Communication

Genetically encoded 2-aryl-5-carboxytetrazoles for site-selective protein photo-cross-linking

Yulin Tian, Marco Paolo Jacinto, Yu Zeng, Zhipeng Yu, Jun Qu, Wenshe R. Liu, and Qing Lin

J. Am. Chem. Soc., **Just Accepted Manuscript** • Publication Date (Web): 19 Apr 2017

Genetically encoded 2-aryl-5-carboxytetrazoles for site-selective protein photo-cross-linking

Yulin Tian,[†] Marco Paolo Jacinto,[†] Yu Zeng,[‡] Zhipeng Yu,[†,#] Jun Qu,[⊥] Wenshe R. Liu,[‡] Qing Lin[†,*]

[†]Department of Chemistry, State University of New York at Buffalo, Buffalo, New York 14260, United States

[‡]Department of Chemistry, Texas A&M University, College Station, Texas 77845, United States

[⊥]Department of Pharmaceutical Sciences, State University of New York at Buffalo, Buffalo, NY 14260, United States

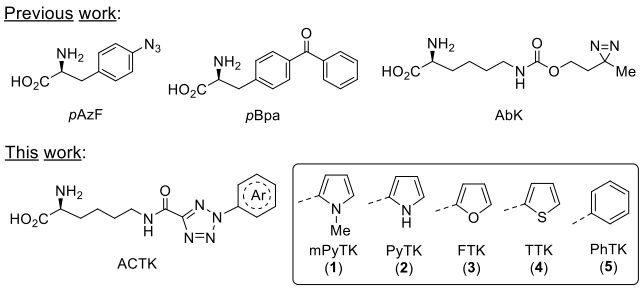
Supporting Information Placeholder

ABSTRACT: The genetically encoded photo-cross-linkers promise to offer a temporally controlled tool to map transient and dynamic protein-protein interaction complexes in living cells. Here we report the synthesis of a panel of 2-aryl-5-carboxytetrazole-lysine analogs (ACTKs) and their site-specific incorporation into proteins via amber codon suppression in *E. coli* and mammalian cells. Among five ACTKs investigated, *N*-methylpyrroletetrazole-lysine (mPyTK) was found to give robust and site-selective photo-cross-linking reactivity in *E. coli* when placed at an appropriate site at the protein interaction interface. A comparison study indicated that mPyTK exhibits higher photo-cross-linking efficiency than a diazirine-based photo-cross-linker, AbK, when placed at the same location of the interaction interface in vitro. When mPyTK was introduced into the adapter protein Grb2, it enabled the photo-capture of EGFR in a stimulus-dependent manner. The design of mPyTK along with the identification of its cognate aminoacyl-tRNA synthetase makes it possible to map transient protein-protein interactions and their interfaces in living cells.

To map dynamic protein-protein interactions in living cells, a powerful chemical strategy involves the use of genetically encoded photo-cross-linkers that permanently link transient protein-protein interaction complexes with a burst of light. Based on the structures, the reported genetically encoded photo-cross-linkers contain one of the three moieties: phenyl azide such as *p*AzF,[1] benzophenone such as *p*Bpa,[2] and diazirine such as AbK[3,4] and their derivatives (Chart 1).[5-8] While these genetically encoded photo-cross-linkers have proven to be valuable in the study of protein structure and function, they invariably cross-link with their interacting protein partners with no selectivity for any particular residue as the photo-generated reactive intermediate, i.e., the nitrene from phenyl azide, the diradical from benzophenone and the carbene from diazirine, inserts into a proximal C–H bond with appropriate distance and angle,[9] making it difficult to predict *a prior* suitable positions for installation of the photo-cross-linker. In addition, the tandem mass spectrometry-based mapping of the interaction interface is complicated as any residue from the interacting protein partner can potentially participate in the photo-cross-linking.

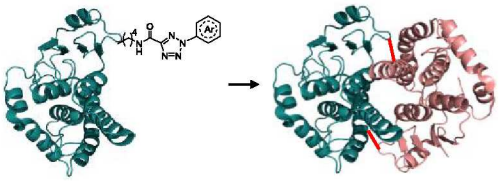
Recently, we reported a new photo-affinity label based on 2-aryl-5-carboxytetrazole (ACT) with a size similar to benzophenone, which cross-links its target proteins via addition with a proximal nucleophile near the active site.[10] Since ACT exhibits ligand-dependent selective photo-cross-linking, we envisioned that ACT may also serve as a new class of genetically encoded photo-cross-linkers for mapping transient protein-protein interaction interfaces.

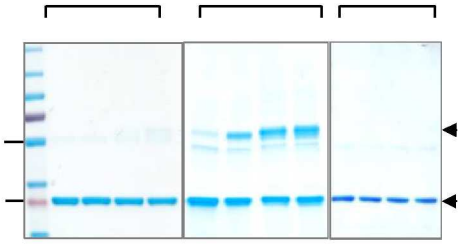
Chart 1. Genetically encoded photo-cross-linkers.

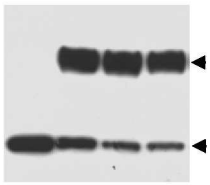


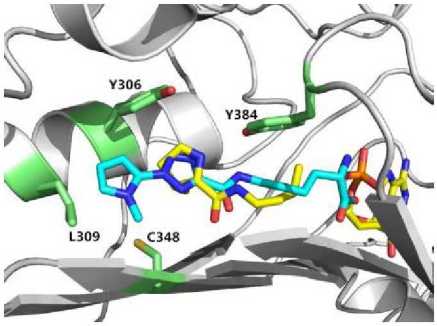
It is noted that a biocompatible proximity-driven nucleophilic substitution reaction between a genetically encoded N^{ε}-fluoroacetyllysine and cysteine was reported recently for mapping the protein-protein interaction interface.[11] However, the occurrence of native cysteine at the protein-protein interaction interface is rather rare.[12] Herein, we report the synthesis of a panel of ACT-lysine analogs (ACTK, Chart 1), and the identification of a new ACTK-specific pyrrolysyl-tRNA synthetase for site-specific incorporation of ACTK into proteins in *E. coli* and in mammalian cells. One of the ACTK analogs, mPyTK, exhibited robust and site-selective photo-cross-linking of a GST dimer in bacteria. In a comparison study, mPyTK showed significantly higher cross-linking efficiency than AbK when both are incorporated at the same location of GST. Moreover, the mPyTK-encoded adapter protein, Grb2, showed a stimulus and position-dependent capture of its transient interaction partner, epidermal growth factor receptor (EGFR), in mammalian cells.

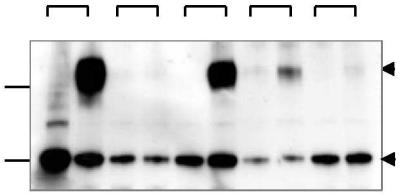
Since pyrrolysyl-tRNA synthetase (PylRS) and its variants have shown tremendous versatility in charging various lysine derivatives into proteins site-selectively in bacteria, yeast and mammalian cells,[13] we decided to append ACT motif onto the ***ε***-amino group via simple acylation reaction. For the synthesis of ACTK analogs **1**-**4** in Chart 1, the key intermediate, ethyl 2-aryl-2*H*-tetrazole-5-carboxylate, was obtained through Cu^{II}-catalyzed cross-coupling of ethyl 2*H*-tetrazole-5-carboxylate with the phenylaryliodonium salt[14] (Schemes S1-S4 in Supporting Information). For PhTK (**5**), the Kakehi tetrazole synthesis was followed to give the ethyl 2-phenyl-2*H*-tetrazole-5-carboxylate intermediate[15] (Scheme S5 in Supporting Information). Subsequent hydrolysis and coupling with Fmoc-lysine•HCl followed by removal of the protecting group afforded the ACT-lysine analogs **1**-**5** (Chart 1). To identify PylRS mutants that efficiently charge mPyTK (**1**), an *Mm*PylRS library, in which four residues surrounding the *N*-methylpyrrole-lysine side chain based on the crystal structure of *M. mazei* PylRS in complex with Pyl-AMP[16] (Y306, L309, C348 and Y384; Figure 1a) were

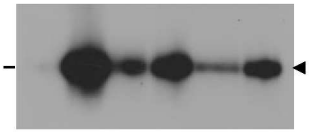


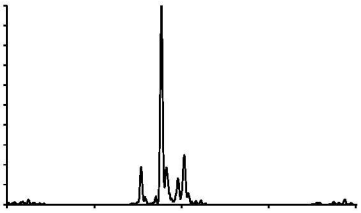





Y306
Y384
L309
C348







(Figure 2b), indicating that the ACT moiety is responsible for dimer cross-linking. For comparison, we expressed the GST-E52AbK mutant using the wild-type PylRS with a yield of 2.5 mg L^{-1} and examined its photo-cross-linking reactivity. To our surprise, AbK exhibited very weak reactivity as the dimer band was detected only by western blot[4] (Figure S5 in SI) but not Coomassie blue (Figure 2b). Interestingly, a higher cross-linking yield (~79%) was obtained when *E. coli* cells expressing GST-E52mPyTK were directly photoirradiated, which could be attributed to higher intracellular concentration of the GST mutant (Figure 2c and Figure S6 in SI). Since mPyTKRS can charge other ACTKs into proteins site-selectively, we expressed *Sj*GST mutants carrying PyTK, FTK, TTK and PhTK, respectively, at position-52 and compared their photo-cross-linking efficiency in *E. coli* cells. Based on western blot analysis, only mPyTK, FTK and TTK showed the cross-linked dimer with the efficiency order of mPyTK > TTK > FTK (Figure 2d), presumably due to the highest electron density at *N*-methyl-pyrrole ring, which helps to stabilize the photo-generated carboxy-nitrile imine and increase its life-time in biological media.

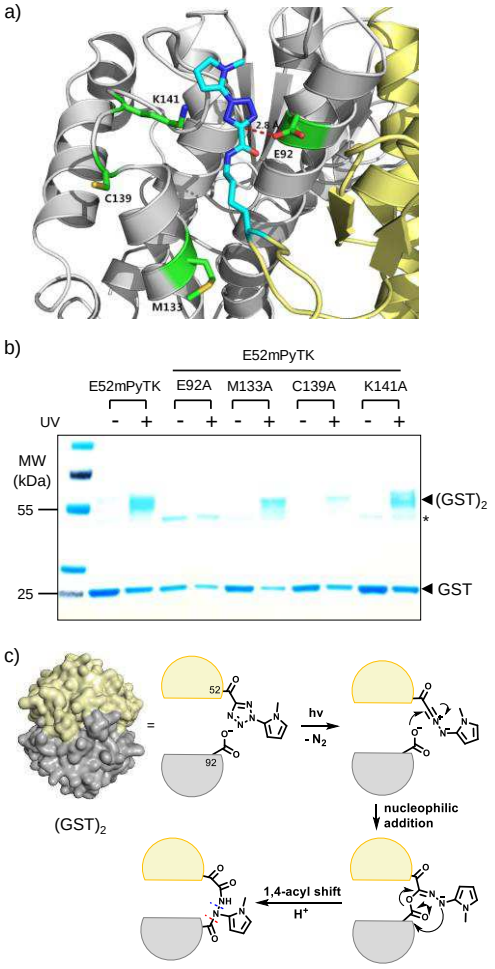


Figure 3. Identifying the mPyTK photo-cross-linking site in GST. (a) A close-up view of the nucleophilic residues from the opposite GST monomer (colored in gray) surrounding mPyTK in GST monomer (colored in yellow). The side chains of proximal resides (E92, M133, C139 and K141) are rendered in tube model. (b) Coomassie blue stained SDS-PAGE gel showing UV-dependent cross-linking of the GST-mPyTK alanine mutants. Asterisk indicates an impurity derived from Ni-NTA affinity purification. The proteins were photoirradiated with a handheld 302-nm UV lamp on ice for 15 min before SDS-PAGE. (c) Proposed mechanism for mPyTK-mediated photo-cross-linking of GST dimer. The two cleavage pathways are marked with blue and red dash lines on the cross-linked structure (see Figure S8 in SI for details).

Since ACT photo-reacts with proximal nucleophilic residues on proteins, we sought to determine which nucleophilic residues on the opposite GST monomer might react with the photo-generated carboxy-nitrile imine intermediate. To this end, we built a model of the GST-E52mPyTK and surveyed the chemical environment surrounding mPyTK. Four nucleophilic residues (E92, M133, C139 and K141) were identified that are located 2.8–13.0 Å from the electrophilic nitrile imine carbon (Figure 3a). To determine which one of these four residues participates the cross-linking reaction, we mutated these residues to alanine and examined the photo-cross-linking activity of the resulting mutants. We found the Glu92→Ala mutation completely abolished the covalent dimer formation while other mutations had no effect (Figure 3b). This result is consistent with the proximity-driven reactivity as E92 is closest to mPyTK with a calculated distance between the carboxylate oxygen and the nitrile imine carbon of 2.8 Å (Figure 3a). Similar results were obtained when the alanine scan was conducted with the GST-E52-FTK mutant (Figure S7 in SI). We propose a photo-cross-linking mechanism in which the E92 carboxylate undergoes nucleophilic addition to the photo-generated carboxy-nitrile imine followed by 1,4-acyl shift (Figure 3c). The rearranged cross-linked structure was supported by tandem mass spectrometry data in which the two fragment ions derived from the two discrete cleavage pathways were positively identified (Figure S8 in SI). For comparison, we performed photo-cross-linking studies with the same set of alanine mutants of the GST-E52AbK (E92A, M133A, C139A and K141A). We did not observe prominent attenuation in GST dimer formation; unexpectedly, two alanine mutants (M133A and K141A) showed greater dimer formation than the GST-E52AbK alone (Figure S7 in SI), presumably due to a remodeling of the interaction interface that alters the distance and/or angle of a suitable proximal C–H bond.[9]

To examine whether the genetically encoded mPyTK can capture protein–protein interaction complexes in mammalian cells, we first confirmed that when mPyTKRS was expressed in HEK293T cells, it allows site-selective incorporation of mPyTK into mCherry-TAG-EGFP based on confocal fluorescence microscopic analysis (Figure S9 in SI). We then introduced mPyTK into Grb2, an adaptor protein that links phosphorylated EGFR to the Ras signaling pathway through guanine nucleotide exchange factor Sos[18] and was used previously for evaluating the genetically encoded photo-cross-linkers in mammalian cells.[19] Inspection of crystal structure of the Grb2 SH2 domain in complex with a phosphotyrosine-containing heptapeptide ligand[20] revealed that 9 residues surrounding the ligand, A91, D104, V105, Q106, F108, K109, L111, W121 and N143, could be mutated to mPyTK for potential photo-cross-linking with the SH2 ligands such as the cytoplasmic domain of EGFR (Figure 4a). Thus, HEK293T cells were co-transfected with pCMV-mPyTKRS-tRNA$^{Pyl}_{CUA}$ encoding mPyTK-specific PylRS and tRNA$^{Pyl}_{CUA}$, pCMV6-Grb2-myc-DDK encoding either wild-type or amber mutant with TAG codon substituted at any of the 9 positions with a C-terminal myc-DDK tag, and pcDNA3-EGFR-EGFP encoding full-length EGFR and a C-terminal EGFP tag, and protein expressions were carried out in DMEM medium supplemented with 10% FBS and 1 mM mPyTK. The cells were starved for 12 h before EGF stimulation and subsequent photoirradiation on ice. The cells were then lysed and the lysates were treated with protein tyrosine phosphatase 1B to hydrolyze the phosphotyrosine to obviate non-covalent Grb2-interacting protein complexes. The cross-linked Grb2-interacting proteins were immunoprecipitated with anti-Flag antibody and analyzed by sequential western blots using anti-myc and anti-EGFR antibodies. The cross-linked Grb2—EGFR complex was detected for 8 out 9 Grb2-mPyTK mutants; the D104mPyTK mutant gave the highest photo-cross-linking yield followed by the V105mPyTK, Q106mPyTK and N143mPyTK mutants (Figure 4b and Figure S10 in SI). The clustering of the four mutants (D104, V105, Q106 and N143) in the same region of Grb2 SH2 domain suggests that the

mPyTK photo-cross-linker in these mutants may react with the same nucleophilic residue on EGFR across the interaction interface. Moreover, the photo-cross-linking of EGFR is EGF stimulation and photoirradiation-dependent and is mediated by mPyTK as the wild-type Grb2 did not exhibit covalent capture of EGFR (Figure S11A in SI). The highest photo-cross-linking yield for Grb2-D104mPyTK was observed when cells were stimulated with EGF for 15 min (Figure S11B in SI), indicating that the Grb2—EGFR interaction is transient and dynamic, resembling some other known EGF-dependent protein-protein interactions.[21]

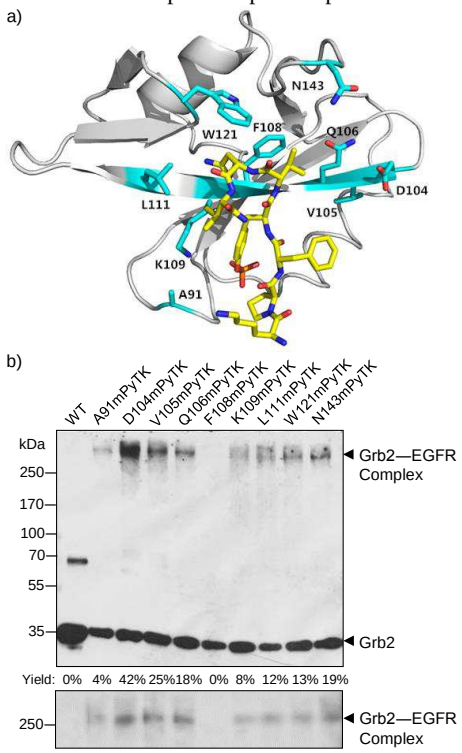


Figure 4. Photo-cross-linking of EGFR by the mPyTK-encoded Grb2 mutants in mammalian cells. (a) A close-up view of Grb2-SH2 domain in complex with phosphotyrosine-containing heptapeptide (PDB code: 1TZE), highlighting the proximal residues surrounding pY (rendered in cyan tube model) that were selected for the mPyTK mutagenesis. (b) Comparison of the efficiency of the mPyTK-containing Grb2-SH2 domain mutants for photo-cross-linking with EGFR in HEK293T cells. Cells were exposed to 302-nm UV light for 5 min before lysis. The Grb2-cross-linked proteins in cell lysates were immunoprecipitated with the anti-FLAG antibody-immobilized agarose beads. The samples were analyzed SDS-PAGE/western blot and probed successively with anti-myc antibody (top panel) and anti-EGFR antibody (bottom panel).

In summary, we have synthesized a panel of 2-aryl-5-carboxy-tetrazole-based photo-cross-linkers and evolved a polyspecific pyrrolysyl-tRNA synthetase variant that charges these ACT-lysine analogs site-selectively into proteins in *E. coli*. One of the most reactive, genetically encoded photo-cross-linkers, mPyTK, allowed site-selective photo-cross-linking of a GST dimer in vitro and in *E. coli* cells. In comparison studies, mPyTK exhibited a significantly higher efficiency than AbK when placed at the same location of the GST dimer interface. Moreover, mPyTK enabled covalent capture of the transient Grb2-interacting protein partner in mammalian cells in a stimulus-dependent manner. In view of their higher photo-cross-linking yields and unique cross-linking mechanism, these genetically encoded ACT-lysines should offer a powerful chemical tool to map transient protein-protein interactions underlining signal transduction pathways. Because tandem MS analysis revealed the characteristic fragments after cross-linking, these ACT-lysines may also find applications in structural analysis of protein complexes.

ASSOCIATED CONTENT

Supporting Information

Supplemental figures and table, synthetic schemes, experimental procedures, characterization of new compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

Qing Lin (qinglin@buffalo.edu).

Present Address

[#]Sichuan University College of Chemistry, China.

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

We gratefully acknowledge the NIH (GM085092 to Q.L., CA 161158 to W.R.L.) for financial support. We thank Prof. John Koland at University of Iowa for providing the pcDNA3-EGFR-EGFP plasmid, and Carlo Ramil in Q.L. lab for cloning assistance.

REFERENCES

(1) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *J Am Chem Soc* **2002**, *124*, 9026-9027.

(2) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proc Natl Acad Sci U S A* **2002**, *99*, 11020-11024.

(3) Ai, H. W.; Shen, W.; Sagi, A.; Chen, P. R.; Schultz, P. G. *Chembiochem* **2011**, *12*, 1854-1857.

(4) Chou, C. J.; Uprety, R.; Davis, L.; Chin, J. W.; Deiters, A. *Chem Sci* **2011**, *2*, 480-483.

(5) Lin, S.; He, D.; Long, T.; Zhang, S.; Meng, R.; Chen, P. R. *J Am Chem Soc* **2014**, *136*, 11860-11863.

(6) Yang, Y.; Song, H.; He, D.; Zhang, S.; Dai, S.; Lin, S.; Meng, R.; Wang, C.; Chen, P. R. *Nat Commun* **2016**, *7*, 12299.

(7) Yanagisawa, T.; Hino, N.; Iraha, F.; Mukai, T.; Sakamoto, K.; Yokoyama, S. *Mol Biosyst* **2012**, *8*, 1131-1135.

(8) 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.

(9) Sato, S.; Mimasu, S.; Sato, A.; Hino, N.; Sakamoto, K.; Umehara, T.; Yokoyama, S. *Biochemistry* **2011**, *50*, 250-257.

(10) Herner, A.; Marjanovic, J.; Lewandowski, T. M.; Marin, V.; Patterson, M.; Miesbauer, L.; Ready, D.; Williams, J.; Vasudevan, A.; Lin, Q. *J Am Chem Soc* **2016**, *138*, 14609-14615.

(11) Kobayashi, T.; Hoppmann, C.; Yang, B.; Wang, L. *J Am Chem Soc* **2016**, *138*, 14832-14835.

(12) Bogan, A. A.; Thorn, K. S. *J Mol Biol* **1998**, *280*, 1-9.

(13) Wan, W.; Tharp, J. M.; Liu, W. R. *Biochim Biophys Acta* **2014**, *1844*, 1059-1070.

(14) An, P.; Yu, Z.; Lin, Q. *Chem Commun (Camb)* **2013**, *49*, 9920-9922.

(15) Wang, Y.; Lin, Q. *Org Lett* **2009**, *11*, 3570-3573.

(16) Kavran, J. M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Soll, D.; Steitz, T. A. *Proc Natl Acad Sci U S A* **2007**, *104*, 11268-11273.

(17) Wang, Z. A.; Zeng, Y.; Kurra, Y.; Wang, X.; Tharp, J. M.; Vatansever, E. C.; Hsu, W. W.; Dai, S.; Fang, X.; Liu, W. R. *Angew Chem Int Ed Engl* **2016**, *55*, 212-216.

(18) Li, N.; Batzer, A.; Daly, R.; Yajnik, V.; Skolnik, E.; Chardin, P.; Bar-Sagi, D.; Margolis, B.; Schlessinger, J. *Nature* **1993**, *363*, 85-88.

(19) Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokoyama, S. *Nat Methods* **2005**, *2*, 201-206.

(20) Rahuel, J.; Gay, B.; Erdmann, D.; Strauss, A.; Garcia-Echeverria, C.; Furet, P.; Caravatti, G.; Fretz, H.; Schoepfer, J.; Grutter, M. G. *Nat Struct Biol* **1996**, *3*, 586-589.

(21) Zheng, Y.; Zhang, C.; Croucher, D. R.; Soliman, M. A.; St-Denis, N.; Pasculescu, A.; Taylor, L.; Tate, S. A.; Hardy, W. R.; Colwill, K.; Dai, A. Y.; Bagshaw, R.; Dennis, J. W.; Gingras, A. C.; Daly, R. J.; Pawson, T. *Nature* **2013**, *499*, 166-171.