Published in final edited form as:

Angew Chem Int Ed Engl. 2012 October 15; 51(42): 10600–10604. doi:10.1002/anie.201205352.

Genetically Encoded Cyclopropene Directs Rapid, Photoclick Chemistry-Mediated Protein Labeling in Mammalian Cells

Dr. Zhipeng Yu[†],

Department of Chemistry, State University of New York at Buffalo, Buffalo, NY 14260 (USA)

Yanchao Pan[†],

Laboratory of Noncoding RNA, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, P.R. China

Dr. Zhiyong Wang,

Department of Chemistry, State University of New York at Buffalo, Buffalo, NY 14260 (USA)

Prof. Dr. Jiangyun Wang^{*}, and

Laboratory of Noncoding RNA, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, P.R. China

Prof. Dr. Qing Lin^{*}

Department of Chemistry, State University of New York at Buffalo, Buffalo, NY 14260 (USA)

Abstract

Genetic incorporation of a cyclopropene amino acid, *N*^ε-(1-methylcycloprop-2-enecarboxamido)-lysine (CpK), into sperm whale myoglobin site-specifically in *E. coli* as well as enhanced green fluorescent protein in mammalian cells was achieved through amber codon suppression employing an orthogonal aminoacyl-tRNA synthetase/tRNA_{CUA} pair. Because of its high ring strain, cyclopropene exhibited fast reaction kinetics (up to $58 \pm 16 \text{ M}^{-1} \text{ s}^{-1}$) in the photoclick reaction and allowed rapid (~ 2 min) bioorthogonal labeling of proteins in mammalian cells.

Keywords

alkene; cyclopropene; tetrazole; bioorthogonal; proteins

The introduction of bioorthogonal organic reporters into proteins site-selectively through genetic,^[1] metabolic,^[2] or enzyme-catalyzed ligation method,^[3] in conjunction with a growing repertoire of bioorthogonal reactions,^[4] has allowed the visualization of proteins and their regulations in their native environment.^[5] Numerous small organic groups such as ketone,^[6] azides,^[7] terminal alkynes,^[8] and terminal alkenes,^[9] as well as larger reactive bioorthogonal groups such as cyclooctyne,^[10] norbornene,^[11] transcyclooctene,^[10b, 11b] tetrazole,^[12] and tetrazine^[13] have been genetically encoded for site-selective protein labeling in vivo. To track fast protein dynamics in vivo, it is imperative that these genetically encoded bioorthogonal reporters direct fast and selective bioorthogonal labeling with the cognate biophysical probes, preferably with a spatiotemporal control.

^{*}Fax: (+86) 010-64871293, jwang@ibp.ac.cn., Fax: (+1) 716-645-6963, qinglin@buffalo.edu.

[†]These two authors contribute equally to the work.

Dedicated to Prof. Andrew D. Hamilton on the occasion of his 60th birthday

Supporting information for this article is available on the WWW under <http://www.angewandte.org> or from the author.

In our continued effort to genetically encode substrates suitable for the photoclick chemistry,^[14] we envisioned that nonnatural amino acids carrying the strained alkenes may show higher rate of cycloaddition without the liability of being Michael acceptors, i.e. electron-deficient alkenes. While we reported recently that norbornene exhibited robust reactivity in the cycloaddition reaction with the macrocyclic tetrazoles,^[15] norbornene is relatively bulky and may perturb the structure of the encoded protein. Therefore, we set out to explore the genetic encoding of cyclopropene because of its small size and inherent large ring strain (54.1 kcal/mol^[16] vs. 21.6 kcal/mol for norbornene^[17]), much of which is released after the cycloaddition reaction (ring strain of cyclopropane = 28.7 kcal/mol).^[18] Here, we report the synthesis of a stable cyclopropene amino acid, the characterization of its reactivity in the photoinduced cycloaddition reaction with two tetrazoles, its site-specific incorporation into proteins both in *E. coli* and in mammalian cells, and its utility in directing bioorthogonal labeling of proteins both in vitro and in vivo.

To design a cyclopropene-containing amino acid suitable for genetic incorporation, we decided to focus on pyrrolysyl-tRNA synthetase (PylRS)/tRNA_{CUA} pair from *Methanosarcina barkeri* (Mb) because: (i) this pair is orthogonal to all the endogenous tRNAs and aminoacyl-tRNA synthetases in *E. coli* and eukaryotic cells;^[19] and (ii) many nonnatural lysine-derived amino acids have been efficiently incorporated into proteins based on this pair.^[20] In our preliminary studies, we found that 3,3-disubstituted cyclopropenes such as **1c** exhibited excellent chemical stability at room temperature. 1-Methylcycloprop-2-enecarboxylic acid (**1c**) can be expediently prepared from the commercially available starting materials ethyl 2-methylacetoacetate through a three-step procedure with an overall yield of 21% (Scheme 1). The cyclopropene carboxylic acid **1c** was then coupled with the ε-amino group of Fmoc-lysine, which upon removal of the protecting group afforded N^ε-(1-methylcycloprop-2-enecarboxamido)-lysine (CpK, **1**) in 74% yield over three steps. To our delight, crystal structure of **1c** was obtained in which a hydrogen-bonded dimer of **1c** was observed (Scheme 1; see Table S1 in Supporting Information for crystal data and structure refinement). As expected, a bond angle of mere 50° at C1-C2-C1A provides a very high angle strain in the ring structure. Importantly, the carbonyl group in **1c** is essentially perpendicular to cyclopropene double bond, preventing the conjugation between these two π systems. Perhaps as a result of this geometry, CpK was found to be very stable towards glutathione—abundant biological nucleophile inside cells: greater than 95% remained after CpK was incubated with 10 mM glutathione for more than 60 hours (Figure S1 in Supporting Information).

Using ethyl 1-methylcycloprop-2-enecarboxylate (**2**) as a model substrate, we examined the reactivity of cyclopropene with two representative tetrazoles: a 302-nm photoreactive tetrazole **3**^[21] and a water-soluble 365-nm photoreactive **4**,^[22] in acetonitrile/PBS (1:1) mixture, and compared it to those of acrylamide and norbornene (Table 1 and Figures S2-S7). In the cycloaddition reactions with tetrazole **3**, cyclopropene **2** showed a k_2 value of $58 \pm 16 \text{ M}^{-1} \text{ s}^{-1}$, similar to acrylamide (entry 1 vs. 3) but significantly faster than norbornene (entry 1 vs. 5). In the cycloaddition reactions with tetrazole **4**, however, cyclopropene **2** showed more than 10-fold drop in k_2 value, which is also 50% slower than that of acrylamide (entry 2 vs. 4) but similar to norbornene (entry 2 vs. 6). The decrease in reaction kinetics at long wavelength is likely due to the filtering effect associated with the formation of fluorescent pyrazoline adducts.^[23] Compared to other bioorthogonal reactions involving the genetically encoded alkenes, the cycloaddition of cyclopropene with tetrazole **3** is about 60 times faster than with the cycloaddition of *O*-allyltyrosine under the same conditions ($k_2 = 0.95 \text{ M}^{-1} \text{ s}^{-1}$),^[21] at least 6 times faster than the ligation between 5-norbornene-2-ol and a pyrimidine-substituted tetrazine ($k_2 = 9 \text{ M}^{-1} \text{ s}^{-1}$ in 95:5 H₂O/MeOH),^[11a] but three orders of magnitude slower than tetrazine ligation with transcyclooctene ($k_2 = 35000 \pm 3000 \text{ M}^{-1} \text{ s}^{-1}$ in biological buffer).^[11b]

To evolve an orthogonal tRNA/aminoacyl-tRNA synthetase pair that selectively charges CpK in response to TAG amber codon in proteins in *E. coli*, an *MbPylRS* library was constructed in which five residues (L266, L270, Y271, L274, and C313) were randomized through overlapping extension PCR using the synthetic oligonucleotide primers (Table S2). After three rounds of positive and two rounds of negative selections, a CpK-specific aminoacyl-tRNA synthetase, termed CpKRS, was identified. Sequencing of this clone revealed the following five mutations: L266M, L270I, Y271L, L274A, and C313I. To test whether CpKRS/*Mb*tRNA_{CUA} pair allows for efficient and selective CpK incorporation in *E. coli*, the expression of sperm whale myoglobin carrying an amber codon at position-4 and a C-terminal His tag was carried out in *E. coli* cells transformed with CpKRS/*Mb*tRNA_{CUA} pair and grown in LB medium supplemented with 1 mM CpK. The CpK-encoded mutant myoglobin proteins (Myo-CpK) were obtained at a yield of 3.0 mg/L. Importantly, Myo-CpK was produced only when CpK was added (Figure 1a), indicating that CpK incorporation is highly specific. ESI-TOF mass spectrometry showed an intact mass of 18476.0 Da for Myo-CpK (Figure S8), matching the theoretic mass of 18476.3 Da. Subsequent tryptic digestion and tandem mass spectrometry analysis confirmed the presence of CpK at position-4 (Figure 1b).

To assess whether CpK can serve as a bioorthogonal reporter for protein labeling, we incubated Myo-CpK with tetrazole **3** in PBS buffer and subjected the mixture to 302-nm photoirradiation with a handheld UV lamp for a period of 1–12 min. In-gel fluorescence analysis revealed a time-dependent appearance of a fluorescent band at the Myo-CpK spot with highest intensity reached at 10 min (Figure 2a), consistent with the formation of fluorescent pyrazoline adduct (Figure 2a).^[24] As a control, photoirradiation of a mixture of wild-type myoglobin and tetrazole **3** did not yield fluorescent bands on the SDS-PAGE gel (Figure S9), confirming that the labeling was mediated through the cyclopropene moiety. Furthermore, ESI-TOF mass spectrometry analysis of the product mixture revealed a greater than 85% conversion with a new mass peak at 18701.87 Da, matching the theoretic mass of the pyrazoline-labeled myoglobin, 18701.56 (Figure S10). Tryptic digestion and tandem mass spectrometry analysis confirmed the presence of the pyrazoline-modified lysine at position-4 by the appearance of expected masses of three fragment ions b_9^{2+} , b_{16}^{2+} , and y_{14}^{3+} (Figure 2b). For comparison, the same photoclick reaction with a mutant myoglobin encoding the recently reported norbornene-modified lysine (NorK)^[11a] at position-4^[25] showed a time-dependent, weakly fluorescent cycloadduct formation (Figure S12), affording the pyrazoline cycloadduct in 60% yield based on mass spectrometry analysis after 10-min photoirradiation (Figure S13). The slightly higher reactivity of CpK relative to NorK in the protein context is consistent with the kinetics data described in Table 1.

To examine whether CpK can direct the bioorthogonal labeling in mammalian cells, we co-transfected human embryonic kidney (HEK) 293 cells with pCMV-CpKRS plasmid in which the transcription of CpKRS is under the control of CMV promoter and the transcription of *Mb*tRNA_{CUA} is under the control of human U6 promoter,^[20c] together with pSwan-EGFP37TAG reporter.^[20d] The cells were allowed to grow in the presence 4 mM CpK for 36 hours, treated with 40 μ M tetrazole **4** for 1.5 hours followed by a brief 365-nm photoirradiation for 2 min, and finally proceeded to microscopic imaging. In the EGFP channel (ex 488 nm, em 499–578), green fluorescent cells were detected only in plates where CpK was included in the culture medium (compare panels b & e to h in Figure 3), indicating that CpKRS/*Mb*tRNA_{CUA} pair supports site-specific incorporation of CpK into EGFP37TAG in HEK293 cells. In the pyrazoline channel (ex 405 nm, em 410–498 nm), only the tetrazole **4**-treated cells expressing CpK-encoded EGFP showed the cyan fluorescence (panel a vs. d & g; see Figure S14 for fluorescence spectrum of the pyrazoline adduct). It is important to note that fluorescent images were acquired in two separate tracks with a single laser source exciting at one wavelength a time to avoid any possible

fluorescence leakage to the unintended channel. As seen in the overlaid images, the cyan fluorescent cells coincided with the cells that showed high green fluorescence (indicated by white arrows in panel c), suggesting that the labeling reaction was indeed directed by the cyclopropene moiety. However, not all green fluorescent cells were labelled, indicating some variability in tetrazole reagent penetration into the highly confluent HEK293 cells. Notably, repeated attempts to incorporate NorK in HEK293 cells using wild-type *MbPylRS* and identical transfection conditions were not successful (Figures S15 and S16), precluding the comparison of the reactivity of these two strained alkenes in photoclick chemistry in vivo.

In conclusion, we have demonstrated the genetic incorporation of a cyclopropene-containing amino acid, CpK, into target proteins site-specifically, and the use of CpK as a bioorthogonal reporter for directing rapid (~ 2 min) fluorescent labeling of the target protein in mammalian cells. Compared to other genetically encoded, bioorthogonal labeling reactions reported recently,^[10–13] the main advantage of the cyclopropene-directed photoclick chemistry lies in its potential in the spatiotemporally controlled protein labeling in mammalian cells, which requires the development of highly reactive laser-activatable tetrazole reagents using either single photon (e.g. 405 nm) or two-photon laser source; work along this line is currently in progress. Because of its small size, cyclopropene moiety such as **1c** can also be readily incorporated into small-molecule substrates and inhibitors for the study of proteomes^[26] and lipids.^[27]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge the National Institutes of Health (GM 085092 to Q.L.), the Major State Basic Research Program of China (2010CB912301 and 2009CB825503 to J.W.), and National Science Foundation of China (30870592 and 90913022 to J.W.) for financial support. We thank William Brennessel at University of Rochester for X-ray crystallography and Reyna K. Lim in Q.L. lab for plasmid preparation, Alan Siegel at Biological Sciences Imaging Facility (supported by National Science Foundation Major Research Instrumentation grant DBI-0923133) for assistance in microscopy. The crystal structure of **1c** has been deposited into the Cambridge Crystallographic Data Centre with deposit number CCDC 808108.

References

1. Liu CC, Schultz PG. *Annu Rev Biochem.* 2010; 79:413. [PubMed: 20307192]
2. (a) Saxon E, Bertozzi CR. *Science.* 2000; 287:2007. [PubMed: 10720325] (b) Dieterich DC, Hodas JJ, Gouzer G, Shadrin IY, Ngo JT, Triller A, Tirrell DA, Schuman EM. *Nat Neurosci.* 2010; 13:897. [PubMed: 20543841]
3. Fernández-Suárez M, Baruah H, Martínez-Hernández L, Xie KT, Baskin JM, Bertozzi CR, Ting AY. *Nat Biotechnol.* 2007; 25:1483. [PubMed: 18059260]
4. (a) Sletten EM, Bertozzi CR. *Angew Chem.* 2009; 121:7108. *Angew Chem Int Ed.* 2009; 48:6974. (b) Lim RK, Lin Q. *Chem Commun.* 2010; 46:1589.
5. Sletten EM, Bertozzi CR. *Acc Chem Res.* 2011; 44:666. [PubMed: 21838330]
6. Wang L, Zhang Z, Brock A, Schultz PG. *Proc Natl Acad Sci USA.* 2003; 100:56. [PubMed: 12518054]
7. Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG. *J Am Chem Soc.* 2002; 124:9026. [PubMed: 12148987]
8. (a) Deiters A, Cropp TA, Mukherji M, Chin JW, Anderson JC, Schultz PG. *J Am Chem Soc.* 2003; 125:11782. [PubMed: 14505376] (b) Nguyen DP, Lusic H, Neumann H, Kapadnis PB, Deiters A, Chin JW. *J Am Chem Soc.* 2009; 131:8720. [PubMed: 19514718] (c) Fekner T, Li X, Lee MM, Chan MK. *Angew Chem.* 2009; 121:1661. *Angew Chem Int Ed.* 2009; 48:1633.

9. (a) Zhang Z, Wang L, Brock A, Schultz PG. *Angew Chem.* 2002; 114:2964. *Angew Chem Int Ed.* 2002; 41:2840. (b) Song W, Wang Y, Qu J, Lin Q. *J Am Chem Soc.* 2008; 130:9654. [PubMed: 18593155] (c) Ai HW, Shen W, Brustad E, Schultz PG. *Angew Chem.* 2010; 122:947. *Angew Chem Int Ed.* 2010; 49:935.
10. (a) Plass T, Milles S, Koehler C, Schultz C, Lemke EA. *Angew Chem.* 2011; 123:3964. *Angew Chem Int Ed.* 2011; 50:3878. (b) Lang K, Davis L, Wallace S, Mahesh M, Cox DJ, Blackman ML, Fox JM, Chin JW. *J Am Chem Soc.* 2012; 134:10317. [PubMed: 22694658]
11. (a) Lang K, Davis L, Torres-Kolbus J, Chou C, Deiters A, Chin JW. *Nat Chem.* 2012; 4:298. [PubMed: 22437715] (b) Plass T, Milles S, Koehler C, Szymanski J, Mueller R, Wießler M, Schultz C, Lemke EA. *Angew Chem.* 2012; 124:4242. *Angew Chem Int Ed.* 2012; 51:4166. (c) Kaya E, Vrabel M, Deiml C, Prill S, Fluxa VS, Carell T. *Angew Chem.* 2012; 124:4542. *Angew Chem Int Ed.* 2012; 51:4466.
12. Wang J, Zhang W, Song W, Wang Y, Yu Z, Li J, Wu M, Wang L, Zang J, Lin Q. *J Am Chem Soc.* 2010; 132:14812. [PubMed: 20919707]
13. Seitchik JL, Peeler JC, Taylor MT, Blackman ML, Rhoads TW, Cooley RB, Refakis C, Fox JM, Mehl RA. *J Am Chem Soc.* 2012; 134:2898. [PubMed: 22283158]
14. (a) Wang Y, Lin Q. *Org Lett.* 2009; 11:3570. [PubMed: 19637915] (b) Lim RK, Lin Q. *Acc Chem Res.* 2011; 44:828. [PubMed: 21609129]
15. Yu Z, Lim RK, Lin Q. *Chem Eur J.* 2010; 16:13325. [PubMed: 21031376]
16. Bach RD, Dmitrenko O. *J Am Chem Soc.* 2004; 126:4444. [PubMed: 15053635]
17. Schleyer, PvR; William, JE.; Blanchard, KR. *J Am Chem Soc.* 1970; 92:2377.
18. Gordon MS. *J Am Chem Soc.* 1980; 102:7419.
19. (a) Blight SK, Larue RC, Mahapatra A, Longstaff DG, Chang E, Zhao G, Kang PT, Green-Church KB, Chan MK, Krzycki JA. *Nature.* 2004; 431:333. [PubMed: 15329732] (b) Ambrogelly A, Gundllapalli S, Herring S, Polycarpo C, Frauer C, Söll D. *Proc Natl Acad Sci USA.* 2007; 104:3141. [PubMed: 17360621]
20. (a) Neumann H, Peak-Chew SY, Chin JW. *Nat Chem Biol.* 2008; 4:232. [PubMed: 18278036] (b) Mukai T, Kobayashi T, Hino N, Yanagisawa T, Sakamoto K, Yokoyama S. *Biochem Biophys Res Commun.* 2008; 371:818. [PubMed: 18471995] (c) Li X, Fekner T, Ottesen JJ, Chan MK. *Angew Chem.* 2009; 121:9348. *Angew Chem Int Ed.* 2009; 48:9184. (d) Chen PR, Groff D, Guo J, Ou W, Cellitti S, Geierstanger BH, Schultz PG. *Angew Chem.* 2009; 121:4112. *Angew Chem Int Ed.* 2009; 48:4052. (e) Takimoto JK, Dellas N, Noel JP, Wang L. *ACS Chem Biol.* 2011; 6:733. [PubMed: 21545173] (f) Wang YS, Fang X, Wallace AL, Wu B, Liu WR. *J Am Chem Soc.* 2012; 134:2950. [PubMed: 22289053]
21. Wang Y, Song W, Hu WJ, Lin Q. *Angew Chem.* 2009; 121:5434. *Angew Chem Int Ed.* 2009; 48:5330.
22. Yu Z, Ho LY, Wang Z, Lin Q. *Bioorg Med Chem Lett.* 2011; 21:5033. [PubMed: 21570845]
23. Wang Y, Hu WJ, Song W, Lim RK, Lin Q. *Org Lett.* 2008; 10:3725. [PubMed: 18671406]
24. Song W, Wang Y, Qu J, Madden MM, Lin Q. *Angew Chem.* 2008; 120:2874. *Angew Chem Int Ed.* 2008; 47:2832.
25. The NorK-encoded myoglobin mutant was expressed using wildtype pYLRS and pBAD-pylT-Myo4TAG reporter in *E. coli* in the presence of 1 mM NorK. An isolated yield of 1.0 mg/L was obtained for the NorK-containing myoglobin: See Figure S11 for characterization data.
26. Martin BR, Wang C, Adibekian A, Tully SE, Cravatt BF. *Nat Methods.* 2011; 9:84. [PubMed: 22056678]
27. During the preparation of this manuscript, a report on the use of 1-methylcyclopropene as a lipid reporter appeared: Yang J, Se kut J, Cole CM, Devaraj NK. *Angew Chem Int Ed.* 2012; 51:7476.

Figure 1.

Incorporation of CpK (1) site-selectively into sperm whale myoglobin: (a) Coomassie blue stained gel of the wildtype and mutant myoglobin expressed in the absence or presence of 1 mM CpK. An isolated yield of 3.0 mg/L was obtained for Myo-CpK. (b) MS2 spectrum of the N-terminal peptide fragment containing CpK (K*). The CpK mass of 207.48 Da was calculated from the b_3 and b_4 ions (mass = $b_4 - b_3$), matching closely to the theoretic mass of 208.13 Da.

Figure 2.

Selective labeling of the CpK-encoded myoglobin via photoclick chemistry. (a) Coomassie blue stain (bottom panel) and inverted fluorescence image (top panel) of the protein gel after the reaction mixtures were resolved by SDS-PAGE. For photoclick reaction, a solution of Myo-CpK (0.5 mg/mL) and tetrazole **3** (100 μ M) in PBS buffer was exposed to 302-nm photoirradiation for a period of 1–12 min. (b) MS2 spectrum of the N-terminal peptide fragment of the pyrazoline-modified myoglobin (Myo-pyr) containing the pyrazoline-linked lysine (K*).

Figure 3.

Confocal micrographs of human embryonic kidney 293 cells transfected with the plasmids expressing CpKRS, *MbtRNA*_{CUA} and EGFP37TAG and grown in the presence of 4 mM CpK (a–f) or the absence of CpK (g–i). The cells in Panels a–c and h–j were treated with 40 μ M tetrazole **4** for 1.5 hours before 365-nm photoirradiation with a handheld UV lamp for 2 minutes. Scale bar = 200 μ m.

Scheme 1.

Synthesis of *N* ϵ -(1-methylcycloprop-2-enecarboxamido)-lysine (CpK, **1**).

Table 1

Kinetic characterization of ethyl 1-methylcycloprop-2-enecarboxylate (**2**) reactivity in photoclick chemistry in acetonitrile/PBS (1:1) mixture.^[a]

entry	alkene	tetrazole	photoactivation wavelength	product	k_2 (M ⁻¹ s ⁻¹)
1		3	302 nm	5	58 ± 16
2		4	365 nm	6	4.6 ± 1.3
3		3	302 nm	7	46 ± 8.6
4		4	365 nm	8	9.2 ± 0.7
5		3	302 nm	9	32 ± 12
6		4	365 nm	10	5.8 ± 1.4

^[a] Reactions were set up by incubating 100 μM of tetrazole and 500 μM of alkene dipolarophile in 0.5 mL PBS/ACN (1:1). The mixtures were exposed to a handheld UV lamp (302 nm: UVM-57, 0.16 AMPS; 365 nm: UVGL-25, 0.16 AMPS) for a specified time under ambient conditions. The product mixtures were analyzed by reverse phase HPLC and quantified by comparing the integration peaks to that of a product standard. The kinetic experiments were repeated three times in order to derive the average kinetic constants along with the standard deviations. See Figures S2–S7 in Supporting Information for details.