bioconjug. Chem.* **2019**, *30*, 793-799.

23 Zheng, Z.; Chen, P.; Li, G.; Zhu, Y.; Shi, Z.; Luo, Y.; Zhao, C.; Fu, Z.; Cui, X.; Ji, C.; Wang, F.; Huang, G.; Liang, G. Mechanistic study of CBT-Cys click reaction and its application for identifying bioactive N-terminal cysteine peptides in amniotic fluid. *Chem. Sci.* **2017**, *8*, 214-222.

24 Keyser, S. G. L.; Utz, A.; Bertozzi, C. R. Computation-Guided Rational Design of a Peptide Motif That Reacts with Cyanobenzothiazoles via Internal Cysteine–Lysine Relay. *J. Org. Chem.* **2018**, *83*, 7467-7479.

25 Nakanishi, M.; Yatome, C.; Ishida, N.; Kitade, Y. Putative ACP phosphodiesterase gene (acpD) encodes an azoreductase. *J. Biol. Chem.* **2001**, 276(49), 46394-46399.

26 Saito, F.; Noda, H.; Bode, J. W. Critical evaluation and rate constants of chemoselective ligation reactions for stoichiometric conjugations in water. *ACS Chem. Biol.* **2015**, *10*, 1026-1033.

27 Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc. Natl. Acad. Sci.* **1999**, *96*, 10068-10073.