

ChemComm

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

A Genetically Encoded Aldehyde for Rapid Protein Labelling

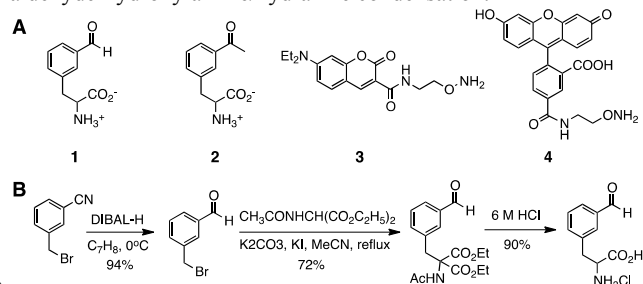
Alfred Tuley,^a Yan-Jiun Lee,^a Bo Wu,^a Zhiyong Wang,^b and Wenshe R. Liu^{*a}

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX

DOI: 10.1039/b000000x

5 Using a mutant pyrrolysyl-tRNA synthetase-tRNA^{Pyl}_{CUA} pair, 3-formyl-phenylalanine is genetically incorporated into proteins at amber mutation sites in *Escherichia coli*. This non-canonical amino acid readily reacts with hydroxylamine dyes, leading to rapid and site-selective protein labelling.

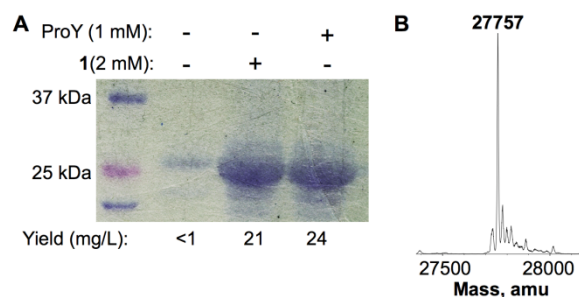
10 Although the aldehyde is one of the most versatile chemical functionalities and participates in a variety of useful chemical reactions,¹ it does not occur in the 20 canonical amino acids. To harness the unique reactivity of aldehydes with hydrazine and hydroxylamine dyes for protein labelling, aldehyde tags have been introduced into proteins by oxidizing an N-terminal serine or via enzymatic modifications of preinstalled substrate peptides.² The N-terminal oxidation approach can only be applied *in vitro* and the enzymatic approach is generally confined for the aldehyde installation at the two protein termini. To genetically encode the keto functionality at any chosen sites in a protein, several ketone-containing non-canonical amino acids (NCAAs) have been designed and incorporated into proteins in living cells using the amber suppression approach.³ However, a method for the genetic incorporation of an aldehyde-containing NCAA has not been developed. We are interested in a genetically encoded, aldehyde-containing NCAA due to its superior reactivity with respect to ketone-containing NCAAs, and the observation that aniline can selectively enhance the reaction kinetics of the aldehyde-hydroxylamine/hydrazine condensation.⁴



30 **Figure 1.** (A) Structures of 1-4. (B) The synthetic route of 1.

In our previous studies, we revealed that a mutant pyrrolysyl-tRNA synthetase with mutations as N346A/C348A (PylRS(N346A/C348A)) recognizes a number of phenylalanine derivatives including *meta*-acetyl-phenylalanine (**2** in **Figure 1A**) and, in coordination with tRNA^{Pyl}_{CUA}, directs their genetic incorporation at amber mutation sites in *Escherichia coli* and mammalian cells.⁵ Given the size similarity between **2** and *meta*-formyl-phenylalanine (**1**), and the ability of this mutant enzyme in accepting diverse *meta*-substituted phenylalanine substrates, we anticipated that this same mutant enzyme would also accept **1** as its substrate. To test this prospect, **1** was readily prepared in a high yield by a three-step synthesis (**Figure 1B**). Its acceptance

as a substrate of PylRS(N346A/C348A) were then examined. An *E. coli* BL21(DE3) cell that harboured two plasmids, pEVOL-pylIT-PylRS(N346A/C348A) and pET-pylIT-sfGFP2TAG, was employed for the investigation. Plasmid pEVOL-pylIT-PylRS(N346A/C348A) contains genes coding both PylRS(N346A/C348A) and tRNA^{Pyl}_{CUA}. Plasmid pET-pylIT-sfGFP2TAG carries a tRNA^{Pyl}_{CUA} coding gene and a sequence-optimized superfolder green fluorescent protein (sfGFP) gene with an amber mutation at its S2 position. Growing cells in M9 minimal medium supplemented with 1% glycerol, 1 mM IPTG and 0.2% arabinose, but without **1**, afforded a minimal expression level of full-length sfGFP (<1 mg/L). Providing **1** at 2 mM into the medium promoted full-length sfGFP expression (**Figure 2A**). The expression level is comparable to that for *para*-propargyloxy-phenylalanine (ProY), and the electrospray ionization mass spectrometry analysis of the purified **1**-containing sfGFP displayed a molecular weight that agreed well with its theoretical value (**Figure 1B**). Moreover, compound **1** is not toxic to *E. coli* cells, as phenotypes and cell pellets are comparable to control experiments.



65 **Figure 2.** (A) The site-specific incorporation of **1** into sfGFP at its S2 position. Cells transformed with pEVOL-PylRS(N346A/C348A) and pET-pylIT-sfGFP2TAG were grown in conditions with or without a NCAA supplement. (B) The deconvoluted ESI-MS spectrum of the purified **1**-containing sfGFP. Its theoretical value is 27,756 Da.

70 Dawson *et al.* previously showed that aniline could serve as an excellent catalyst for speeding the imine formation of aldehydes with hydroxylamines and hydrazines. To demonstrate this catalytically rapid reaction on proteins, sfGFP with **1** incorporated at its N149 position (sfGFP-**1**) was expressed and then applied to a Förster resonance energy transfer (FRET)-based kinetic investigation of its reaction with a coumarin dye **3**,^{3d, 6} in the presence of 100 mM aniline and at pH 7 using a fluorescence spectrophotometer (**Figure 1A**). **3** has an excitation wavelength at 417 nm and emits at 476 nm. Its emitted light falls in the range of excitation lights of sfGFP that emits at 510 nm.⁷ The reaction of sfGFP-**1** with **3** will place **3** in a close locality of the sfGFP

fluorophore for FRET. All kinetic analyses were carried out in pseudo-first-order conditions in which **3** was at least in 5-fold excess in comparison to sfGFP-1. Data collected under these conditions were well fitted to a single exponential increase equation $F = F_1 - F_2 \times e^{(-k' \times t)}$, where F was the detected fluorescent signal at a given time, F_1 was the final fluorescence, $F_1 - F_2$ was the background fluorescent signal, and k' was the apparent pseudo first-order rate constant (Figure 3A). The determined k' values were plotted against the concentrations of **3** and fitted to the equation $k' = k \times [3] + C$, where k was a second-order rate constant for the reaction of sfGFP-1 with **3** (inset of Figure 3A). The calculated second-order rate constant was $182 \pm 12 \text{ M}^{-1} \text{ s}^{-1}$. This determined rate constant is comparable to what Dawson *et al.* reported for aniline-catalyzed bioconjugation, and is among one of the most rapid bioorthogonal reactions.^{4, 8} We also carried out the kinetic analysis in the absence of aniline. However, no obvious signal changes were observed when $0.1 \mu\text{M}$ sfGFP-1 reacted with $4 \mu\text{M}$ **3** in a period of 2 h, indicating very slow kinetics of the sfGFP-1 reaction with **3** in the absence of aniline and at pH 7. We also tried much higher concentrations of **3** to speed up the reaction rate. However, the background fluorescence from **3** in these conditions was too high to collect reliable data. As a comparison, a sfGFP variant with **2** incorporated at its N149 position (sfGFP-2) was also expressed and used for kinetic investigation of its reaction with **3** in the presence of 100 mM aniline and at pH 7. As shown in Figure 3A, at the tested conditions of $0.05 \mu\text{M}$ sfGFP-2 and $10 \mu\text{M}$ **3**, no obvious signal increase due to the sfGFP-2 reaction with **3** was observed in a range of 30 min. The fluorescence decrease was caused by the bleaching effect of the excitation light.

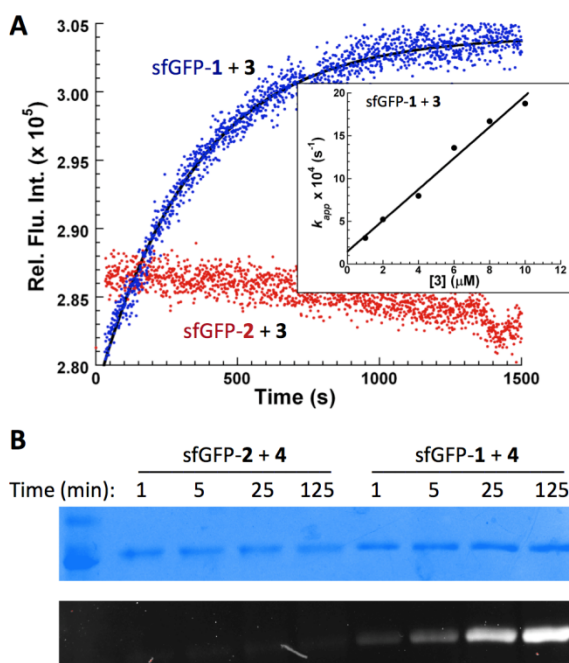


Figure 3. (A) Fluorescent intensities as a function of time for reactions of $0.05 \mu\text{M}$ sfGFP-1 (coloured in blue) and $0.05 \mu\text{M}$ sfGFP-2 (coloured in red) with $10 \mu\text{M}$ **3** in the presence of 100 mM aniline and at pH 7. Presented in the inset is the linear dependence of the determined pseudo first-order rate constants for the reaction of $0.1 \mu\text{M}$ sfGFP-1 with **3** on the concentrations of **3**. (B) Gel imaging analysis of labelling of sfGFP-1 and sfGFP-2 with **4** with different incubation times. For both sfGFP-1 and sfGFP-2, labelling was carried out in conditions with $8 \mu\text{M}$ protein, $10 \mu\text{M}$ **4**, 100 mM aniline, and pH 7. Reactions were quenched with the addition of 2 mM benzaldehyde at indicated times and analyzed by

denaturing SDS-PAGE. The top panel shows proteins with Coomassie blue staining and the bottom panel presents the fluorescent imaging of the same gel before Coomassie blue staining.

The different labelling kinetics of sfGFP-1 and sfGFP-2 were also examined by fluorescent imaging of SDS-PAGE analyzed **4**-labelled products. Both sfGFP-1 and sfGFP-2 at $8 \mu\text{M}$ were incubated with $10 \mu\text{M}$ **4** in the presence of 100 mM aniline and at pH 7. Reactions were quenched at different times and analyzed by denaturing SDS-PAGE to remove the intrinsic fluorescence of sfGFP. Fluorescent imaging of two proteins at different labelling times clearly showed that sfGFP-1 was efficiently labelled at 25 min but sfGFP-2 was minimally labelled even at 2 h. Labelling could also be achieved in the absence of aniline, though reaction times were significantly longer (Figure S2). Once again, sfGFP-1 showed superior labelling with respect to sfGFP-2. This result shows that, in systems where aniline may be toxic, labelling can still be achieved without it or at much lower concentrations.

With the demonstration of fast labelling kinetics of **1** with hydroxylamine dyes, we then proceeded to test the conditions for labelling proteins that bear site-specifically incorporated **1** in living cells. Plasmid pEVOL-pyIT-PylRS(N346A/C348A) and a previously constructed plasmid, pETDuet-OmpXTAG, were used to transform *E. coli* BL21 cells. The latter plasmid contained a gene coding for an *E. coli* outer membrane protein OmpX,^{8b} with an AAAXAA (A denotes alanine and X denotes an amber mutation) insertion between two extracellular residues, 53 and 54. The transformed cells were grown in M9 minimal medium supplemented with 1% glycerol, 1 mM IPTG, 0.2% arabinose, and 2 mM **1** or **2** to express OmpX with **1** or **2** incorporated (OmpX-1 or OmpX-2). Cells were then harvested and labelled with **4** in the presence of 100 mM aniline and at pH 7 for 1 h. After labelling, cells were washed with PBS buffer for six times to remove the residual dye and then imaged by fluorescent microscopy. Cells grown in the absence of **1** or **2** were used as control. As shown in Figure 4, **4** specifically labelled cells expressing OmpX-1 and OmpX-2 were observed, but not for cells in the control experiment. Although cells expressing OmpX-2 were fluorescently labelled, their intensities are weaker than those of cells expressing OmpX-1, indicating a slower labelling kinetics of OmpX-2. However, we have found that this concentration of aniline is toxic to *E. coli*, which underscores the need for developing less toxic catalysts for this reaction.

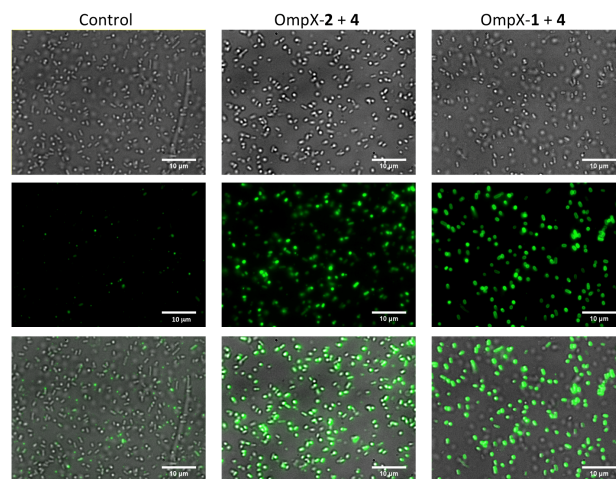


Figure 4. Selective labelling of *E. coli* cells expressing OmpX-1 and OmpX-2. The labelling was carried out in the presence of 100 mM aniline and at pH 7 for 1 h. The top panel shows bright field imaging of *E. coli* cells, the middle panel shows green fluorescent imaging of same cells, and the bottom panel shows their composite images (scale bars are 10 μm).

In summary, we reported the genetic incorporation of a readily synthesized aldehyde-containing NCA and demonstrated its fast labelling kinetics with hydroxylamine dyes in the presence of the aniline catalyst. This rapid labelling approach was also successfully applied to label a membrane protein on the *E. coli* extracellular surface. Although genetically encoded ketone-containing NCAs were reported previously, labelling of proteins with these NCAs typically suffers low labelling efficiencies, attributing to slow labelling kinetics of ketones. This work resolves this obstacle. Given that a large variety of hydroxylamine and hydrazine dyes are commercially available, as well as the existence of numerous aldehyde-based bioconjugation strategies,⁹ we anticipate this approach will be quickly adopted by others for studies such as protein folding/dynamics, protein-ligand interactions, high throughput drug discovery, etc.

We thank the financial support from the National Institute of Health (grant 1R01CA161158), the National Science Foundation (grant CHE-1148684), and the Welch Foundation (grant A-1715). We also thank Dr. Yohannes H. Rezenom from the Laboratory for Biological Mass Spectrometry at Texas A&M University for characterizing our proteins with electrospray ionization mass spectrometry.

Notes and references

^a Department of Chemistry, Texas A&M University, College Station, TX 77843; wliu@chem.tamu.edu.

^b Department of Chemistry, Troy University, Troy, Alabama, 36082. Electronic Supplementary Information (ESI) available: Synthesis, protein expression, protein labelling, and mass spectrometry analysis. See DOI: 10.1039/c000000x/

- (a) J. M. McFarland and M. B. Francis, *J Am Chem Soc*, 2005, **127**, 13490; (b) J. M. McFarland, N. S. Joshi and M. B. Francis, *J Am Chem Soc*, 2008, **130**, 7639; (c) M. J. Han, D. C. Xiong and X. S. Ye, *Chem Commun (Camb)*, 2012, **48**, 11079; (d) P. Agarwal, R. Kudirka, A. E. Albers, R. M. Barfield, G. W. de Hart, P. M. Drake, L. C. Jones and D. Rabuka, *Bioconjug Chem*, 2013, **24**, 846; (e) P. Agarwal, J. van der Weijden, E. M. Sletten, D. Rabuka and C. R. Bertozzi, *Proc Natl Acad Sci U S A*, 2013, **110**, 46; (f) L. S. Witus, C. Netirojjanakul, K. S. Palla, E. M. Muehl, C. H. Weng, A. T. Iavarone and M. B. Francis, *J Am Chem Soc*, 2013, **135**, 17223.
- (a) I. I. Geschwind and C. Hao Li, *Biochimica et Biophysica Acta*, 1954, **15**, 442; (b) I. S. Carrico, B. L. Carlson and C. R. Bertozzi, *Nat Chem Biol*, 2007, **3**, 321; (c) J. S. Rush and C. R. Bertozzi, *J Am Chem Soc*, 2008, **130**, 12240; (d) P. Wu, W. Shui, B. L. Carlson, N. Hu, D. Rabuka, J. Lee and C. R. Bertozzi, *Proc Natl Acad Sci U S A*, 2009, **106**, 3000; (e) J. E. Hudak, H. H. Yu and C. R. Bertozzi, *J Am Chem Soc*, 2011, **133**, 16127; (f) O. El-Mahdi and O. Melnyk, *Bioconjug Chem*, 2013, **24**, 735.
- (a) L. Wang, Z. Zhang, A. Brock and P. G. Schultz, *Proc Natl Acad Sci U S A*, 2003, **100**, 56; (b) Z. Zhang, B. A. Smith, L. Wang, A. Brock, C. Cho and P. G. Schultz, *Biochemistry*, 2003, **42**, 6735; (c) Y. Huang, W. Wan, W. K. Russell, P. J. Pai, Z. Wang, D. H. Russell and W. Liu, *Bioorg Med Chem Lett*, 2010, **20**, 878; (d) B. Wu, Z. Wang, Y. Huang and W. R. Liu, *ChemBiochem*, 2012, **13**, 1405; (e) H. Zeng, J. Xie and P. G. Schultz, *Bioorg Med Chem Lett*, 2006, **16**, 5356.
- (a) A. Dirksen, T. M. Hackeng and P. E. Dawson, *Angew Chem Int Ed Engl*, 2006, **45**, 7581; (b) A. Dirksen and P. E. Dawson, *Bioconjug Chem*, 2008, **19**, 2543; (c) Y. Zeng, T. N. Ramya, A. Dirksen, P. E. Dawson and J. C. Paulson, *Nat Methods*, 2009, **6**, 207; (d) M. Wendeler, L. Grinberg, X. Wang, P. E. Dawson and M. Baca, *Bioconjug Chem*, 2014, **25**, 93.
- (a) Y. S. Wang, W. K. Russell, Z. Wang, W. Wan, L. E. Dodd, P. J. Pai, D. H. Russell and W. R. Liu, *Mol Biosyst*, 2011, **7**, 714; (b) Y. S. Wang, X. Fang, A. L. Wallace, B. Wu and W. R. Liu, *J Am Chem Soc*, 2012, **134**, 2950; (c) Y. S. Wang, X. Fang, H. Y. Chen, B. Wu, Z. U. Wang, C. Hilty and W. R. Liu, *ACS Chem Biol*, 2013, **8**, 405; (d) J. M. Tharp, Y. S. Wang, Y. J. Lee, Y. Yang and W. R. Liu, *ACS Chem Biol*, 2014, DOI: 10.1021/cb400917a; (e) A. Tuley, Y. S. Wang, X. Fang, Y. Kurra, Y. H. Rezenom and W. R. Liu, *Chem Commun (Camb)*, 2014, **50**, 2673.
- S. C. Reddington, E. M. Tippmann and D. D. Jones, *Chem Commun (Camb)*, 2012, **48**, 8419.
- J. D. Pedelacq, S. Cabantous, T. Tran, T. C. Terwilliger and G. S. Waldo, *Nat Biotechnol*, 2006, **24**, 79.
- (a) X. S. Wang, Y. J. Lee and W. R. Liu, *Chem Commun (Camb)*, 2014, **50**, 3176; (b) Y. J. Lee, B. Wu, J. E. Raymond, Y. Zeng, X. Fang, K. L. Wooley and W. R. Liu, *ACS Chem Biol*, 2013, **8**, 1664; (c) K. Lang, L. Davis, S. Wallace, M. Mahesh, D. J. Cox, M. L. Blackman, J. M. Fox and J. W. Chin, *J Am Chem Soc*, 2012, **134**, 10317; (d) A. Borrmann, S. Milles, T. Plass, J. Dommerholt, J. M. Verkade, M. Wiessler, C. Schultz, J. C. van Hest, F. L. van Delft and E. A. Lemke, *ChemBiochem*, 2012, **13**, 2094; (e) T. Plass, S. Milles, C. Koehler, J. Szymanski, R. Mueller, M. Wiessler, C. Schultz and E. A. Lemke, *Angew Chem Int Ed Engl*, 2012, **51**, 4166; (f) M. L. Blackman, M. Royzen and J. M. Fox, *J Am Chem Soc*, 2008, **130**, 13518.
- (a) G. R. Lorello, M. C. Legault, B. Rakic, K. Bisgaard and J. P. Pezacki, *Bioorg Chem*, 2008, **36**, 105; (b) J. Y. Byeon, F. T. Limpoco and R. C. Bailey, *Langmuir*, 2010, **26**, 15430; (c) D. A. Ossipov, X. Yang, O. Varghese, S. Kootala and J. Hilborn, *Chem Commun (Camb)*, 2010, **46**, 8368; (d) V. Raindlova, R. Pohl, M. Sanda and M. Hocek, *Angew Chem Int Ed Engl*, 2010, **49**, 1064; (e) B. H. Rotstein, V. Rai, R. Hili and A. K. Yudin, *Nat Protoc*, 2010, **5**, 1813; (f) G. Iyer, F. Pinaud, J. Xu, Y. Ebenstein, J. Li, J. Chang, M. Dahan and S. Weiss, *Bioconjug Chem*, 2011, **22**, 1006; (g) J. D. Cohen, P. Zou and A. Y. Ting, *ChemBiochem*, 2012, **13**, 888; (h) P. Crisalli, A. R. Hernandez and E. T. Kool, *Bioconjug Chem*, 2012, **23**, 1969; (i) V. Raindlova, R. Pohl and M. Hocek, *Chemistry*, 2012, **18**, 4080; (j) P. K. Dhal, S. C. Polomoscanik, D. A. Gianolio, P. G. Starremans, M. Busch, K. Alving, B. Chen and R. J. Miller, *Bioconjug Chem*, 2013, **24**, 865; (k) D. Ossipov, S. Kootala, Z. Yi, X. Yang and J. Hilborn, *Macromolecules*, 2013, **46**, 4105; (l) K. Tanaka, Y. Nakamoto, E. R. Siwu, A. R. Pradipta, K. Morimoto, T. Fujiwara, S. Yoshida, T. Hosoya, Y. Tamura, G. Hirai, M. Sodeoka and K. Fukase, *Org Biomol Chem*, 2013, **11**, 7326; (m) S. Wang, O. P. Oommen, H. Yan and O. P. Varghese, *Biomacromolecules*, 2013, **14**, 2427.