Bioorganic & Medicinal Chemistry Letters 20 (2010) 878–880

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Genetic incorporation of an aliphatic keto-containing amino acid into proteins for their site-specific modifications

Ying Huang [a], Wei Wan [b], William K. Russell [a], Pei-Jing Pai [a], Zhiyong Wang [a], David H. Russell [a], Wenshe Liu [a,*]

[a] Department of Chemistry, Texas A&M University, College Station, TX 77843, United States
[b] Department of Biology, Texas A&M University, College Station, TX 77843, United States

ARTICLE INFO

Article history:
Received 17 November 2009
Accepted 18 December 2009
Available online 24 December 2009

Keywords:
Keto
2-Amino-8-oxononanoic acid
Genetic code expansion
Amber suppression
Ne-acetyl-lysine

ABSTRACT

2-Amino-8-oxononanoic acid has been genetically incorporated into proteins in *Escherichia coli* by the use of an evolved pyrrolysyl-tRNA synthetase/pylT pair. The direct usage of the exclusive reactivity of the keto group in this amino acid with hydrazide- and alkoxyamine-bearing compounds to site-specifically label proteins under a mild condition close to physiological pH exhibited a very high efficiency.

Published by Elsevier Ltd.

The keto group is one of the most versatile functional groups in organic chemistry and participates in a large number of reactions ranging from addition reactions to aldol condensations. A unique feature of the keto group that has been extensively explored and applied to biopolymer modifications is its reactivity with hydrazide- and alkoxyamine-bearing compounds under physiological conditions (Fig. 1). Although ubiquitous in nature, the keto group is absent from proteins unless it is posttranslationally installed or occurs in protein cofactors.[1–3]

A variety of approaches have been developed to achieve the installation of the keto group into proteins posttranslationally. An N-terminal serine in a purified protein can be specifically oxidized chemically by periodates to form an aldehyde.[4] The keto group has also been installed by enzymatic modifications of specific tags that are genetically linked to proteins and utilized to site-specifically label proteins on cell surface.[5–9] Although chemical and enzymatic keto installations provide great tools for protein labeling, they have one common limitation. The site for keto installation is generally limited to the two termini of proteins and installation of the keto at internal sites is hard to achieve. Keto incorporation into proteins can also be achieved cotranslationally. This approach relies on the read-through of an in-frame amber (UAG) stop codon in mRNA by an amber suppressor tRNA ($tRNA_{CUA}$) acylated specifically with a keto-containing amino acid. An essential advantage of this approach is that the keto group can be installed at any desired site of a protein regardless of the protein size. Using a chemically acylated-$tRNA_{CUA}$, Schultz et al. first demonstrated the incorporation of an aliphatic keto-containing amino acid, *O*-(2-oxopropyl)-L-tyrosine, into T4 lysozyme in vitro for its site-specific labeling with fluorescent hydrazide dyes under physiological pH.[10] However, the synthesis of chemically acylated $tRNA_{CUA}$ is time-consuming and the expressed protein amount is limited. To reach protein amount that is more accessible for most biochemical studies, the genetic incorporation of a phenolic keto-containing amino acid, *p*-acetyl-L-phenylalanine (*p*-Ac-Phe), into proteins in vivo was later developed by using either an evolved orthogonal *Methanococcus jannaschii* tyrosyl-tRNA synthetase/$tRNA_{CUA}$ pair in *Escherichia coli* or an evolved orthogonal *E. coli* tyrosyl-tRNA synthetase/$tRNA_{CUA}$ pair in *Saccharomyces cerevisiae* and feeding cells with *p*-Ac-Phe.[11,12] An elegant application of the genetic incorporation of *p*-Ac-Phe was recently demon-

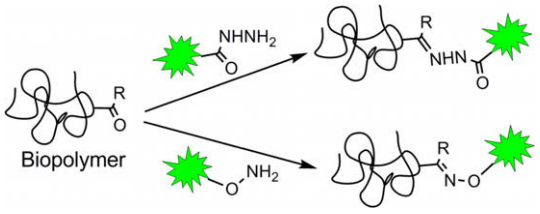


Figure 1. Reactions of the keto group in biopolymers with hydrazide- and alkoxyamine-bearing compounds.

* Corresponding author. Tel.: +1 979 845 1746; fax: +1 979 845 4719.
E-mail address: wliu@mail.chem.tamu.edu (W. Liu).

0960-894X/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.bmcl.2009.12.077

strated.[13] Using orthogonal reactions with a genetically incorporated *p*-Ac-Phe and a cysteine residue, a FRET pair was introduced in T4 lysozyme for single-molecule FRET analysis of protein folding. However, the efficient labeling of *p*-Ac-Phe in T4 lysozyme with alkoxyamine dyes could only be achieved at pH 4 with overnight incubation and the labeling with hydrazide dyes exhibited very low efficiency at all pH tested.[13] The low reactivity of the keto group in *p*-Ac-Phe is possibly due to its conjugation with an aromatic phenyl ring. The conjugative electron-donating effect of the phenyl group may reduce the electrophilicity of the carbonyl carbon of the keto and decrease its reactivity. For this reason, approaches allowing the genetic incorporation of aliphatic keto-containing amino acids into proteins still need to be developed. Herein, we report the genetic incorporation of one of these amino acids, 2-amino-8-oxononanoic acid (KetoK in Fig. 2), into proteins in *E. coli* and its application in efficient labeling of proteins with different probes with a site-specific manner under a mild condition close to the physiological pH.

The genetic incorporation of KetoK takes advantage of an N^{ϵ}-acetyl-L-lysyl-tRNA synthetase (AcKRS)/$tRNA_{CUA}$ (pylT) pair that was originally generated by Chin et al. for the incorporation of N^{ϵ}-acetyl-lysine (AcK) into proteins in *E. coli*.[14] AcKRS was evolved from *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase[15] and contains six mutations D76G/L266V/L270I/Y271F/L273A/C313F. Since the N^{ϵ}-amide NH group of AcK is not involved in any apparent hydrogen bonding interactions with vicinal residues of AcKRS when bound to the active site,[16] we reason that changing the amino acid substrate from AcK to KetoK should not significantly affect the substrate's binding potential to AcKRS and we may be able to use the AcKRS/pylT pair to genetically incorporate KetoK into proteins in *E. coli*.

Starting from 6-bromohexanoic acid, recemic KetoK was conveniently synthesized in gram quantities over a 6-step sequence in 50% overall yield (Fig. 2A). The genetic encoding of KetoK by amber codon in *E. coli* was subsequently tested using the AcKRS/pylT pair. A plasmid pAcKRS-pylT-GFP1Amber bearing the genes encoding AcKRS, pylT and GFP_{UV} with an amber mutation at position 149 and a 6×His tag at the C-terminus was constructed from the pETduet-1 vector (Stratagene Inc.). In this plasmid, AcKRS and GFP_{UV} are both under the control of IPTG-inducible T7 promoters and pylT is flanked by the lpp promoter and the rrnC terminator. The transformation of this plasmid into BL21 cells and subsequent growth in LB medium supplemented with 2 mM KetoK and 500 μM IPTG afforded full-length GFP_{UV} incorporated with KetoK (GFP-KetoK) with a yield of ~0.5 mg/L, which was comparable to the expression yield of full-length GFP_{UV} incorporated with AcK (GFP-AcK) (~0.8 mg/L) when the supplemented amino acid was changed to 5 mM AcK (Fig. 2B). The deconvoluted electrospray ionization mass spectrometry (ESI-MS) spectrum of GFP-KetoK revealed two intense peaks

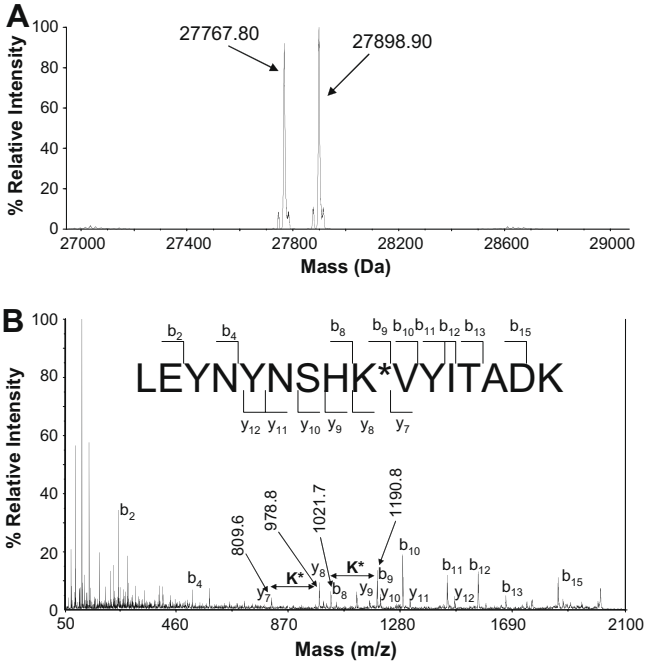
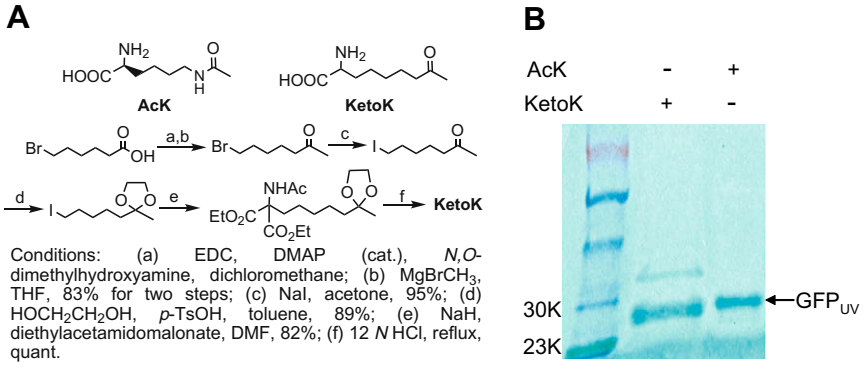


Figure 3. (A) The deconvoluted ESI-MS spectrum of GFP-KetoK and (B) tandem MS spectrum of LEYNYNSHK*VYITADK from GFP-KetoK.

(Fig. 3A) corresponding to the full-length GFP-KetoK with and without N-terminal methionine, respectively (detected: 27,898.9 Da, 27,767.8 Da; calculated: 277,896 Da, 27,765 Da). The site-specific incorporation of KetoK at position 149 was further validated by the tandem mass spectral analysis of the tryptic KetoK-containing fragment of LEYNYNSHK*VYITADK (K* denotes KetoK). The K*-containing ions (y_8 to y_{12} and b_9 to b_{15}) all had the expected mass (Fig. 3B).

To test the efficiency of using the genetically incorporated KetoK to specifically label proteins with fluorescent dyes under a mild condition, the purified GFP-KetoK and GFP-AcK were both reacted with 1 equiv Texas Red hydrazide in PBS buffer at 37 °C, pH 6.3 overnight and then analyzed on a SDS-PAGE gel. The gel was visualized by both silver staining and fluorescent imaging (Fig. 4A&B). Whereas both GFP-KetoK and GFP-AcK showed a strong protein band after silver staining, only labeled GFP-KetoK had an intense fluorescent band. This indicates that labeling only occurred on GFP-KetoK and the reaction between the keto group and a hydrazide is specific.



Conditions: (a) EDC, DMAP (cat.), *N,O*-dimethylhydroxyamine, dichloromethane; (b) $MgBrCH_3$, THF, 83% for two steps; (c) NaI, acetone, 95%; (d) $HOCH_2CH_2OH$, *p*-TsOH, toluene, 89%; (e) NaH, diethylacetamidomalonate, DMF, 82%; (f) 12 *N* HCl, reflux, quant.

Figure 2. (A) The synthetic route of KetoK and (B) the expression of GFP_{UV} with amber mutation at 149 in cells containing the AcKRS/pylT pair and in the presence of KetoK or AcK.

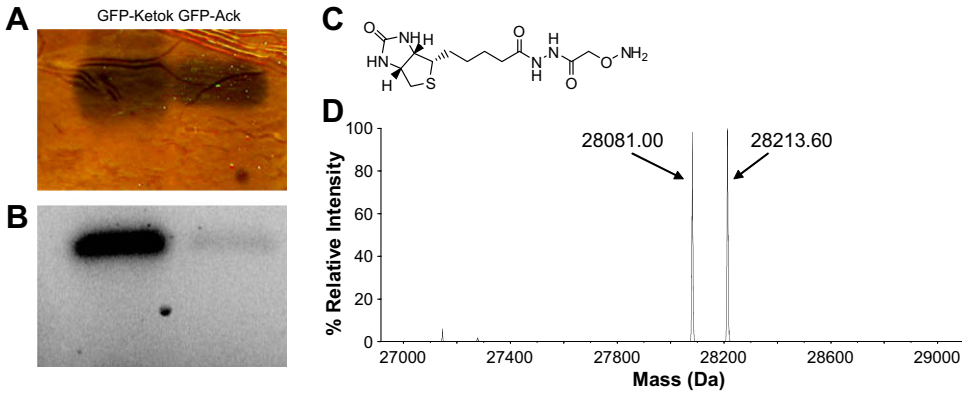


Figure 4. (A) Silver staining and (B) fluorescent imaging of GFP-KetoK and GFP-AcK after their reactions with Texas Red hydrazide; (C) the structure of biotin alkoxyamine; (D) The deconvoluted ESI-MS spectrum of GFP-KetoK after its reaction with biotin alkoxyamine.

To further characterize the labeling efficiency, the purified GFP-KetoK was reacted with 5 equiv of biotin alkoxyamine in PBS buffer at 37 °C, pH 6.3 overnight and the modified protein was analyzed by ESI-MS (Fig. 4D). Two intense peaks (28,213.6 Da, 28,081.0 Da) were shown in the deconvoluted ESI-MS spectrum and they matched the biotin labeled full-length GFP-KetoK with and without N-terminal methionine (calculated mass: 28,209 Da, 28,078 Da), respectively. Given the fact that no original full-length GFP-KetoK was detected, the labeling reaction was highly efficient. As a control, the same labeling reaction was carried out on wild type GFP_{UV}. The deconvoluted ESI-MS spectra of wild type GFP_{UV} before and after the labeling reaction showed no difference (Supplementary Fig. 1), demonstrating the reaction specificity between KetoK and alkoxyamine probes.

In summary, we have achieved the genetic incorporation of KetoK into proteins in *E. coli* and applied it to label proteins with different probes. The labeling features high specificity and high efficiency under a mild reaction condition. One potential application of our method is to site-specifically introduce various biochemical and biophysical probes into proteins, including biotin tag, fluorescent labels, NMR probes, EPR probes, etc. Unlike the previously developed phenolic keto-containing amino acids, KetoK is more reactive toward the conjugation with hydrazide- and alkoxyamine-bearing compounds. This high reactivity may allow the labeling of proteins incorporated with KetoK on live cell surface. Since conjugation of therapeutic proteins with polyethylene glycols has been successfully used to prolong the half life of these protein drugs in human patients,[17] the incorporation of KetoK into therapeutic proteins may provide a more straight forward method to modify these proteins with polyethylene glycols that contain either hydrazide or alkoxyamine group. Moreover, KetoK is an unhydrolysable analogue of AcK that represents one of the most important posttranslationally modified amino acids in eukaryotic cells. Since protein acetylation is a reversible process catalyzed by enzymes like histone acetyltransferases and histone deacetyltransferases, direct incorporation of KetoK into proteins in vivo at sites with naturally occurring lysine acetylation will permanently install an AcK mimic and may provide an efficient way to decipher the regulation roles of acetylation in histones, p53 and other transcription regulatory proteins.[18,19]

Acknowledgement

This work was supported by Welch Grant A-1715 and Texas A&M Chemistry Department. We thank Prof. Jiong Yang and his group for their help in the KetoK synthesis.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2009.12.077.

References and notes

1. van Poelje, P. D.; Snell, E. E. *Annu. Rev. Biochem.* **1990**, *59*, 29–59.
2. Eliot, A. C.; Kirsch, J. F. *Annu. Rev. Biochem.* **2004**, *73*, 383–415.
3. Schmidt, B.; Selmer, T.; Ingendoh, A.; von Figura, K. *Cell* **1995**, *82*, 271–278.
4. Geoghegan, K. F.; Emery, M. J.; Martin, W. H.; McColl, A. S.; Daumy, G. O. *Bioconjugate Chem.* **1993**, *4*, 537–544.
5. Mahal, L. K.; Yarema, K. J.; Bertozzi, C. R. *Science* **1997**, *276*, 1125–1128.
6. Hang, H. C.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2001**, *123*, 1242–1243.
7. Chen, I.; Howarth, M.; Lin, W.; Ting, A. Y. *Nat. Methods* **2005**, *2*, 99–104.
8. Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R. *Nat. Chem. Biol.* **2007**, *3*, 321–322.
9. Wu, P.; Shui, W.; Carlson, B. L.; Hu, N.; Rabuka, D.; Lee, J.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 3000–3005.
10. Cornish, V. W.; Mendel, D.; Schultz, P. G. *J. Am. Chem. Soc.* **1996**, *118*, 8150–8151.
11. Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 56–61.
12. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. *Science* **2003**, *301*, 964–967.
13. Brustad, E. M.; Lemke, E. A.; Schultz, P. G.; Deniz, A. A. *J. Am. Chem. Soc.* **2008**, *130*, 17664–17665.
14. Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Chem. Biol.* **2008**, *4*, 232–234.
15. Srinivasan, G.; James, C. M.; Krzycki, J. A. *Science* **2002**, *296*, 1459–1462.
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. Bailon, P.; Won, C. Y. *Expert Opin. Drug Deliv.* **2009**, *6*, 1–16.
18. Yang, X. J.; Seto, E. *Mol. Cell.* **2008**, *31*, 449–461.
19. Kruse, J. P.; Gu, W. *Cell* **2009**, *137*, 609–622.