J|A|C|S
COMMUNICATIONS

Published on Web 07/11/2002

Addition of *p*-Azido-L-phenylalanine to the Genetic Code of *Escherichia coli*

Jason W. Chin,[†] Stephen W. Santoro,[†] Andrew B. Martin,[†] David S. King,[‡] Lei Wang,[†] and Peter G. Schultz*,[†]

Department of Chemistry and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, and Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720-1406

Received May 22, 2002

Although proteins are involved in almost every biological process, they are biosynthesized from the same 20 common amino acids and contain a limited set of functional groups (nitrogen bases, carboxylic acids, amides, alcohols, and thiols). The ability to genetically incorporate into proteins amino acids containing "unnatural" functional groups would provide powerful tools for reporting on, and rationally controlling, protein function. One particulary interesting group for incorporation into proteins is the aryl-azide, a photocrosslinking agent first described in a biological context by Fleet and co-workers.[1] Upon irradiation with UV light at wavelengths below 310 nm, simple aryl-azides extrude molecular nitrogen to form short-lived (~1 ns) singlet nitrenes which rearrange to form dehydroazepines;[2] the latter react rapidly with amines to form robust adducts. Polyfluorination extends the lifetime of singlet aryl nitrenes to tens and hundreds of nanoseconds, and these species react efficiently with C−H and N−H bonds to form covalent adducts.[3]

Aryl-azides are among the most widely used photocrosslinking agents. They have been used to photochemically label antibodies with hapten,[1] irreversibly inactivate enzymes,[4] and probe protein−peptide[5] and protein−protein interactions.[6,7] The site-specific incorporation of aryl azides has previously been achieved in an in vitro translation system using N^{ϵ}-(5-azido-2-nitrobenzoyl)-Lys tRNA.[6] Despite the utility of aryl-azides in mapping protein interactions, no easy, general, high-yield route exists for their site-specific incorporation into proteins. We recently demonstrated that additional amino acids could be site-specifically incorporated with high translational fidelity into proteins in vivo[8] by the addition of new components to the translational machinery of *Escherichia coli* (*E.coli*). We now report the selection of a new orthogonal aminoacyl tRNA synthetase/tRNA pair for the in vivo incorporation of the photocrosslinker, *p*-azido-L-phenylalanine, into proteins in response to the amber codon, TAG. The amino acid is incorporated in good yield with high fidelity and can be used to crosslink interacting proteins.

The *Methanococcus jannaschii* tyrosyl-tRNA synthetase (Mj TyrRS) was used as a starting point for the generation of an orthogonal synthetase that incorporates *p*-azido-L-phenylalanine, but not any of the 20 common amino acids. The Mj TyrRS does not aminoacylate any endogenous *E. coli* tRNAs with tyrosine,[9] but aminoacylates a mutant tyrosine amber suppressor ($mutRNA_{CUA}$).[10] We used a Mj TyrRS library of mutants in which five residues (Tyr 34, Glu 107, Asp 158, Ile 159, Leu 162) were randomized.[8] Active synthetase variants (Table 1) were identified from the mutant TyrRS library using two different methods. AzPheRS variants **1**−**4** were identified using chloramphenicol acetyl transferase (CAT) and barnase reporter systems for positive and negative selection, respectively, as described.[11] AzPheRS variants **5**−**7** were identified using a CAT reporter for positive selection and a T7 RNA polymerase/green fluorescent protein reporter for negative screening using fluorescence activated cell sorting, as described.[12]

Table 1. AzPheRS Sequences

residue #	32	107	158	159	162
Mj TyrRS[a]	Tyr	Glu	Asp	Ile	Leu
AzPheRS-1	Thr	Asn	Pro	Leu	Gln
AzPheRS-2	Thr	Ser	Pro	Ser	Gln
AzPheRS-3	Thr	Ser	Pro	Leu	Gln
AzPheRS-4	Leu	Thr	Pro	Val	Gln
AzPheRS-5	Ala	Arg	Val	Ile	Asp
AzPheRS-6[b]	Gly	Thr	Thr	Tyr	Leu
AzPheRS-7[b]	Leu	Pro	Gln	Ile	Ser

[a] All proteins have an IC_{50} of chloramphenicol resistance of 110 mg/L with 1 mM *p*-azido-L-phenylalanine and less than 5 mg/L in the absence of added *p*-azido-L-phenylalanine. [b] These synthetases contained additional mutations (AzPheRS-6: His160Tyr; AzPheRS-7: Tyr161Ser).

Selected synthetases were characterized using an in vivo assay based on suppression of the Asp112TAG codon in the CAT gene. *E. coli* expressing the selected AzPhe-RS/$mutRNA_{CUA}$ pairs survived on chloramphenicol with IC_{50} values of 110 mg/L and less than 5 mg/L in the presence and absence of 1 mM *p*-azido-L-phenylalanine, respectively. The large difference in chloramphenicol resistance in the presence and absence of *p*-azido-L-phenylalanine suggests a substantial in vivo specificity of the selected synthetase/tRNA pairs for insertion of *p*-azido-L-phenylalanine over all 20 natural amino acids found in the cell in response to an amber codon.

To measure the fidelity and efficiency of *p*-azido-L-phenylalanine incorporation, we incorporated *p*-azido-L-phenylalanine in response to an amber codon at position four in a C-terminally hexahistidine tagged mutant sperm whale myoglobin gene[11] (Figure 1). In the presence of both the AzPhe-RS-1/$mutRNA_{CUA}$ pair and 1 mM *p*-azido-L-phenylalanine, full length myoglobin was produced in GMML (glycerol minimal media with leucine). The protein was purified via Ni-NTA affinity chromatography under a red photographic light, to avoid photolysis of the aryl-azide. The yield of purified protein was 2 mg/L, comparable to that found when the the Mj TyrRS/$mutRNA_{CUA}$ pair suppresses the same amber codon. No myoglobin protein was detectable by silver stain or Western blot against the His6 tag on myoglobin if any of the three components responsible for specific amber suppression with *p*-azido-L-phenylalanine (amino acid, synthetase, or tRNA) were withheld. These data provide further evidence that the selected synthetase is very selective for *p*-azido-L-phenylalanine.

* To whom correspondence should be addressed. E-mail: schultz@scripps.edu.
[†] The Scripps Research Institute.
[‡] HHMI, UC Berkeley.

10.1021/ja027007w CCC: $22.00

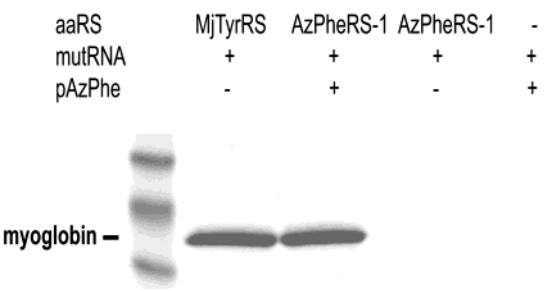

Figure 1. *p*-Azido-L-phenylalanine dependent expression of sperm whale myoglobin in response to an amber codon at the fourth position in the gene for this protein.

Electrospray-ionization ion trap mass spectrometry of the mutant myoglobin gave a mass of 18456.40 ± 0.47, which is identical to the calculated mass of 18456.21 for the *p*-azido-L-phenylalanine containing myoglobin. This confirms the incorporation of *p*-azido-L-phenylalanine at a single site in the protein. No masses were observed in the mass spectra corresponding to natural amino acid incorporation or reduction of the azide,[13] providing additional evidence for the high fidelity, stable incorporation of *p*-azido-L-phenylalanine.

To demonstrate the utility of incorporating *p*-azido-L-phenylalanine for mapping protein−protein interactions, model crosslinking experiments were carried out with glutathione-*S*-transferase, a dimer of two identical subunits.[14] An AzPheRS/mutRNA$_{CUA}$ pair was used to site-specifically incorporate *p*-azido-L-phenylalanine into C-terminally hexahistidine tagged SjGST at one of two sites, in response to the corresponding amber codon within the SjGST gene.

SjGST containing *p*-azido-L-phenylalanine at residue 52, which lies at the dimer interface,[11] was purified by Ni-NTA affinity chromatography before crosslinking with a handheld UV lamp. The crosslinked products were resolved by SDS−PAGE and conversion to the covalently linked homodimer assayed by Western blot with an anti-GST antibody (Figure 2). These experiments revealed crosslinking of between 10 and 30% of the mutant protein after 5 min. In contrast, control experiments using either wild-type SjGST or SjGST containing *p*-azido-L-phenylalanine at residue 198, which lies outside the dimer interface, show no detectable crosslinking in response to UV irradiation. To investigate the generality of the method with respect to the crosslinking site, *p*-azido-L-phenylalanine was incorporated at three additional sites at the dimer interface of SjGST. All mutants were crosslinked upon exposure to UV light (Figure S1). These results demonstrate that site-specific *p*-azido-L-phenylalanine substitution can be used to define amino acids involved in a protein−protein interaction.

In conclusion, we have developed a tool for site-specifically incorporating *p*-azido-L-phenylalanine into proteins expressed in *E. coli* with high translational fidelity and efficiency. The in vivo method reported herein provides a route to several milligrams of purified protein per liter of bacterial culture and can, in principle, be applied to any protein that is expressed in *E. coli*, irrespective of its size or sequence. Together with our recent report of the addition of *p*-benzoyl-L-phenylalanine to the genetic code of *E. coli*,[11] this tool should facilitate the study of protein function in vitro and potentially in vivo.

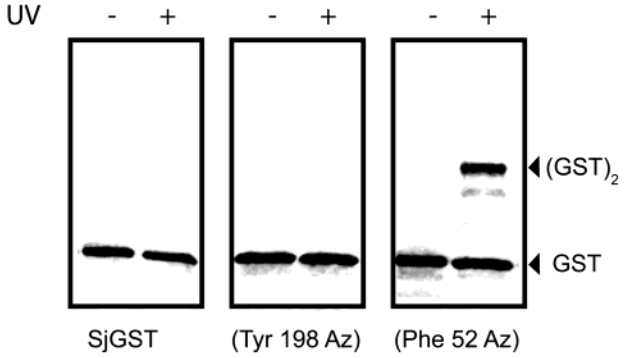

Figure 2. Surface specific photocrosslinking of *p*-azido-L-phenylalanine containing SjGST. SjGST Phe-52-AzPhe, SjGST Tyr-198-AzPhe, or SjGST was irradiated for 0 or 5 min with a handheld UV lamp (254 nm) and monomer and covalent dimer resolved by SDS−PAGE.

Acknowledgment. J.W.C. and S.W.S. are fellows of the Damon Runyon Cancer Research Foundation (DRG# 1707-02) and the Jane Coffin Childs Memorial Fund, respectively. J.W.C. thanks Stephanie Chu (UC San Diego) for technical assistance. This work was supported by the Department of Energy and the Skaggs Institute. This is manuscript #15031-CH of the Scripps Research Institute.

Supporting Information Available: Crosslinking data for additional sites in GST (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Fleet, G. W.; Porter, R. R.; Knowles, J. R. *Nature* **1969**, *224*, 511.
(2) Hermanson, G. T. *Bioconjugate Techniques*; Academic Press: San Diego, 1996.
(3) (a) Keana, J. F. W.; Cai, S. X. *J. Org. Chem.* **1990**, *55*, 3640. (b) Gritsan, N. P.; et al. *J. Am. Chem. Soc.* **2001**, *123*, 1951−1962. (c) Borden, W. T.; et al. *Acc. Chem. Res.* **2000**, *33*, 765−811. Gritsan, N. P.; Platz, M. S. *Adv. Phys. Org. Chem.* **2001**, *36*, 255−304.
(4) Escher, E.; Jost, R.; Zuber, H.; Schwyzer, R. *Isr. J. Chem.* **1974**, *12*, 129−138.
(5) (a) Ong, H.; et al. *Endocrinology* **1998**, *139*, 432−435. (b) Bodart, V.; et al. *Circ. Res.* **1999**, *85*, 796−802.
(6) Krieg, U. C.; Walter, P.; Johnson, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8604−8608.
(7) Chen, Y.; Ebright, Y. W.; Ebright, R. H. *Science* **1994**, *265*, 90−92.
(8) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498−500.
(9) Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. *J. Am. Chem. Soc.* **2000**, *122*, 5010−5011.
(10) Wang, L.; Schultz, P. G. *Chem. Biol.* **2001**, *8*, 883−890.
(11) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, in press.
(12) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. **2002**, in preparation.
(13) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 19−24.
(14) (a) Maru, Y.; Afar, D. E.; Witte, O. N.; Shibuya, M. *J. Biol. Chem.* **1996**, *271*, 15353−15357. (b) McTigue, M. A.; Williams, D. R.; Tainer, J. A. *J. Mol. Biol.* **1995**, *246*, 21−27.

JA027007W