Published in final edited form as:

Angew Chem Int Ed Engl. 2017 April 24; 56(18): 5096–5100. doi:10.1002/anie.201611841.

Protein Crosslinking by Genetically Encoded Noncanonical Amino Acids with Reactive Aryl Carbamate Side Chains

Weimin Xuan, Sida Shao, and Peter G. Schultz

Department of Chemistry, the Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037 (USA)

Abstract

The use of genetically encoded noncanonical amino acids (ncAAs) to construct crosslinks within or between proteins has emerged as a useful method to enhance protein stability, investigate protein-protein interactions and improve the pharmacological properties of proteins. Here we report ncAAs with aryl carbamate side chains (PheK and FPheK) that can react with proximal nucleophilic residues to form intra- or intermolecular protein crosslinks. We evolved a pyrrolysyl-tRNA synthetase that can site-specifically incorporate PheK and FPheK into proteins in both *E. coli* and mammalian cells. PheK and FPheK when incorporated into proteins showed good stability during protein expression and purification; crosslinking required incubation under weakly basic condition. We demonstrated that FPheK can react with adjacent Lys, Cys and Tyr residues in thioredoxin in high yields. In addition, crosslinks could be formed between FPheK and Lys residue of two interacting proteins, including the heavy chain and light chain of an antibody Fab. The high efficiency and specificity of FPheK crosslinking may make it a practical tool to engineer proteins with unnatural covalent linkages.

Noncanonical Amino Acid (ncAA)

Reactive carbamate-containing ncAA (FPheK) was genetically encoded into recombinant proteins. It displayed high reactivity to proximal nucleophilic residues at weakly basic condition, including

Lys, Cys and Tyr as efficient targets. Highly efficient and site-specific intra- and intermolecular protein crosslinks were demonstrated with FPheK.

Keywords

noncanonical amino acid; crosslink; protein engineering; carbamate; proximity

Intra- or intermolecular crosslinks within peptides and proteins play an important role in stabilizing polypeptide tertiary or quaternary structures, including high energy backbone conformations. Examples include disulfide-bonded insulin and antibodies, and highly post-translationally processed thiopeptides containing lanthionine linkages.^[1] However, the natural chemistries that can be used to constrain polypeptides and the chemical structures are limited, and as a consequence, “synthetic” crosslinking strategies have been developed. The latter includes bifunctional chemical reagents,^[2] olefin metathesis^[3] and intein-mediated cyclization.^[4]

Our lab and others have developed recombinant methods to site-specifically engineer proteins with new chemistries by reprogramming the protein translational machinery with an orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pair, a unique nonsense or frameshift codon, and a noncanonical amino acid (ncAA).^[5] A number of ncAAs with electrophilic side chains that can react with adjacent nucleophilic residues (usually Cys or Lys) in a target protein to form intra- or intermolecular crosslinks, have been developed. These reactive groups include reactive halides,^[6] Michael acceptors^[7] and an aryl isothiocyanate.^[8] In addition, ncAAs with long side-chain thiols that form extended disulfide bonds in proteins have been reported.^[9] These crosslinks are characterized by good efficiency and selectivity, and mild crosslinking conditions. To further expand the number and nature of co-translational protein crosslinking chemistries, we now report a genetically encode carbamate containing ncAA that can readily form intra- and intermolecular crosslinks in proteins under mild conditions.

Aryl carbamates are moderate electrophiles, and have been used as cleavable linkers in antibody drug conjugates that undergo proximity assisted nucleophilic attack to release free drug.^[10] To incorporate a carbamate side chain into a ncAA, we appended either a O-phenyl or O-4-fluorophenyl side chains to Lys to afford PheK and FPheK, as described in Scheme 1. Phenol or 4-fluorophenol will be released to directly signal crosslink formation. We synthesized these two ncAAs in a facile two-step process in ~30% overall yield (see SI). To genetically encode PheK and FPheK, the Mb pyrrolysyl-tRNA synthetase (PylRS)/tRNA_{CUA} pair that has been used to incorporate a large number of Lys derivatives, was employed.^[11] A Mb-PylRS library with residues Leu270, Tyr271, Leu274, and Cys313 randomized to NNK (N = A, T, G or C, K = T or G) and residue Try349 fixed to Phe was used to evolve a synthetase that can specifically recognize PheK.^[12] After alternating positive and negative rounds of selection in *E. coli* DH10B, two unique synthetase mutants were obtained (PheKRS1 and PheKRS2, Fig S1b). To determine relative suppression efficiencies, we expressed a sfGFP variant in *E. coli* DH10B harboring pLeiG-GFP-Asp134TAG (sfGFP with a Asp134TAG mutation at a nonessential site) and pBK-PheKRS1

or pBK-PheKRS2 in LB media supplemented with 2 mM PheK. PheKRS1 (Leu270Leu, Tyr271Tyr, Leu274Ala, Cys313Phe, Tyr349Phe) showed much stronger fluorescence enhancement (Fig S1c). It's noteworthy that PheKRS1 has the same sequence as the synthetase previously evolved to incorporate ϵ -*N*-crotonyllysine, indicating its polyspecific nature.^[12b] We also tested the ability of PheKRS1 to incorporate FPheK, and found it to be incorporated into sfGFP(Tyr151TAG) in excellent yield (Fig 1A). The suppression difference may be partially due to the modest toxicity of PheK to *E. coli* growth, whereas FPheK had little effect on *E. coli* growth (Fig S2). To confirm the selective incorporation of PheK and FPheK, C-terminal His-tagged mutants of sfGFP (Y151TAG) were expressed in *E. coli* DH10B in the presence of 2 mM PheK or FPheK, and purified by Ni-NTA column in 3 and 8 mg/L isolated yield (wild type: 65 mg/L). The mutant proteins were analyzed by SDS-PAGE gel (Fig 1B) and electrospray ionization quadrupole time-of-flight mass spectrometry (ESI-QTOF Mass); the observed mass was in good agreement with the expected mass for both sfGFP mutants (Fig 1C), and no free Cys or Lys adduct was observed, suggesting good stability of both ncAAs to physiological conditions.

Previously, it was shown that the PylRS/tRNA_{CUA} pair is also orthogonal mammalian cells.^[7b, 13] Therefore we attempted to incorporate PheK and FPheK into an EGFP mutant (Tyr39TAG) in HEK293T cells. As shown in Fig 1D, bright green fluorescence was only observable in transfected cells in the presence of 1 mM PheK or FPheK, indicating the production of full-length EGFP (Tyr39TAG) through ncAA insertion. Furthermore, ncAA incorporation and stability was verified by ESI-QTOF analysis of the purified C-terminal His-tagged mutant of EGFP expressed in HEK293T cells (Fig S3). Due to its higher suppression efficiency and lower cytotoxicity, FPheK was used in all the following experiments.

Next we asked whether FPheK when incorporated into proteins can react with proximal Lys residues to form a urea crosslink. The distance between two surface residues Gln98 and Lys90 in *E. coli* thioredoxin is 7.1 Å (Fig S4), suggesting that substitution of Gln98 with FPheK might result in a urea linkage between these two sites. To test this hypothesis, Gln98FPheK mutant was generated in a C-terminal His-tagged thioredoxin. The mutant was successfully expressed and purified by Ni-NTA chromatography in reasonable yield (3 mg/L); incorporation of FPheK was confirmed by ESI-QTOF mass (Exp. 12634, Obs. 12634). The purified thioredoxin mutant showed an intact FPheK even in the presence of an adjacent Lys in PBS buffer, pH 7.4. This is likely due to the high pKa (10.5) of the Lys ϵ -amino group. Therefore, to trigger crosslinking, we incubated this thioredoxin mutant in pH 8.5 HEPES buffer at 37 °C for 8 hours, and then analyzed the resulting product by ESI-QTOF mass. The original mass peak at 12634 Da completely disappeared and a new peak at 12521 Da appeared, consistent with the loss of 4-fluorophenol and formation of a urea crosslink (Fig 2A). The thioredoxin mutants were also analyzed by SDS-PAGE gel, showing a slightly increased mobility for the crosslinked product. (Fig 2B). As negative control, we expressed a thioredoxin mutant with a stable carbamate-containing ncAA, BocK (*N*_ε-Boc-L-lysine) at Gln98 followed by incubation at pH 8.5 for 24 hours, no crosslinked product was observed; Furthermore, when Lys90 was mutated to Asn, no crosslinked product was observed (Fig S5). These results strongly confirmed that the crosslink was formed between FPheK and Lys90.

To test whether FPheK can react with other nucleophilic residues to form a stable crosslink, we independently mutated Lys90 in thioredoxin to Cys, Tyr, His and Ser in the context of the Gln98PheK mutant. All mutants were expressed in *E. coli* DH10B, purified by Ni-NTA chromatography and analyzed by SDS-PAGE (Fig S6) and ESI-QTOF mass. All purified mutants showed mass spectra consistent with an intact FPheK. When we incubated these four mutants in pH 8.5 HEPES buffer at 37 °C for 12 hours, ESI-QTOF analysis showed that both Cys and Tyr can form thiocarbamate or carbamate crosslinks with FPheK with excellent efficiency (Fig 3A and B). A small peak (12530 Da) observed for Tyr may result from competitive hydrolysis of FPheK that generates carbamic acid with loss of CO₂. Crosslinks were also observed with His and Ser but with low reaction efficiency and competitive hydrolysis.

Next, we explored whether FPheK can form an intermolecular crosslink between two interacting proteins. Z protein binds Z_{spa} affibody (Afb) with high affinity.^[14] In addition, we previously demonstrated *p*NCSF (*p*-isothiocyanate phenylalanine) at Glu25 in Z protein can form a stable thiourea linkage with Lys8 in an affibody fusion protein (Afb-Trx) (Fig 4A).^[8] Therefore, we incorporated FPheK at Glu25 in Z protein; the Z protein mutant (Glu25FPheK) was expressed in *E. coli* DH10B and purified by Ni-NTA chromatography in a yield of 4 mg/L. SDS-PAGE gel and ESI-QTOF mass was used to confirm FPheK incorporation (Fig S7); the observed mass (7744.4 Da) was consistent with the expected mass (7744.5 Da). Afb-Trx (6 μM) was co-incubated with 2 equiv of Z protein mutant at 37 °C in pH 8.5 HEPES buffer, and samples were taken at different time points. On the basis of SDS-PAGE analysis (Fig 4B), a new band was observable at 2 h with expected size of 27 kDa, consistent with the formation of Z protein/Afb-Trx complex. The formation of the Z protein/Afb-Trx complex was confirmed by ESI-QTOF analysis (Exp. 26694, Obs. 26693, Fig S8). As a negative control, incubation of the Z protein Glu25PheK mutant and Afb-Trx Lys8Asn mutant at pH 8.5 did not afford crosslinked product (Fig S9). Finally, crosslinking between FPheK and Lys8 of Afb-Trx was further confirmed by Glu-C protease digestion and LC-MS/MS analysis (Fig 4C, Fig. S10 and Table S2).

Finally, we attempted to crosslink the heavy and light chains of an antibody Fab. The antigen-binding fragment of the therapeutic monoclonal antibody Herceptin has a size of over 47 kDa and contains five disulfide bonds, including the interchain disulfide bond between Cys223 of the heavy chain and Cys214 of the light chain. To form a urea crosslink between the two chains, FPheK was substituted for Gln100 on the light chain which is in close proximity to Lys43 on the heavy chain (Fig 5A). To avoid competitive crosslinking within light chain, Lys103 on the light chain was mutated to Gln. The Herceptin Fab mutant was expressed in *E. coli* DH10B and purified by periplasmic lysis and protein G antibody affinity chromatography in 5 mg/L isolated yield. ESI-QTOF mass (Fig 5C) confirmed FPheK incorporation. This mutant was then incubated in pH 8.5 HEPES buffer at 37 °C for 24 hours, and analyzed by SDS-PAGE and ESI-QTOF (Fig 5B and C). Wild type Herceptin Fab was completely cleaved by DTT on SDS-PAGE gel, while the crosslinked Fab mutant displayed a reduced mobility in the presence of DTT on a SDS-PAGE gel, consistent with a urea crosslink between the two chains. ESI-QTOF mass further confirmed crosslink formation (loss of 112 Da as expected). Both SDS-PAGE gel and ESI-QTOF indicate that crosslink formation is nearly quantitative. As a negative control, Lys43 on the heavy chain

was mutated to Gln, and this control mutant was expressed and purified in 4 mg/L yield. After incubation under basic condition, SDS-PAGE gel showed a complete cleavage of this control mutant by DTT, and ESI-QTOF mass further confirmed that most of the control mutant remained intact with only a small amount of byproduct likely due to hydrolysis and nonspecific crosslinking within the light chain (Fig S11). These results demonstrate that a site-specific and nearly quantitative crosslink has been successfully constructed between FPheK at Gln100 in the heavy chain and Lys43 in the light chain in Herceptin Fab. Reaction times for crosslinking experiments with FPheK ranged from 4 hours to 24 hours, suggesting that besides pH, protein microenvironment can also affect the kinetics.

In summary, we genetically incorporated FPheK with a reactive aryl carbamate side chain into recombinant proteins in both *E. coli* and mammalian cells. FPheK when incorporated into proteins showed good stability and necessitated only weakly basic conditions to react with adjacent nucleophilic residues with the release of 4-fluorophenol. We demonstrated highly efficient crosslink formation between FPheK and the proximal Lys, Cys or Tyr residues within thioredoxin. Furthermore, we showed that site-specific and highly efficient urea crosslinks can be formed between two interacting proteins. This work provides a useful nAA to engineer proteins with highly efficient and site-specific crosslinks.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge Kristen Williams for her assistance in manuscript preparation. This work is supported by NIH grant R01 GM062159 (P.G.S.). This is manuscript 29441 of The Scripps Research Institute.

References

1. a) Bagley MC, Dale JW, Merritt EA, Xiong X. Chem. Rev. 2005; 105:685–714. [PubMed: 15700961] b) Arnison PG, et al. Nat. Prod. Rep. 2013; 30:108–160. [PubMed: 23165928]
2. a) Yang P-Y, Zou H, Chao E, Sherwood L, Nunez V, Keeney M, Gharthey-Tagoe E, Ding Z, Quirino H, Luo X, Welzel G, Chen G, Singh P, Woods AK, Schultz PG, Shen W. Proc. Natl. Acad. Sci. U.S.A. 2016; 113:4140–4145. [PubMed: 27035989] b) Muppidi A, Zou H, Yang PY, Chao E, Sherwood L, Nunez V, Woods AK, Schultz PG, Lin Q, Shen W. ACS Chem. Biol. 2016; 11:324–328. [PubMed: 26727558] c) Lautrette G, Touti F, Lee HG, Dai P, Pentelute BL. J. Am. Chem. Soc. 2016; 138:8340–8343. [PubMed: 27332147] d) Zhao L, Ehrt C, Koch O, Wu Y-W. Angew. Chem. Int. Ed. 2016; 55:8129–8133. e) Kim JW, Cochran FV, Cochran JR. J. Am. Chem. Soc. 2015; 137:6–9. [PubMed: 25486381] f) Muppidi A, Doi K, Edwardraja S, Drake EJ, Gulick AM, Wang H-G, Lin Q. J. Am. Chem. Soc. 2012; 134:14734–14737. [PubMed: 22920569]
3. a) Wang D, Chen K, Kulp JL, Arora PS. J. Am. Chem. Soc. 2006; 128:9248–9256. [PubMed: 16834399] b) Speltz TE, Fanning SW, Mayne CG, Fowler C, Tajkhorshid E, Greene GL, Moore TW. Angew. Chem. Int. Ed. 2016; 55:4252–4255.
4. Shah NH, Muir TW. Chem. Sci. 2014; 5:446–461. [PubMed: 24634716]
5. a) Xiao H, Schultz PG. Cold Spring Harb Perspect Biol. 2016; 8b) Liu CC, Schultz PG. Annu. Rev. Biochem. 2010; 79:413–444. [PubMed: 20307192]
6. a) Xiang Z, Ren H, Hu YS, Coin I, Wei J, Cang H, Wang L. Nat. Meth. 2013; 10:885–888. b) Xiang Z, Lacey VK, Ren H, Xu J, Burban DJ, Jennings PA, Wang L. Angew. Chem. Int. Ed. 2014; 53:2190–2193. c) Hoppmann C, Maslennikov I, Choe S, Wang L. J. Am. Chem. Soc. 2015;

- 137:11218–11221. [PubMed: 26301538] d) Chen X-H, Xiang Z, Hu YS, Lacey VK, Cang H, Wang L. ACS Chem. Biol. 2014; 9:1956–1961. [PubMed: 25010185]
7. a) Lee Y-J, Wu B, Raymond JE, Zeng Y, Fang X, Wooley KL, Liu WR. ACS Chem. Biol. 2013; 8:1664–1670. [PubMed: 23735044] b) Li F, Zhang H, Sun Y, Pan Y, Zhou J, Wang J. Angew. Chem. Int. Ed. 2013; 52:9700–9704. c) Furman JL, Kang M, Choi S, Cao Y, Wold ED, Sun SB, Smider VV, Schultz PG, Kim CH. J. Am. Chem. Soc. 2014; 136:8411–8417. [PubMed: 24846839]
8. Xuan W, Li J, Luo X, Schultz PG. Angew. Chem. Int. Ed. 2016; 55:10065–10068.
9. Liu T, Wang Y, Luo X, Li J, Reed SA, Xiao H, Young TS, Schultz PG. Proc. Natl. Acad. Sci. U.S.A. 2016; 113:5910–5915. [PubMed: 27162342]
10. Nani RR, Gorka AP, Nagaya T, Kobayashi H, Schnermann MJ. Angew. Chem. Int. Ed. 2015; 54:13635–13638.
11. Wan W, Tharp JM, Liu WR. Biochim. Biophys. Acta, Proteins Proteomics. 2014; 1844:1059–1070.
12. a) Groff D, Chen PR, Peters FB, Schultz PG. ChemBioChem. 2010; 11:1066–1068. [PubMed: 20422671] b) Kim CH, Kang M, Kim HJ, Chatterjee A, Schultz PG. Angew. Chem. Int. Ed. 2012; 51:7246–7249. c) Xiao H, Xuan W, Shao S, Liu T, Schultz PG. ACS Chem. Biol. 2015; 10:1599–1603. [PubMed: 25909834]
13. a) Luo X, Zambaldo C, Liu T, Zhang Y, Xuan W, Wang C, Reed SA, Yang P-Y, Wang RE, Javahishvili T, Schultz PG, Young TS. Proc. Natl. Acad. Sci. U.S.A. 2016; 113:3615–3620. [PubMed: 26976568] b) Chen PR, Groff D, Guo J, Ou W, Cellitti S, Geierstanger BH, Schultz PG. Angew. Chem. Int. Ed. 2009; 48:4052–4055. c) Elliott TS, Bianco A, Chin JW. Curr. Opin. Chem. Biol. 2014; 21:154–160. [PubMed: 25159020] d) Hancock SM, Uprety R, Deiters A, Chin JW. J. Am. Chem. Soc. 2010; 132:14819–14824. [PubMed: 20925334]
14. Högbom M, Eklund M, Nygren P-Å, Nordlund P. Proc. Natl. Acad. Sci. U.S.A. 2003; 100:3191–3196. [PubMed: 12604795]

Fig 1. Genetic incorporation of PheK and FPheK into proteins in *E. coli* and mammalian cells. A) Fluorescence analysis of the ability of PheKRS1 to incorporate PheK and FPheK into sfGFP (Tyr151TAG) in *E. coli* DH10B. 2 mM ncAA was used; B) SDS-PAGE analysis of sfGFP (Tyr151TAG) expressed in the presence or absence of 2 mM PheK and FPheK; C) Mass spectrum of sfGFP mutants with PheK or FPheK at Tyr151; D) Fluorescence microscopy of HEK293T cells expressing an EGFP mutant (Tyr39TAG) in the presence or absence of 1 mM ncAAs.

Fig 2.

Intramolecular crosslinking of FPheK with Lys in thioredoxin. A) ESI-QTOF analysis of thioredoxin Gln98FPheK mutant before and after crosslinking. The mass loss of 113 Da was consistent with the release of 4-fluorophenol (Exp. 112 Da); B) SDS-PAGE analysis of thioredoxin Gln98FPheK mutant before and after crosslinking.

Fig 3.

Intramolecular crosslinking of FPheK at residue 98 with Cys and Tyr at residue 90 respectively in thioredoxin. A) Mass analysis of crosslinking between FPheK and Cys. The appearance of 12496.98 Da indicated the loss of 4-fluorophenol and formation of a thiocarbamate bond; B) Mass analysis of crosslinking between FPheK and Tyr. The appearance of 12556.72 Da indicated the loss of 4-fluorophenol and the formation of a new carbamate bond; small amount of byproduct (12530.67 Da) indicated the formation of free lysine due to hydrolysis.

Fig 4.

Intermolecular crosslinking of Z protein (Glu25FPheK) to Afb-Trx. A) Crystal structure of the Z protein-Afb complex (PDB: 1LP1) reveals that Glu25 in Z protein and Lys8 in Afb are in close proximity at the protein interface; B) SDS-PAGE analysis of crosslink between Z protein (Glu25FPheK) and Afb-Trx shows that the reaction was complete in 8 hours; C) MS/MS fragmentation of Glu-C digested peptides derived from the Z protein/Afb-Trx crosslinking product. A complete list of observed fragments is available in table S2; the peak at 737.0576 corresponds to the crosslinked peptide (+ 3H) after digestion (Cal. 737.05887). X denotes FPheK.

Fig 5.

Construction of a urea crosslink between the heavy chain and light chain of Herceptin Fab.

A) Crystal structure of Herceptin Fab (PDB: 1N8Z) reveals that Gln100 of light chain and Lys43 of heavy chain are in close proximity at the chain interface; B) SDS-PAGE analysis of wild type Herceptin Fab and its FPheK mutant. The crosslink was induced by incubating the Fab mutant in pH 8.5 HEPES buffer at 37 °C for 24 hours; C) ESI-QTOF analysis of Herceptin Fab FPheK mutant before and after crosslinking along with the interchain disulfide bond cleaved by DTT. The mass variation observed here was likely due to the partial cleavage of intra-chain disulfide bonds by DTT.

scheme 1.

ncAAs with aryl carbamate side chains used in this work and a crosslinking reaction demonstrated with Lys.