Noncanonical Amino Acids

Deutsche Ausgabe: DOI: 10.1002/ange.201604891
Internationale Ausgabe: DOI: 10.1002/anie.201604891

Genetic Incorporation of a Reactive Isothiocyanate Group into Proteins

*Weimin Xuan, Jack Li, Xiaozhou Luo, and Peter G. Schultz**

Abstract: *Methods for the site-specific modification of proteins are useful for introducing biological probes into proteins and engineering proteins with novel activities. Herein, we genetically encode a novel noncanonical amino acid (ncAA) that contains an aryl isothiocyanate group which can form stable thiourea crosslinks with amines under mild conditions. We show that this ncAA (pNCSF) allows the selective conjugation of proteins to amine-containing molecular probes through formation of a thiourea bridge. pNCSF was also used to replace a native salt bridge in myoglobin with an intramolecular crosslink to a proximal Lys residue, leading to increased thermal stability. Finally, we show that pNCSF can form stable intermolecular crosslinks between two interacting proteins.*

Methods for the site-specific introduction of biophysical probes and post-translational modifications into proteins have become important tools for the study of protein function both inside and outside the cell.[1] These methods have also been used to engineer proteins with improved pharmacological, biological or enzymatic activities, for example by the selective conjugation of polyethylene glycols, drugs or cofactors to proteins.[2] A number of different methods have been developed to selectively modify proteins, including the use of enzymes to modify defined sequence motifs within a protein,[3] the modification of uniquely reactive cysteine residues in proteins[4] and the introduction of noncanonical amino acids (ncAAs) through semisynthetic[5] or recombinant methods.[6] In general, highly selective methods that minimally perturb the structure of the protein are most desirable, so as not to affect protein function, introduce immunogenic epitopes, or create heterogeneous mixtures of proteins with distinct activities.

We and others have exploited recombinant methods to introduce ncAAs site-specially into proteins in response to unique nonsense or frameshift codons by means of orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pairs. This method has been used to introduce amino acids with reactive side chains into proteins that can selectively react with exogenous agents or form intramolecular crosslinks within proteins. Examples include amino acids with reactive halides (e.g., α-haloketones and haloalkanes),[7] aryl ketones[8] and boronate groups,[9] Michael acceptors (e.g., α, β-unsaturated carboxylamide and sulfonamide groups),[10] alkynes[11] and azides[12] for "click" reactions, and photoreactive side chains (e.g., aryl azides, benzophenone and diazirines).[12a,13] Here, we report the genetic encoding in *E. coli* of an unnatural amino acid with a phenyl isothiocyanate side chain, and show that this amino acid can be selectively modified with biophysical probes and form stable intra- or intermolecular crosslinks with native Lys residues in proteins under mild conditions.

Phenyl isothiocyanates have long been used to modify the N-terminal amino group of proteins in the Edman degradation.[14] This functional group has attenuated reactivity compared to other electrophilic groups that rapidly react with glutathione (GSH) and other cellular thiol-containing molecules. Indeed, the second-order reaction rate constant between fluorescein isothiocyanate and propylamine was determined to be $(2.9 \pm 0.2) \times 10^{-2}\,\text{M}^{-1}\text{s}^{-1}$ at 37 °C in pH 7.4 DPBS (Supporting Information (SI)). Furthermore, in the presence of 1 mM pNCSF. The growth rate of *E. coli* in LB broth was only slightly affected (Figure S1 in the SI). To incorporate this functional group into proteins, we synthesized the structurally similar analogue of phenylalanine, *para*-isothiocyanate phenylalanine (*p*NCSF, Figure 1 a), in two steps in 73 % overall yield (see SI). To genetically encode *p*NCSF, we screened several polyspecific aaRSs that have previously been shown to incorporate a number of structur-

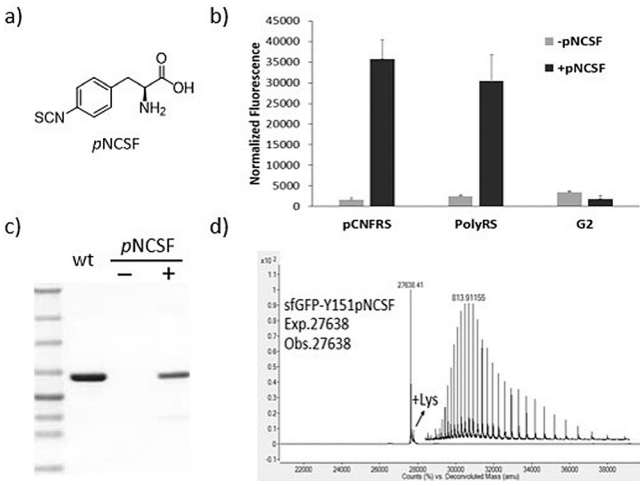

Figure 1. a) Structure of *p*NCSF. b) Fluorescence screening of three poly-specific tRNA synthetase for their ability to incorporate *p*NCSF into proteins in *E. coli*. c) SDS-PAGE analysis of sfGFP (151TAG) expressed with *p*CNFRS in the presence or absence of *p*NCSF. d) Mass spectrum of sfGFP (Y151*p*NCSF). The inset is the mass spectrum before deconvolution.

[*] Dr. W. Xuan, J. Li, X. Luo, Prof. P. G. Schultz
Department of Chemistry, The Scripps Research Institute
10550 N. Torrey Pines Road, La Jolla, CA 92037 (USA)
E-mail: schultz@scripps.edu

Supporting information for this article can be found under: http://dx.doi.org/10.1002/anie.201604891.

ally related ncAAs, but not the common twenty canonical amino acids. Specifically, we screened three polyspecific aaRSs derived from the *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase, (*p*CNFRS,[15] PolyRS[10b,16] and G2[17]) for their ability to aminoacylate *p*NCSF. To assess suppression efficiency, the fluorescent protein sfGFP with an amber stop codon (TAG) at a permissive site (Y151) was co-transformed into *E. coli* DH10B along with the aaRS/$tRNA_{CUA}$ pair. As shown in Figure 1b, cells transformed with *p*CNFRS or PolyRS showed significant fluorescence in the presence of 1 mM *p*NCSF; in the absence of *p*NCSF, no increase in fluorescence was observed. These results suggest that these aaRSs can utilize *p*NCSF as a substrate. We chose *p*CNFRS for all subsequent experiments. To confirm the selective incorporation of *p*NCSF, a C-terminal His-tagged mutant of sfGFP (Y151*p*NCSF) was expressed in *E. coli* DH10B, and purified by Ni-NTA column in 16 mg L^{-1} isolated yield. The mutant protein was analyzed by SDS-PAGE gel (Figure 1c) and electrospray ionization quadrupole time-of-flight mass spectrometry (ESI-QTOF Mass); the observed mass (27638 Da) was consistent with the expected mass (27638 Da) (Figure 1d). A small amount of the sfGFP mutant was observed as a Lys adduct (Figure 1d).[18] After 3 month storage at 4 °C, no obvious multimerization was observed.

*p*NCSF should serve as a facile and selective handle for conjugation with molecules that contain primary or secondary amines. To demonstrate this notion, we attempted to conjugate two amine-containing small molecules, glucosamine and dansyl-KG, to a sfGFP mutant in which the surface Y151 was substituted with *p*NCSF (Figure 2b). After co-incubation

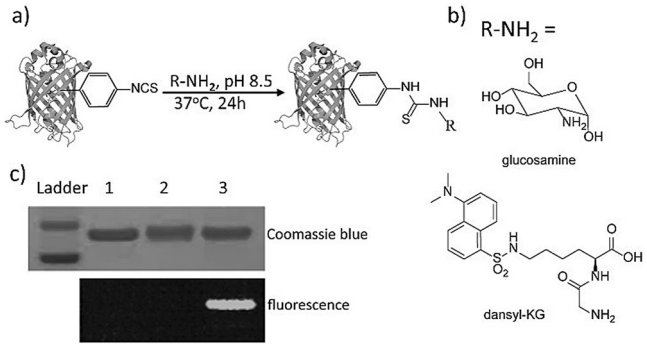

Figure 2. a) Labelling of sfGFP by the genetic incorporation of *p*NCSF at Y151. b) Structure of glucosamine and dansyl-KG used in labelling experiments. c) SDS-PAGE gel was visualized by coomassie blue staining (top) and fluorescence emission (bottom). Lane 1: wt sfGFP; Lane2: sfGFP labelled with glucosamine; Lane 3: sfGFP labelled with dansyl-KG.

of the sfGFP mutant (7.5 μM) and glucosamine or dansyl-KG (100 equiv) at pH 8.5, 37 °C for 24 hours, successful labelling was confirmed by ESI-QTOF Mass; the efficiency was estimated to be over 80% (Figure S2). SDS-PAGE was performed, and the dansyl-KG labelled sfGFP was further verified by on-gel fluorescence (Figure 2c). Importantly, this method avoids the low pH or catalysts required for oxime formation, and either the metal catalysts or ring-strained alkynes that are required for efficient click reactions. Using this approach, it should be possible to use a large number of existing amine-containing molecules directly for the site-specific modification of *p*NCSF-encoded proteins.

Next we explored whether *p*NCSF could be used to form intramolecular thiourea crosslinks with Lys residues within a protein. Myoglobin has a salt bridge between K17 and D123 (Figure S3). Replacement of this salt bridge with a covalent bond might be expected to increase the thermal stability of the protein. To test this hypothesis, D123 was substituted with *p*NCSF in C-terminal His-tagged myoglobin. As a negative control, the D123OMeY (OMeY: O-Methyl tyrosine) and K17N/D123*p*NCSF myoglobin mutants were also expressed. All mutants were successfully expressed and purified by Ni-NTA chromatography in good yields (8 mg L^{-1}, 13 mg L^{-1} and 5 mg L^{-1}, respectively); incorporation of ncAAs was confirmed by ESI-QTOF Mass (Figure S4), and no Lys adduct was observed in two *p*NCSF mutants. SDS-PAGE (Figure 3a)showed a higher mobility band for the D123*p*NCSF

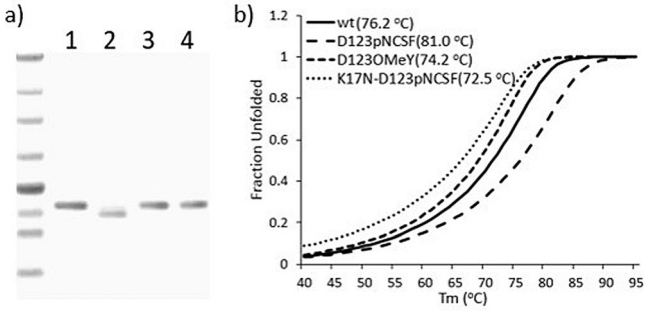

Figure 3. Intramolecular crosslinking within myoglogin. a) SDS-PAGE of wt myoglobin and its mutants. Lane 1: wt; Lane 2: D123*p*NCSF; Lane 3: D123OMeY; Lane 4: K17N-D123*p*NCSF. b) Circular dichroism (CD) melting curves of wt myoglobin and its mutants. The Tm value was determined based on the maximum slope.

mutant compared to wt myoglobin and the other two mutants, consistent with a more compact structure under denaturing conditions. Because intramolecular protein crosslinks are known to enhance protein thermal stability, we measured the melting temperature of myoglobin and its mutants by circular dichroism (CD) (Figure 3b). The *p*NCSF mutant had a CD melting temperature of 81.0 °C compared to 76.2 °C for wt myoglobin. Conversely, both the D123OMeY and K17N/D123*p*NCSF mutants showed a reduced CD melting temperature likely due to the loss of this salt bridge. It is noteworthy that this crosslinking occurred during expression and purification, and no extra treatment was necessary. Unfortunately, attempts to proteolyze and confirm crosslinking of *p*NCSF to K17 were unsuccessful (likely due to the crosslink).

Finally, we explored whether *p*NCSF can also form an intermolecular crosslink between two interacting proteins. Z protein and the Z_{SPA} affibody (Afb) are two small proteins (< 8 *k*D) that form a complex with a K_d value of 6 μM.[19] On the basis of the x-ray co-crystal structure, we hypothesized that substitution of E25 with *p*NCSF in Z protein would lead to formation of a thiourea bond with K8 in Afb (Figure 4a).

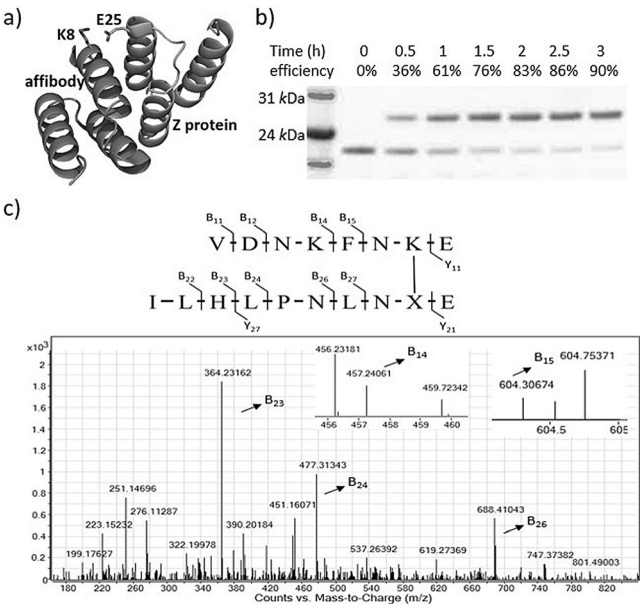

Figure 4. Intermolecular crosslinking of Z protein (E25*p*NCSF) to Afb-Trx. a) Crystal structure of the Z protein-Afb complex (PDB: 1LP1) reveals that E25 in Z protein and K8 in Afb are in close proximity at the protein interface. b) SDS-PAGE analysis of crosslink between Z protein (E25*p*NCSF) and Afb-Trx fusion. Crosslinking efficiency was measured based on relative band intensities of Afb-Trx. c) MS/MS fragmentation of Glu-C digested peptides derived from the Z protein/Afb-Trx crosslinking product. The spectra confirm that crosslinking occurred between *p*NCSF and K8 in Afb-Trx. X denotes *p*NCSF.

We fused *E. coli* thioredoxin (Trx) to the C-terminal of Afb (Afb-Trx) followed by a His-tag to simplify purification. The Z protein E25*p*NCSF mutant was expressed and purified by Ni-NTA chromatography (9 mg L^{-1}). *p*NCSF incorporation was further verified by SDS-PAGE and ESI-QTOF Mass (Figure S5 and S6), and no Lys adduct was observed. To form an intermolecular crosslink, Afb-Trx (5 μM) was co-incubated with 3 equiv of Z protein mutant at 37 °C in 200 mM HEPES buffer at pH 7.4, and samples were taken at different time points. Based on SDS-PAGE analysis (Figure 4b), a new band appeared at 0.5 h with an expected size of 27 kD, consistent with the formation of Z protein/Afb-Trx complex; after 3 hours, the crosslinking efficiency was calculated to be 90%. The formation of the Z protein/Afb-Trx complex was also confirmed by ESI-QTOF analysis with a peak (26745.48 kD) corresponding to the molecular weight of the Z protein-Afb-Trx heterodimer (26746.15 kD) (Figure S7). As a negative control, reaction of the E25*p*NCSF Z protein mutant with the Afb-Trx K8N mutant did not afford a crosslinked product (Figure S8), indicating that both *p*NCSF and the corresponding Lys are required for crosslink formation. Finally, crosslinking between *p*NCSF and K8 of Afb-Trx was further confirmed by Glu-C protease digestion and LC-MS/MS analysis (Figure 4c, Figure S9 and Table S3). Because Lys is highly abundant on protein surfaces, the use of *p*NCSF may provide a useful tool to study protein/protein interactions.

In summary, we have genetically incorporated a novel ncAA that can react with amine-containing molecules with excellent selectivity under weakly basic conditions. *p*NCSF can also efficiently form intra- or inter-molecular crosslinks with proximal Lys residues in proteins. This work provides another useful tool for the selective modification of proteins, and is especially useful for crosslinking to native lysine residues in proteins.

Acknowledgements

This work was supported by Department of Energy Grant DE-SC0011787. This is paper no 29348 at The Scripps Research Institute.

Keywords: crosslink · isothiocyanate · lysine · noncanonical amino acid · thiourea

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 10065–10068
Angew. Chem. **2016**, *128*, 10219–10222

[1] N. Krall, F. P. da Cruz, O. Boutureira, G. J. L. Bernardes, *Nat. Chem.* **2016**, *8*, 103–113.
[2] a) T. Liu, J. Du, X. Luo, P. G. Schultz, F. Wang, *Curr. Opin. Chem. Biol.* **2015**, *28*, 66–74; b) S. B. Sun, P. G. Schultz, C. H. Kim, *ChemBioChem* **2014**, *15*, 1721–1729.
[3] a) I. S. Carrico, B. L. Carlson, C. R. Bertozzi, *Nat. Chem. Biol.* **2007**, *3*, 321–322; b) A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, *Nat. Biotechnol.* **2003**, *21*, 86–89; c) M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner, H. L. Ploegh, *Nat. Chem. Biol.* **2007**, *3*, 707–708; d) L. W. Miller, J. Sable, P. Goelet, M. P. Sheetz, V. W. Cornish, *Angew. Chem. Int. Ed.* **2004**, *43*, 1672–1675; *Angew. Chem.* **2004**, *116*, 1704–1707; e) M. Fernández-Suárez, H. Baruah, L. Martínez-Hernández, K. T. Xie, J. M. Baskin, C. R. Bertozzi, A. Y. Ting, *Nat. Biotechnol.* **2007**, *25*, 1483–1487; f) J. Yin, F. Liu, X. Li, C. T. Walsh, *J. Am. Chem. Soc.* **2004**, *126*, 7754–7755.
[4] a) C. Zhang, M. Welborn, T. Zhu, N. J. Yang, M. S. Santos, T. Van Voorhis, B. L. Pentelute, *Nat. Chem.* **2016**, *8*, 120–128; b) C. P. Ramil, P. An, Z. Yu, Q. Lin, *J. Am. Chem. Soc.* **2016**; c) M. D. Simon, F. Chu, L. R. Racki, C. C. de la Cruz, A. L. Burlingame, B. Panning, G. J. Narlikar, K. M. Shokat, *Cell* **2007**, *128*, 1003–1012; d) P. M. S. D. Cal, G. J. L. Bernardes, P. M. P. Gois, *Angew. Chem. Int. Ed.* **2014**, *53*, 10585–10587; *Angew. Chem.* **2014**, *126*, 10758–10760.
[5] a) N. H. Shah, T. W. Muir, *Chem. Sci.* **2014**, *5*, 446–461; b) S. Bondalapati, M. Jbara, A. Brik, *Nat. Chem.* **2016**, *8*, 407–418.
[6] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–444.
[7] a) Z. Xiang, H. Ren, Y. S. Hu, I. Coin, J. Wei, H. Cang, L. Wang, *Nat. Methods* **2013**, *10*, 885–888; b) Z. Xiang, V. K. Lacey, H. Ren, J. Xu, D. J. Burban, P. A. Jennings, L. Wang, *Angew. Chem. Int. Ed.* **2014**, *53*, 2190–2193; *Angew. Chem.* **2014**, *126*, 2222–2225.
[8] L. Wang, Z. Zhang, A. Brock, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 56–61.
[9] E. Brustad, M. L. Bushey, J. W. Lee, D. Groff, W. Liu, P. G. Schultz, *Angew. Chem. Int. Ed.* **2008**, *47*, 8220–8223; *Angew. Chem.* **2008**, *120*, 8344–8347.
[10] a) Y.-J. Lee, B. Wu, J. E. Raymond, Y. Zeng, X. Fang, K. L. Wooley, W. R. Liu, *ACS Chem. Biol.* **2013**, *8*, 1664–1670; b) J. L. Furman, M. Kang, S. Choi, Y. Cao, E. D. Wold, S. B. Sun, V. V. Smider, P. G. Schultz, C. H. Kim, *J. Am. Chem. Soc.* **2014**, *136*, 8411–8417.
[11] A. Deiters, P. G. Schultz, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1521–1524.
[12] a) J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 9026–9027; b) D. P.

Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 8720 – 8721.
[13] a) J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11020 – 11024; b) S. Lin, Z. Zhang, H. Xu, L. Li, S. Chen, J. Li, Z. Hao, P. R. Chen, *J. Am. Chem. Soc.* **2011**, *133*, 20581 – 20587; c) E. M. Tippmann, W. Liu, D. Summerer, A. V. Mack, P. G. Schultz, *ChemBioChem* **2007**, *8*, 2210 – 2214.
[14] P. Edman, *Acta Chem. Scand.* **1950**, *4*, 283 – 293.
[15] D. D. Young, T. S. Young, M. Jahnz, I. Ahmad, G. Spraggon, P. G. Schultz, *Biochemistry* **2011**, *50*, 1894 – 1900.
[16] H. Xiao, F. Nasertorabi, S.-h. Choi, G. W. Han, S. A. Reed, R. C. Stevens, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6961 – 6966.
[17] R. B. Cooley, P. A. Karplus, R. A. Mehl, *ChemBioChem* **2014**, *15*, 1810 – 1819.
[18] B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, J. D. Rabinowitz, *Nat. Chem. Biol.* **2009**, *5*, 593 – 599.
[19] M. Högbom, M. Eklund, P.-Å. Nygren, P. Nordlund, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3191 – 3196.

Received: May 18, 2016
Revised: June 17, 2016
Published online: July 15, 2016

