J|A|C|S COMMUNICATIONS

Published on Web 06/10/2009

Genetic Encoding and Labeling of Aliphatic Azides and Alkynes in Recombinant Proteins *via* a Pyrrolysyl-tRNA Synthetase/$tRNA_{CUA}$ Pair and Click Chemistry

Duy P. Nguyen,[†] Hrvoje Lusic,[‡] Heinz Neumann,[†] Prashant B. Kapadnis,[†] Alexander Deiters,*[,‡] and Jason W. Chin*[,†]

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 0QH, U.K., and North Carolina State University, Department of Chemistry, Raleigh, North Carolina 27695

Received January 23, 2009; E-mail: alex_deiters@ncsu.edu; chin@mrc-lmb.cam.ac.uk

The site-specific and homogeneous modification of recombinant proteins under physiological conditions is an important challenge. Cysteines and other amino acid residues in proteins can be specifically labeled by several methods,[1] but site-specificity, as opposed to residue specificity,[2,3] is difficult to achieve.

Several phenylalanine derivatives can be site-specifically introduced into recombinant proteins in response to an amber codon inserted into the corresponding gene using an evolved tyrosyl-tRNA synthetase/$tRNA_{CUA}$ pair that is orthogonal in *E. coli*.[4] Phenylalanine derivatives bearing alkynyl, azido, and keto groups, that are bio-orthogonal in their chemical reactivity have been incorporated.[5−8] However the introduction of aromatic amino acids at the majority of sites in proteins where aliphatic amino acids are naturally found may cause misfolding or loss of protein function; therefore there is a pressing need for methods to site-specifically incorporate aliphatic amino acids that contain bio-orthogonal chemical handles for use in protein labeling.

The *Methanosarcina barkeri* MS pyrrolysyl tRNA synthetase/$tRNA_{CUA}$ (*Mb*PylRS/*Mb*$tRNA_{CUA}$) pair is a new orthogonal pair in *E. coli*.[9,10] We demonstrated that the *Mb*PylRS/$tRNA_{CUA}$ pair can be evolved to direct the efficient incorporation of unnatural amino acids into genetically determined sites in recombinant proteins[10] and several unnatural amino acids have now been incorporated by evolving this pair.[11,12] Here we report the synthesis and genetic incorporation of aliphatic azides and alkynes into proteins using the natural *Mb*PylRS/$tRNA_{CUA}$ pair and the efficient bio-orthogonal labeling of these amino acids using [3+2] cycloaddition ("click") chemistry.[13]

Since unnatural amino acids destined for incorporation into recombinant proteins are added to cell cultures at 1−10 mM[9] it is important that they can be synthesized in gram quantities via concise, efficient syntheses. Yokoyama and co-workers recently reported the genetic incorporation of the aromatic, photoreactive lysine derivative N^{ϵ}-(o-azidobenzyloxycarbonyl-lysine) using a mutant pyrrolysine synthetase/tRNA pair.[11] However, the synthetic route, yield, and NMR characterization of this amino acid were not reported. Very recently Chan and co-workers reported the incorporation of a direct pyrrolysine analogue with an appended alkyne.[14] The pyrrolysine analogue was synthesized in 17% yield after 16 steps. We designed and synthesized aliphatic, photostable amino acids **1** and **2** that link the alkyne functional group to a lysine residue via an amide or carbamate bond (Figure 1). These amino acids were synthesized in two steps and 70−80% yield from commercially available material (Supplementary Scheme 1, Supplementary Figure 1, and Supplementary Methods).

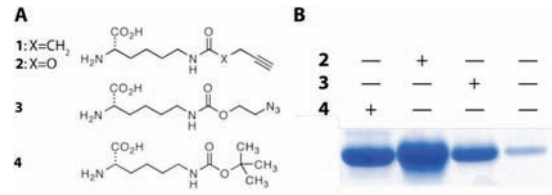

Figure 1. (A) Alkyne **1**, **2**, or azide **3**. (B) Ni-NTA Purified myoglobin-his6 from cells containing the PylRS/$tRNA_{CUA}$ orthogonal pair. **4** is a known efficient substrate for the PylRS/$tRNA_{CUA}$ pair.

To investigate whether **1** and **2** are substrates for the *Mb*PylRS/$tRNA_{CUA}$ pair, we transformed *E. coli* with pBKPylS[10] (which encodes *Mb*PylRS) and p*Myo*4TAGPylT-*his6*[10] (which encodes *Mb*$tRNA_{CUA}$ and a C-terminally hexahistidine tagged myoglobin gene with an amber codon at position 4). We added **1** or **2** (1 mM) to log phase cells and induced myoglobin-his6 expression. While in the presence of **1**, only background levels of myoglobin-his6 were purified by Ni-NTA chromatography; full-length myoglobin was purified in good yield (2 mg/L, comparable to that for other unnatural amino acids[5,10]) after expression in the presence of **2**, indicating that **2** but not **1** is incorporated by the *Mb*PylRS/$tRNA_{CUA}$ pair. This may reflect the greater flexibility of the carbamate linkage. The yield of protein containing **2** was not improved by efforts to evolve the enzyme but was increased 5-fold by increasing the concentration of **2** 7.5-fold (Supplementary Figure 2A).

Previous work on genetically encoding alkynes in *E. coli* used LC-coupled to MS or MS/MS of tryptic fragments to demonstrate the incorporation.[7,11,14] Since the ionization of closely related tryptic peptides may be very different, it is difficult to confirm the fidelity of incorporation *via* these methods. To demonstrate that **2** is incorporated with high fidelity and without modification by the cell[10] we used total protein electrospray ionization mass spectrometry (ESI-MS). Myoglobin-his6 incorporating **2** has the expected mass (found: 18477.5 ± 1 Da, expected: 18478.2 Da, Supplementary Figure 2B). These experiments demonstrate that **2** can be site-specifically incorporated into recombinant proteins in good yield and with high selectivity.

To investigate if the carbamate linkage provides a general route to the incorporation of other functional groups suitable for bioconjugation, we synthesized a simple aliphatic azide **3** (2 steps, 80% yield, Supplementary Scheme 1 and Methods). Protein expression and ESI-MS (Figure 1, Supplementary Figure 2C) experiments demonstrate that **3** is site-specifically incorporated into proteins in good yield (3 mg/L) using the *Mb*PylRS/*Mb*$tRNA_{CUA}$ pair.

To demonstrate that recombinant proteins containing the alkyne amino acid **2** can be site-specifically labeled with azido probes (via a Cu-catalyzed Huisgen [3+2] cycloaddition reaction[13]), myoglo-

[†] Medical Research Council Laboratory of Molecular Biology.
[‡] North Carolina State University.

10.1021/ja900553w CCC: $40.75

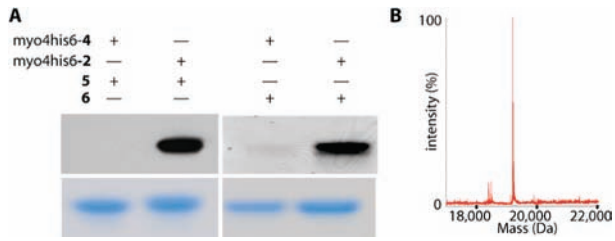

Figure 2. (A) Efficient and specific labeling of genetically encoded **2** with azido probes. (Left) The biotin azide **5** labeling reaction was performed on myoglobin containing **4** or **2** at position 4 (myo-4his6-4 and myo-4his6-2). Proteins were probed for biotin (top). (Right) By3 labeling with **6** was imaged directly. Coomassie stained protein gels (bottom) demonstrate equal protein recovery in the samples. (B) ESI-MS of the myoglobin-his6 containing **2** labeled with biotin azide **5** (Found: 19199.5 ± 1.5 Da, expected: 19198.2 Da).

bin-his6 bearing **2** at position 4 was treated with the biotin azide **5** or a fluorophore (By3) azide **6** (Supplementary Figure 3), in the presence of $CuSO_4$, 4,7-diphenyl-1,10-phenanthrolinedisulfonic acid disodium salt, and ascorbate in sodium phosphate buffer (pH 8.3).[15] In control experiments myoglobin-his6 bearing **4** at position 4 was treated identically. After 18 h, the purified labeling reactions were probed and analyzed by SDS-PAGE (Figure 2A). These experiments demonstrate the specific labeling of the alkyne containing protein.

Previous work has visualized protein labeling by gel-based methods alone,[7,11,14] which does not allow quantification of labeling efficiency. ESI-MS of our purified labeling reaction (Figure 2B) demonstrates a labeling efficiency of 90−100%. Quantification of the ratio of biotin or By3 to protein in purified samples provides independent confirmation of the labeling efficiency (Supplementary Methods).

In conclusion, we report the efficient synthesis and site-specific, genetically encoded incorporation of aliphatic amino acids bearing a carbon−carbon triple bond and an azido group into recombinant proteins and the near-quantitative on-protein labeling of the alkyne. In contrast to previous work[11,14] the amino acids can be synthesized in just two steps in excellent yield and site-specifically incorporated. Unlike Schultz and coworkers we began with an orthogonal synthetase/tRNA pair that does not use any amino acid naturally present in *E. coli*, allowing us to discover useful unnatural amino acids that the synthetase will use as substrates without the need for a series of enzyme evolution steps. In contrast to the aromatic azides previously incorporated by the Schultz[5] or Yokoyama[11] group the aliphatic azide we have incorporated is photostable and therefore easy to handle.

The genetically encoded alkyne **2** can be specifically and efficiently labeled with azides that introduce biotin or fluorescent groups, and in contrast to previous reports[7,11,14] we have explicitly demonstrated and quantified the efficient conjugation of probes to the genetically encoded amino acid. Since many protein therapeutics are conjugated in a residue specific manner to polyethylene glycols through lysine,[16] the method reported may provide a direct route to discovering site-specifically modified versions of these therapeutics that are more efficacious. The labeling method is compatible with and orthogonal to cysteine labeling and will prove useful for introducing two distinct labels into a single protein for fluorescence resonance energy transfer (FRET) experiments to probe protein function, structure, and dynamic behavior. Moreover, since this synthetase tRNA pair is functional and orthogonal in eukaryotic cells,[12] it will be possible to extend the approach reported here to the labeling of proteins produced in, and displayed on, eukaryotic cells.

The alkyne **2** and azide **3** are incorporated using a synthetase and tRNA pair that is mutually orthogonal in its aminoacylation specificity to the *Mj*TyrRS/$tRNA_{CUA}$ pair (JWC unpublished) that has been used to incorporate a range of aromatic unnatural amino acids.[4] This suggests it will be possible to incorporate **2** or **3** in combination with genetically encoded aromatic amino acids, including previously incorporated azides[5] and alkynes,[7] at distinct sites in recombinant proteins using suitably altered combinations of synthetase/tRNA pairs and evolved orthogonal ribosomes.[17] This will enable the formation of directional intramolecular cross-links to constrain protein structure and may allow for the genetic selection of enhanced protein stability and function.

Acknowledgment. A.D. acknowledges support by North Carolina State University and is a Beckman Young Investigator and Cottrell Scholar. J.W.C. acknowledges the MRC and European Research Council for support.

Supporting Information Available: . Experimental protocols and supplementary figures. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hermanson, G. T. *Bioconjugate Techniques*; Academic Press: 1996.
(2) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 19–24.
(3) Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652–3.
(4) Xie, J.; Schultz, P. G. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775–82.
(5) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 9026–7.
(6) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. *Science* **2003**, *301*, 964–7.
(7) Deiters, A.; Schultz, P. G. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1521–4.
(8) Zhang, Z.; Smith, B. A.; Wang, L.; Brock, A.; Cho, C.; Schultz, P. G. *Biochemistry* **2003**, *42*, 6735–46.
(9) Ambrogelly, A.; Gundllapalli, S.; Herring, S.; Polycarpo, C.; Frauer, C.; Soll, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 3141–6.
(10) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Chem. Biol.* **2008**, *4*, 232–4.
(11) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. *Chem. Biol.* **2008**, *15*, 1187–97.
(12) Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818–22.
(13) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2001**, *40*, 2004–2021.
(14) Fekner, T.; Li, X.; Lee, M. M.; Chan, M. K. *Angew. Chem.* **2009**, *121*, 1661–1663.
(15) Schoffelen, S.; Lambermon, M. H.; van Eldijk, M. B.; van Hest, J. C. *Bioconjugate Chem.* **2008**, *19*, 1127–31.
(16) Kozlowski, A.; Harris, J. M. *J. Controlled Release* **2001**, *72*, 217–24.
(17) Wang, K.; Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Biotechnol.* **2007**, *25*, 770–7.

JA900553W