ChemComm

Cite this: *Chem. Commun.*, 2011, **47**, 4502–4504

www.rsc.org/chemcomm

COMMUNICATION

A readily synthesized cyclic pyrrolysine analogue for site-specific protein "click" labeling†

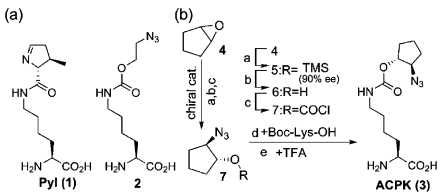
Ziyang Hao,[a] Yanqun Song,[a] Shixian Lin,[a] Maiyun Yang,[a] Yujie Liang,[a] Jing Wang[b] and Peng R. Chen*[a]

Received 3rd January 2011, Accepted 16th February 2011
DOI: 10.1039/c1cc00024a

A concise route was developed for the facile synthesis of a cyclic pyrrolysine analogue bearing an azide handle. Directed evolution enabled the encoding of this non-natural amino acid in both prokaryotic and eukaryotic cells, which offers a highly efficient approach for the site-specific protein labeling using click chemistry.

The bioorthogonal ligation chemistry, such as "click" reaction, has emerged as a valuable tool for labeling of various biomolecules in living systems.[1] Unique functional groups involved in these reactions have been specifically incorporated into target biomolecules, which allows the subsequent bioorthogonal conjugation with probes bearing complementary functionality. Extensive efforts have been focused on applying this strategy for specific labeling of proteins, the most abundant biomolecules in a cell. Bioorthogonal moieties such as azide and alkyne were added to proteins by diverse methods.[2] In particular, a genetic code expansion approach has allowed the incorporation of unnatural amino acids (UAAs) containing these functional groups into proteins in a site-specific as opposed to the residue specific manner.[3] However, an on-going challenge for this strategy is to produce large amounts of recombinant proteins carrying easily accessible UAAs cost-effectively, especially for proteins expressed in mammalian cells.

Pyrrolysine (Pyl, Scheme 1a), the 22nd naturally occurring amino acid encoded by a pyrrolysyl-tRNA synthetase (PylRS)–$tRNA^{Pyl}_{CUA}$ pair in archaea species, had recently been adapted for encoding various UAAs in response to an in-frame amber codon in bacteria, mammalian cells, and yeast.[4] A unique feature of this PylRS–$tRNA^{Pyl}_{CUA}$ pair is its orthogonality in both prokaryotic and eukaryotic organisms, making it an attractive facile system for directly utilizing aminoacyl-tRNA synthetases evolved in *E. coli* to expand the genetic code of eukaryotic cells.[4,5] Employing this method, a range of UAAs containing alkyne or azide groups have been synthesized and incorporated into proteins including a direct cyclic mimic of Pyl.[2e] However, the 16-step long synthesis with an overall yield of 17% limits its application.[6] The usage of an aromatic azido-Pyl derivative has also been limited for the similar reason.[2c] Chin and Deiters *et al*. showed that aliphatic Pyl analogues containing azide (**2**) or alkyne groups can be incorporated into proteins by the wild-type (wt) PylRS.[2d] The synthesis is less tedious and the incorporation efficiency is improved. This linear alkyne analogue was further utilized for making glycosylated proteins through the click reaction.[7] Despite these progresses, since the parent compound Pyl contains a cyclic moiety that is selectively recognized by the natural PylRS, we surmise that certain cyclic analogues of Pyl could still be more sterically favoured and advantageous over linear Pyl analogues in terms of amber suppression efficiency.



Scheme 1 (a) The structures of pyrrolysine (Pyl, **1**) and the pyrrolysine analogue **2**. (b) Synthetic route to ACPK (**3**).

We envision that, if properly designed, a cyclic mimic of Pyl could be readily synthesized and then incorporated into proteins with a higher yield. In order to simplify the synthesis and install a clickable functionality, we propose a new azide-bearing cyclic Pyl analogue N^{ε}-(((1*R*,2*R*)-2-azidocyclopentyloxy)carbonyl)-L-lysine (ACPK; Scheme 1b) that can be readily obtained by a concise synthetic route. Directed evolution allowed us to identify a mutant PylRS to encode this UAA efficiently in prokaryotic and eukaryotic cells, using *E. coli* acid-stress chaperone HdeA and mammalian tumor suppressor p53 as model proteins.

The idea was inspired by the fact that a series of carbamate-linker based Pyl analogues including N^{ε}-cyclopentyloxycarbonyl-L-lysine (Cyc) can be recognized by wt or mutant PylRS.[6,8] This carbamate linkage can therefore serve as a general route to connect N^{ε}-lysine with diverse alcohol containing molecules including cyclic alcohols. In previous work by Jacobsen and

[a] *Beijing National Laboratory for Molecular Sciences (BNLMS), Department of Chemical Biology, College of Chemistry, Peking University, Beijing 100871, China. E-mail: pengchen@pku.edu.cn; Fax: +86 10-62767433; Tel: +86 10-62755773*
[b] *Department of Chemistry, The University of Chicago, Chicago, IL 60637, USA*

† Electronic supplementary information (ESI) available: See DOI: 10.1039/c1cc00024a

co-workers, a practical approach for highly efficient synthesis of enantiopure cyclic 1,2-amino alcohols *via* catalytic asymmetric ring opening (ARO) of meso epoxides has been developed.[9] The chiral (salen) Cr(III) complex catalyzes the formation of azido silyl ether with high enantioselectivity (>90% ee; Fig. S1, ESI†) and no detectable byproducts. Adopting this method, we synthesized *trans*-1,2-azido alcohol (**6**) from an inexpensive, readily available five-membered ring-fused meso epoxide (**4**), which subsequently reacted with triphosgene to yield **7** (Scheme 1b). Coupling of Boc-Lys-OH with **7** followed by the treatment with TFA gave the final product **3** with 70% overall yield (5 steps). The readily accessible starting materials, high volumetric productivity, ease of product isolation and the potential to recycle the catalysts with no loss of catalytic activity render this ARO method a facile, cost-effective and easily scalable strategy for the synthesis of **3**.

We next tested whether **3** can be incorporated into proteins in *E. coli* by the wt-PylRS from *M. barkeri* (MbPylRS). A C-terminal His-tagged GFP harboring an amber mutation (TAG) at residue Asp149 (GFP-N149TAG) was employed as the model protein. We found by SDS-PAGE analysis that although **3** can be recognized by wt-MbPylRS, the yield of full-length GFP containing **3** at the 149 position (GFP-N149-**3**) is much lower than GFP containing **2** at the same position using wt-MbPylRS (Fig. 1a). This result is consistent with the low recognition efficiency of Cyc by wt-MmPylRS (<1 mg L^{-1} GFP in the presence of 1 mM Cyc) reported previously.[4a] In order to improve the incorporation efficiency, we turned to alter the MbPylRS active site pocket for optimal recognition of **3**. Five residues (Leu270, Tyr271, Leu274, Cys313, Tyr349) surrounding the binding site of the methyl pyrroline ring on Pyl were randomized to afford a mutant library (3×10^7 diversity) and directed evolution was applied according to the previously published protocol (ESI†).[4a] Two MbPylRS mutants were enriched after successfully passing through the selection and they are designated as ACPK-RS and ACPK-RS-A, respectively (Table S1, ESI†). Because ACPK-RS exhibited a higher amber suppression efficiency relative to ACPK-RS-A, it was chosen for further exploration.

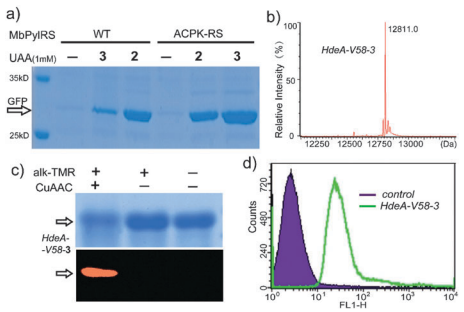


Fig. 1 Genetic incorporation of ACPK (**3**) into proteins by the evolved MbPylRS–$tRNA^{Pyl}_{CUA}$ mutant pair in *E. coli*. (a) SDS-PAGE analysis demonstrating the incorporation efficiencies of **2** and **3** into GFP in *E. coli*. (b) ESI-MS analysis of the full-length HdeA-His_6 protein containing **3** at residue V58. (c) Fluorescent labeling of the purified HdeA-V58-**3** protein by CuAAC and alk-TMR. (d) Flow cytometry results on biocompatible CuAAC-mediated fluorescent labeling of living *E. coli* cells expressing HdeA-V58-**3** or expressing wt-HdeA as a control.

The incorporation efficiency and fidelity of ACPK-RS on **3** were determined in *E. coli*. Protein expression was carried out in BL21-DE3 cells co-transformed with plasmids expressing ACPK-RS–$tRNA^{Pyl}_{CUA}$ pair and GFP-N149TAG, and the LB medium was supplemented with 1 mM **3**, or 1 mM **2**, or without UAA. SDS-PAGE analysis on the purified proteins from Ni-NTA chromatography showed that the expression level of GFP-N149-**3** is about two-fold higher than that of GFP-N149-**2** produced by ACPK-RS (Fig. 1a; the calculated yield of full-length GFP-N149-**3** is ~10 mg L^{-1}). Furthermore, GFP-N149-**2** produced by ACPK-RS is slightly less than that by wt-MbPylRS, consistent with the previous report that no mutant MbPylRS was identified with improved incorporation efficiency on **2** than wt-MbPylRS.[6] Taken together, these data demonstrate that **3** can be incorporated into proteins by the newly evolved ACPK-RS with a higher yield than the previously reported azido-Pyl analogues and wt-MbPylRS.

In order to demonstrate that **3** is a functional handle for protein click labeling *in vitro* and *in vivo*, we choose to incorporate **3** into HdeA, an *E. coli* periplasmic chaperone, preventing the acid induced protein aggregation. HdeA plays essential roles for enteric bacterial pathogens to pass through the extremely acidic mammalian stomach (pH 1–3) before reaching their primary infection sites.[10] We introduced **3** at different sites on HdeA and this didn't affect HdeA's dimer formation at pH = 7 as confirmed by native PAGE gel (Fig. S2, ESI†). Electrospray ionization mass spectrometry (ESI-MS) of purified full-length HdeA-V58-**3** revealed a single peak at 12811 Da (expected: 12812 Da; Fig. 1b, ESI†), confirming that **3** was not modified inside of *E. coli*. We then performed copper-induced azide–alkyne cycloaddition (CuAAC) on HdeA-V58-**3** with an alkyne-modified tetramethylrhodamine fluorophore *in vitro* (alk-TMR; Fig. S3, ESI†).[11] Indeed, we only observed a fluorescence band in the presence of alk-TMR (two-fold excess), sodium ascorbate and $CuSO_4$ (Fig. 1c, ESI†). Furthermore, by employing the recently developed biocompatible Cu(I) ligand BTTES, we conducted CuAAC on ACPK-incorporated HdeA in living conditions.[12] BTTES had been recently shown to not only accelerate CuAAC significantly, but also it converted this original cytotoxic reaction to be adaptable for living mammalian cells.[12] *E. coli* cells expressing HdeA-V58-**3** were treated with an alkyne-functionalized coumarin dye and CuAAC reagents (BTTES, sodium ascorbate and $CuSO_4$), followed by flow cytometric analysis (Fig. 1d, ESI†). The fluorescent labeled cells have median fluorescence intensity 10 fold greater than the background (wt-*E. coli* control), while no apparent toxicity was observed on *E. coli* cells under this reaction condition. This result was further confirmed by the SDS-PAGE analysis on *E. coli* cell lysates showing a single fluorescent band corresponding to the fluorescent labeled HdeA-V58-**3** (Fig. S4, ESI†). Therefore, this BTTES-assisted CuAAC can be applied in conjunction with the ACPK handle for protein click labeling in living bacterial cells.

The evolved ACPK-RS–$tRNA^{Pyl}_{CUA}$ pair from *E. coli* was then transferred into mammalian cells and we choose to incorporate **3** into the tumor suppressor protein p53 for proof-of-concept. The MbPylRS synthetase gene from the previously

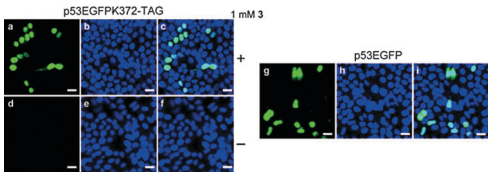


Fig. 2 Encoding ACPK (**3**) in mammalian cells. Confocal microscopy of the expression of p53EGFP-K372-**3** in HEK293T cells harboring ACPK-RS–$tRNA^{Pyl}_{CUA}$ pair in the presence (a–c) or absence (d–f) of **3**. Wt-p53EGFP expressed in HEK293T cells was used as a control (g–i). (a, d, g): GFP channel; (b, e h): blue channel; (c, f, i): merging of GFP and blue channels. All scale bars are 20 μm.

developed mammalian expression plasmid pCMV-NBK-1[4a] was mutated to afford the plasmid pCMV-ACPK-RS. This plasmid was co-transfected into HEK293T cells with a plasmid encoding p53EGFP harboring a single amber mutation (TAG) at the C-terminal Lys372 site. Confocal microscopy was applied to HEK293T cells expressing this mutant protein (p53EGFP-K372-**3**) produced by ACPK-RS–$tRNA^{Pyl}_{CUA}$ pair. Green fluorescence was only observed when cells were supplemented with 1 mM **3** (Fig. 2). The nucleus stain DAPI was then applied to these cells and HEK293T cells expressing wt-p53EGFP were used as control. Similar to p53EGFP, p53EGFP-K372-**3** was located within the cell nucleus, confirming that the introduction of **3** at this position on p53 didn't disrupt its sub-cellular localization. The incorporation efficiency of **3** onto p53 was estimated to be around 50% of the wt-p53 as measured by immunoblotting analysis of H1299 cells (p53-null) expressing ACPK-incorporated p53 at the Ser20 or Met340 site (Fig. S6, ESI†). Furthermore, we evaluated ACPK as a bioorthogonal handle for chemically attaching proteins onto the living cell surface. ACPK was introduced into EGFP as a model protein at residue 149 that is well exposed on the protein surface. HEK 293T cells were incubated with 50 μM peracetylated *N*-(4-pentynoyl)mannosamine (Ac4ManNAl) to metabolically incorporate the corresponding alkynyl sialic acid (SiaNAl) into their cell-surface glycoconjugates.[11b] The resulting cells bearing alkynyl groups on the surface were reacted with 15 μM ACPK-incorporated EGFP by BTTES-assisted CuAAC. Confocal fluorescence microscopy showed robust GFP fluorescence on the cell surface (Fig. S7, ESI†). Given the generic nature of this strategy, other proteins can also be specifically attached onto the mammalian cell surface.

Finally, we employed this newly evolved ACPK-RS-tRNA pair to introduced 3 into the heme recognition domain (NEAT) of a bacterial heme-transfer protein IsdA, which allowed us to conjugate DBCO-Fluor 488 near the heme binding pocket by the strain promoted azide–alkyne cycloaddition. A reversible fluorescence quenching effect was observed, caused by the direct energy transfer between heme and Fluor 488 (Fig. S5, ESI†).

In conclusion, we have developed a concise and highly efficient approach for the facile synthesis of an azide-bearing cyclic pyrrolysine analogue ACPK. Directed evolution enabled us to genetically encode this UAA in prokaryotic and eukaryotic cells, which allows the site-specific incorporation of ACPK into target proteins that are readily accessible by alkyne probes through click reactions. In particular, by adopting the recently developed biocompatible CuAAC, this click labeling strategy was realized on ACPK-incorporated proteins in living *E. coli* cells and on the surfaces of living mammalian cells. To further demonstrate the application, the ACPK-attached IsdA-NEAT protein was conjugated with DBCO-488 by the "copper free" click reaction with great efficiency. This allowed us to use fluorescent energy transfer to monitor the heme binding and release on hemoproteins. Such a readily synthesized cyclic pyrrolysine mimic might also permit the large-scale preparation of recombinant proteins containing "clickable" tags for *in vitro* and *in vivo* labeling.

This work was supported by the National Key Basic Research Foundation of China (2010CB912300), National Natural Science Foundation of China (91013005). We thank Prof. Peter Schultz for the aaRS/tRNA plasmids.

Notes and references

1 J. A. Prescher and C. R. Bertozzi, *Nat. Chem. Biol.*, 2005, **1**, 13–21.
2 (*a*) A. J. Link, M. L. Mock and D. A. Tirrell, *Curr. Opin. Biotechnol.*, 2003, **14**, 603–609; (*b*) L. Wang, J. Xie and P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.*, 2006, **35**, 225–249; (*c*) T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto and S. Yokoyama, *Chem. Biol.*, 2008, **15**, 1187–1197; (*d*) D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters and J. W. Chin, *J. Am. Chem. Soc.*, 2009, **131**, 8720–8721; (*e*) T. Fekner, X. Li, M. M. Lee and M. K. Chan, *Angew. Chem., Int. Ed.*, 2009, **48**, 1633–1635.
3 (*a*) L. Wang, A. Brock, B. Herberich and P. G. Schultz, *Science*, 2001, **292**, 498–500; (*b*) C. C. Liu and P. G. Schultz, *Annu. Rev. Biochem.*, 2010, **79**, 413–444.
4 (*a*) P. R. Chen, D. Groff, J. T. Guo, W. J. Ou, S. Cellitti, B. H. Geierstanger and P. G. Schultz, *Angew. Chem., Int. Ed.*, 2009, **48**, 4052–4055; (*b*) S. M. Hancock, R. Uprety, A. Deiters and J. W. Chin, *J. Am. Chem. Soc.*, 2010, **132**, 14819–14824; (*c*) T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto and S. Yokoyama, *Biochem. Biophys. Res. Commun.*, 2008, **371**, 818–822; (*d*) H. Neumann, S. Y. Peak-Chew and J. W. Chin, *Nat. Chem. Biol.*, 2008, **4**, 232–234; (*e*) Y. Huang, W. Wan, W. K. Russell, P. J. Pai, Z. Y. Wang, D. H. Russell and W. S. Liu, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 878–880.
5 A. Gautier, D. P. Nguyen, H. Lusic, W. A. An, A. Deiters and J. W. Chin, *J. Am. Chem. Soc.*, 2010, **132**, 4086–4088.
6 T. Fekner, X. Li and M. K. Chan, *Eur. J. Org. Chem.*, 2010, 4171–4179.
7 E. Kaya, K. Gutsmiedl, M. Vrabel, M. Muller, P. Thumbs and T. Carell, *ChemBioChem*, 2009, **10**, 2858–2861.
8 (*a*) J. M. Kavran, S. Gundliapalli, P. O'Donoghue, M. Englert, D. Soell and T. A. Steitz, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 11268–11273; (*b*) C. R. Polycarpo, S. Herring, A. Berube, J. L. Wood, D. Soll and A. Ambrogelly, *FEBS Lett.*, 2006, **580**, 6695–6700.
9 S. E. Schaus, J. F. Larrow and E. N. Jacobsen, *J. Org. Chem.*, 1997, **62**, 4197–4199.
10 K. S. Gajiwala and S. K. Burley, *J. Mol. Biol.*, 2000, **295**, 605–612.
11 (*a*) V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2002, **41**, 2596–2597; (*b*) P. V. Chang, X. Chen, C. Smyrniotis, A. Xenakis, T. S. Hu, C. R. Bertozzi and P. Wu, *Angew. Chem., Int. Ed.*, 2009, **48**, 4030–4033.
12 D. S. del Amo, W. Wang, H. Jiang, C. Besanceney, A. C. Yan, M. Levy, Y. Liu, F. L. Marlow and P. Wu, *J. Am. Chem. Soc.*, 2010, **132**, 16893–16899.