Published in final edited form as:

Chem Commun (Camb). 2018 April 26; 54(35): 4449–4452. doi:10.1039/c8cc02431f.

Genetic Encoding of 2-Aryl-5-carboxytetrazole-Based Protein Photo-cross-linkers†

Yulin Tian and Qing Lin

Department of Chemistry, State University of New York at Buffalo, Buffalo, New York 14260-3000, United States

Abstract

Three γ -heteroatom-substituted *N*-methylpyrroletetrazole-lysines (mPyTXKs) were synthesized and subsequently incorporated into proteins site-specifically via genetic code expansion. The γ -seleno-substituted derivative, mPyTSeK, showed excellent incorporation efficiency in *Escherichia coli* and allowed site-selective photo-cross-linking of the GST dimer. Furthermore, the mPyTSeK-cross-linked GST dimer can be cleaved under a mild oxidative condition. The incorporation of mPyTXKs into proteins in mammalian cells was also demonstrated. Lastly, the recombinantly expressed mPyTSeK-encoded Grb2 was shown to covalently capture its interaction partner, EGFR, in mammalian cell lysates, which was subsequently released after treatment with H₂O₂.

Graphical Abstract

We report the genetic encoding of three γ -heteroatom-substituted *N*-methylpyrroletetrazole-lysines (mPyTXKs) as site-selective photo-cross-linkers, with mPyTSeK cleavable upon H₂O₂ treatment.

The genetically encoded photo-cross-linkers have provided a powerful tool to study dynamic protein–protein interactions in cellular systems because they enable rapid formation of the covalent linkages among proteins with a burst of light.^{1–11} Among them, those containing a cleavable moiety^{12–16} are particularly useful because they combine photo-capture and subsequent release of the interacting partners in a single moiety, facilitating mass spectrometry-based characterization of the interaction interfaces. Recently, we reported a new class of genetically encoded protein photo-cross-linkers, 2-aryl-5-carboxytetrazole-lysines (ACTK) that cross-link the target proteins by reacting with a proximal nucleophilic

†Electronic Supplementary Information (ESI) available: Supplementary figures, experimental procedures, and NMR spectra. See DOI: 10.1039/x0xx00000x

Correspondence to: Qing Lin.

Conflicts of interest

There are no conflicts to declare.

residue.¹⁷ The most active analogue, *N*-methylpyrroletetrazole-lysine (**1** in Chart 1), exhibited robust and site-selective photo-cross-linking reactivity in *Escherichia coli* and mammalian cells. However, the expression yield was low due to modest activity of the mPyTKRS, an evolved ACTK-specific pyrrolysyl-tRNA synthetase that contains the Y306V/L309A/C348F/Y384F mutations.

To overcome this limitation, herein we report the design and synthesis of three new ACTK derivatives by replacing the γ -carbon of mPyTK with a heteroatom such as oxygen, sulphur, and selenium to increase hydrophilicity, which may also enhance substrate properties toward mPyTKRS. The resulting photo-cross-linkers are referred to thereafter as mPyTOK (**2**), mPyTSK (**3**), and mPyTSeK (**4**), or collectively as mPyTXK (X = O, S, Se). The calculated LogP values for these three analogues are -3.627 , -2.882 and -2.666 , respectively, lower than that of mPyTK (cLogP = -2.397) (Chart 1). All three mPyTXK analogues were genetically encoded into proteins in *E. coli* and mammalian cells. Two of the analogues, mPyTSK and mPyTSeK, enabled site-selective photo-cross-linking of the GST dimer at an efficiency comparable to mPyTK. Moreover, the cleavable photo-cross-linker, mPyTSeK, allowed rapid cleavage of the GST dimer upon treatment with a mild oxidizing agent, H₂O₂, which could be potentially useful in resolving the captured client proteins from the bait protein.^{15, 18–20}

We started by synthesizing mPyTXK using three discrete routes (Scheme 1). For the synthesis of mPyTOK, the aziridine ring-opening by a Boc-protected β -amino alcohol was employed as a key step to introduce γ -oxygen.²¹ In brief, the trityl-protected aziridine **6** was synthesized in three steps from L-serine benzyl ester. After switching the trityl protecting group to the Cbz group, the resulting aziridine **7** reacted with the Boc-protected β -aminoethanol in a Lewis acid catalysed nucleophilic ring opening reaction to afford the fully protected γ -oxo-lysine **8**. The Boc protecting group was then removed by treatment of HCl, and the resulting γ -oxo-lysine **9** was coupled with tetrazole carboxylic acid **10**, followed by deprotection to yield mPyTOK (**2**). For the synthesis of mPyTSK, the Boc-protected 2-bromoethanamine was conjugated with L-cysteine followed by Fmoc protection to give the protected γ -sulfur-lysine **12**. Removal of the Boc group, coupling with tetrazole **10**, followed by the Fmoc deprotection afforded mPyTSK (**3**). For the synthesis of mPyTSeK, 2-bromoethanamine was first coupled with tetrazole **10** to produce tetrazole bromide **14**, which was then used to alkylate L-selenocysteine followed by the Boc removal to yield mPyTSeK (**4**).²²

Because of the high structural similarity between mPyTXK and mPyTK, the previously reported mPyTKRS together with its cognate tRNA^{Pyl} CUA were employed to charge mPyTXK into superfolder green fluorescent protein (sfGFP) carrying an amber mutation at position-204 in *Escherichia coli*.^{17a} To our delight, all three mPyTXK analogues were incorporated into sfGFP with expression yield following the order of mPyTSeK > mPyTSK > mPyTOK ~ mPyTK (Fig. 1a). The fidelity of mPyTKRS in charging mPyTXK was assessed by ESI-MS and found to be high for mPyTOK and eroded to a small extent for mPyTSK and mPyTSeK. (Fig. 1b–d).

Next, we examined the efficiency of mPyTXK in photo-cross-linking *Schistosoma japonicum* glutathione-*S*-transferase (*Sj*GST) homodimer, a system used widely for evaluating the genetically encoded photo-cross-linkers.^{1,2,5} The GST proteins carrying mPyTXK at E52 position were expressed in BL21(DE3) cells and purified by Ni-NTA affinity chromatography. All three mPyTXK analogues were incorporated into the GST protein as evidenced by SDS-PAGE analysis (Fig. 2b) as well as LC-MS data (Fig. S1). The expression yield for GST-E52mPyTOK, GST-E52mPyTSK and GST-E52mPyTSeK was 1.2, 2.4, and 3.3 mg/L, respectively, in agreement with the sfGFP expression studies (Fig. 1a). We then photoirradiated the GST-E52mPyTXK proteins on ice with a handheld 302 nm UV lamp for 15 min and analysed the product mixtures by SDS-PAGE. As expected, we detected the GST dimer formation for all three mutants (Fig. 2b). Compared with mPyTK, mPyTSK and mPyTSeK gave comparable cross-linking yields whereas mPyTOK gave significantly lower yield, presumably due to greater solvation of the mPyTOK side chain which leads to more efficient water-quenching of the in situ generated carboxynitrile imine.

To probe whether the Se-containing mPyTSeK can mediate oxidative cleavage of the cross-linked proteins, GST-E52mPyTSeK was treated with 8 mM H₂O₂ at 30 °C for 1 h after initial photoirradiation (Fig. 2a). SDS-PAGE analysis indicated that the GST dimer disappeared completely (lane 10 in Fig. 2b), confirming that mPyTSeK is a cleavable protein photo-cross-linker. We then examined the incorporation of mPyTSeK into other positions (F53, L66 and R74) located at the interface of GST dimer. Unlike mPyTK, which can only be introduced into the E52 position,^{17a} mPyTSeK was incorporated at the other three positions (Fig. 2c), reaffirming that mPyTSeK is a superior substrate for mPyTKRS. However, the photo-cross-linking efficiency of the GST-F53mPyTSeK, GST-L66mPyTSeK or GST-R74mPyTSeK monomer to form the GST dimer was extremely low compared to GST-E52mPyTSeK (Figure 2c). We reasoned that the nucleophilic residues surrounding these positions are likely not in close proximity to the in situ generated carboxy-nitrile imine in order to form the covalent linkage. This result also confirms that mPyTSeK is a proximity-driven site-selective photo-cross-linker, similar to mPyTK.^{17a}

Since the mPyTKRS/tRNAPyl CUA pair are orthogonal to their mammalian counterparts,^{23–25} we examined the incorporation of mPyTXK into proteins in human embryonic kidney (HEK) cells. Briefly, HEK293T cells were transfected with two plasmids: a reporter construct bearing an *N*-terminal mCherry gene, a linker containing TAG amber codon, a C-terminal EGFP gene, and a HA-tag coding sequence (Fig. 3a), and a pCMV-mPyTKRS-tRNAPyl CUA expression cassette. The transfected cells were cultured in DMEM medium supplemented with 10% FBS in the absence and presence of mPyTK/mPyTXK for 48 h prior to imaging. As expected, the EGFP fluorescence was detected only when the unnatural amino acid was present in the medium (Fig. 3b). Compared to mPyTK, mPyTOK and mPyTSK showed similar levels of EGFP expression based on both the number of green fluorescent cells and the relative EGFP fluorescence intensity. Unexpectedly, mPyTSeK exhibited a modest cytotoxicity at concentrations 30 μM (data not shown), which is presumably due to oxidative stress caused by redox-active selenium-containing compounds.²⁶ To avoid this problem, a lower concentration of mPyTSeK (20 μM) was used and as a result lower incorporation efficiency was observed (Fig. 3b). The anti-HA western blot of the cell lysates confirmed the expression of the full-length fusion protein (Fig. 3c). Together,

these results confirm the poly-specificity of mPyTKRS in charging mPyTK analogues in both bacterial and mammalian cells.

To further demonstrate the utility of the cleavable protein photo-cross-linker mPyTSeK, we expressed the mPyTSeK-encoded Grb2, Grb2–143mPyTSeK (Fig. S2), an adaptor protein transducing signals from the activated EGFR to the Ras signaling pathway in mammalian cells.^{3,17a} After photo-cross-linking, the cross-linked Grb2–EGFR complex was detected by anti-His and anti-EGFR antibodies in the western blots, indicating that mPyTSeK-encoded Grb2 can covalently capture its interaction partner, EGFR, efficiently in the cell lysate (Fig. 4). Moreover, upon treatment with H₂O₂, the cross-linked Grb2–143mPyTSeK–EGFR complex disappeared, indicating the release of EGFR as a result of the cleavage of Se–C bond. In contrast, the cross-linked Grb2–143mPyTK–EGFR complex remained unchanged. This result confirms the function of the cleavable photo-cross-linker mPyTSeK and demonstrates a new way to use mPyTSeK for *in vitro* capturing of the protein–protein interaction complexes in mammalian cells lysates.

In summary, we have synthesized three γ -heteroatom-substituted *N*-methylpyrroletetrazole-lysine derivatives (mPyTXKs) as new protein photo-cross-linkers. All three mPyTXK analogues can be incorporated into proteins site-specifically in *E. coli* as well as in mammalian cells, with mPyTSK and mPySeK showing photo-cross-linking efficiency comparable to that of mPyTK. Furthermore, the mPyTSeK photo-cross-linked GST dimer can be cleaved under a mild oxidative condition. The cleavage site of mPyTSeK was located at the Se– γ -C₈ bond, which is different from other reported γ -seleno-substituted lysines. In addition, the mPyTSeK-encoded Grb2 protein can covalently capture and subsequent release of EGFR in mammalian cell lysates. The development of mPyTSK and mPyTSeK should add to a growing repertoire of the genetically encoded protein photo-cross-linkers and further enhance our capability in using chemical tools to map transient protein–protein interactions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge NIH (GM085092) for financial support.

Notes and references

1. Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG. J Am Chem Soc. 2002; 124:9026–9027. [PubMed: 12148987]
2. Chin JW, Martin AB, King DS, Wang L, Schultz PG. Proc Natl Acad Sci U S A. 2002; 99:11020–11024. [PubMed: 12154230]
3. Hino N, Okazaki Y, Kobayashi T, Hayashi A, Sakamoto K, Yokoyama S. Nat Methods. 2005; 2:201–206. [PubMed: 15782189]
4. Hino N, Oyama M, Sato A, Mukai T, Iraha F, Hayashi A, Kozuka-Hata H, Yamamoto T, Yokoyama S, Sakamoto K. J Mol Biol. 2011; 406:343–353. [PubMed: 21185312]
5. Chou C, Uprety R, Davis L, Chin JW, Deiters A. Chem Sci. 2011; 2:480–483.

6. Ai HW, Shen W, Sagi A, Chen PR, Schultz PG. *ChemBioChem*. 2011; 12:1854–1857. [PubMed: 21678540]
7. Yanagisawa T, Hino N, Iraha F, Mukai T, Sakamoto K, Yokoyama S. *Mol Biosyst*. 2012; 8:1131–1135. [PubMed: 22294092]
8. Zhang M, Lin S, Song X, Liu J, Fu Y, Ge X, Fu X, Chang Z, Chen PR. *Nat Chem Biol*. 2011; 7:671–677. [PubMed: 21892184]
9. Lin S, Zhang Z, Xu H, Li L, Chen S, Li J, Hao Z, Chen PR. *J Am Chem Soc*. 2011; 133:20581–20587. [PubMed: 22084898]
10. Pham ND, Parker RB, Kohler JJ. *Curr Opin Chem Biol*. 2013; 17:90–101. [PubMed: 23149092]
11. Preston GW, Wilson AJ. *Chem Soc Rev*. 2013; 42:3289–3301. [PubMed: 23396550]
12. Tamura T, Tsukiji S, Hamachi I. *J Am Chem Soc*. 2012; 134:2216–2226. [PubMed: 22220821]
13. Leriche G, Chisholm L, Wagner A. *Bioorg Med Chem*. 2012; 20:571–582. [PubMed: 21880494]
14. Yang Y, Hahne H, Kuster B, Verhelst SH. *Mol Cell Proteomics*. 2013; 12:237–244. [PubMed: 23028061]
15. Lin S, He D, Long T, Zhang S, Meng R, Chen PR. *J Am Chem Soc*. 2014; 136:11860–11863. [PubMed: 25084056]
16. Yang Y, Song H, He D, Zhang S, Dai S, Lin S, Meng R, Wang C, Chen PR. *Nat Commun*. 2016; 7:12299. [PubMed: 27460181]
17. (a) Tian Y, Jacinto MP, Zeng Y, Yu Z, Qu J, Liu WR, Lin Q. *J Am Chem Soc*. 2017; 139:6078–6081. [PubMed: 28422494] (b) Herner A, Marjanovic J, Lewandowski TM, Marin V, Patterson M, Miesbauer L, Ready D, Williams J, Vasudevan A, Lin Q. *J Am Chem Soc*. 2016; 138:14609–14615. [PubMed: 27740749]
18. Buchardt O, Elsner HI, Nielsen PE, Petersen LC, Suenson E. *Anal Biochem*. 1986; 158:87–92. [PubMed: 3026203]
19. Guo J, Wang J, Lee JS, Schultz PG. *Angew Chem Int Ed Engl*. 2008; 47:6399–6401. [PubMed: 18624319]
20. Wang ZU, Wang YS, Pai PJ, Russell WK, Russell DH, Liu WR. *Biochemistry*. 2012; 51:5232–5234. [PubMed: 22697363]
21. (a) Baldwin JE, Spivey AC, Schofield CJ, Sweeney JB. *Tetrahedron*. 1993; 49:6309–6330. (b) Larsson U, Carlson R. *Acta Chem Scand*. 1994; 48:511–516.
22. Siebum AHG, Woo WS, Raap J, Lugtenburg J. *Eur J Org Chem*. 2004; 2004:2905–2913.
23. Neumann H, Peak-Chew SY, Chin JW. *Nat Chem Biol*. 2008; 4:232–234. [PubMed: 18278036]
24. Chen PR, Groff D, Guo J, Ou W, Cellitti S, Geierstanger BH, Schultz PG. *Angew Chem Int Ed Engl*. 2009; 48:4052–4055. [PubMed: 19378306]
25. Schmied WH, Elsasser SJ, Uttamapinant C, Chin JW. *J Am Chem Soc*. 2014; 136:15577–15583. [PubMed: 25350841]
26. (a) Misra S, Boylan M, Selvam A, Spallholz JE, Bjornstedt M. *Nutrients*. 2015; 7:3536–3556. [PubMed: 25984742] (b) Barton WA, Tzvetkova-Robev D, Erdjument-Bromage H, Tempst P, Nikolov DB. *Protein Sci*. 2006; 15:2008–2013. [PubMed: 16823037]

Fig. 1.

Incorporation of mPyTXK into proteins in *E. coli*. (a) Coomassie blue stained SDS-PAGE gel of sfGFP-Q204 mutants encoding mPyTK, mPyTOK, mPyTSK, and mPyTSeK. Asterisk indicates an impurity derived from Ni-NTA affinity purification. (b) Deconvoluted mass spectrum of sfGFP-Q204mPyTOK: calcd, 27887.1 Da [M – Met + H⁺], found, 27891.2 ± 3.0 Da. (c) Deconvoluted mass spectrum of sfGFP-Q204mPyTSK: calcd, 27903.1 Da [M – Met + H⁺], found, 27904.7 ± 2.3 Da; calcd, 28034.1 Da [M + H⁺], found, 28037.2 ± 3.3 Da. (d) Deconvoluted mass spectrum of sfGFP-Q204mPyTSeK: calcd, 27951.1 Da [M – Met + H⁺], found, 27951.6 ± 1.6 Da; calcd, 28082.1 Da [M + H⁺], found, 28081.2 ± 3.2 Da. The smaller mass peaks around 27768 Da and 27710 Da correspond to sfGFP-Q204W and sfGFP-Q204Q, respectively, the products of near-cognate suppression by Trp and Gln, respectively.

Fig. 2.

(a) Scheme for photo-cross-linking of GST-mPyTSeK to form covalent GST dimer and the cleavage of GST dimer via H₂O₂-mediated oxidative reaction. The cross-linking sites are marked as red lines between the two monomers. (b) Coomassie blue stained SDS-PAGE gel after GST-E52mPyTK/mPyTOK/mPyTSK/mPyTSeK mutants were photoirradiated with a 302 nm UV lamp for 15 min. GST-E52mPyTSeK dimer was treated with 8 mM H₂O₂ at 30 °C for 1 h before SDS-PAGE analysis. (c) Coomassie blue stained SDS-PAGE gel after GST-E52mPyTSeK, GST-F53mPyTSeK, GST-L66mPyTSeK, GST-R74mPyTSeK mutants were photoirradiated with a 302 nm UV lamp for 15 min. Asterisk indicates an impurity derived from Ni-NTA affinity purification.

Fig. 3. (a) Structure diagram of the mCherry-TAG-EGFP reporter construct. (b) Bright field and fluorescence micrographs of HEK293T cells transfected with the plasmids expressing mCherry-TAG-EGFP and mPyTKRS-tRNAPyl CUA and cultured in DMEM medium supplemented with 10% FBS and 1 mM mPyTK/mPyTXK (X = O, S) or 20 μ M mPyTSeK. (c) Western blot analysis of HEK293T cell lysates probed with anti-HA tag antibody.

Fig. 4.

Photo-cross-linking of EGFR by the mPyTK/mPyTSeK-encoded Grb2 mutants. Purified wt-Grb2, Grb2-143mPyTK/mPyTSeK was mixed with the EGFR-expressing HEK293T cell lysates, photoirradiated with a 302 nm UV lamp for 15 min, and then treated with 8 mM H₂O₂ at 30 °C for 1 h before SDS-PAGE analysis and western blot probed with anti-His tag antibody (top panel) and anti-EGFR antibody (bottom panel).

Scheme 1.

Synthesis of (a) mPyTOK, (b) mPyTSK, and (c) mPyTSeK.

X = CH ₂	mPyTK (1)	cLoP = -2.397
O	mPyTOK (2)	cLoP = -3.627
S	mPyTSK (3)	cLoP = -2.882
Se	mPyTSeK (4)	cLoP = -2.666

Chart 1.
Structures of mPyTK and mPyTXX