J|A|C|S
COMMUNICATIONS

Published on Web 06/20/2006

A Genetically Encoded Fluorescent Amino Acid

Jiangyun Wang, Jianming Xie, 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

Received April 17, 2006; E-mail: schultz@scripps.edu

The ability to selectively modify proteins with fluorescent probes has greatly facilitated both in vitro and in vivo studies of protein structure and function.[1,2] For example, green fluorescent protein (GFP) is a powerful probe of protein expression, localization, and bimolecular interactions. However, GFP fusions are limited to the C- or N- terminus of the target protein, relatively insensitive to environment, and GFP can cause significant perturbations due to its size.[2] Chemical methods also can be used to selectively modify proteins with a variety of synthetic fluorophores[1,3] but are generally limited to uniquely reactive surface accessible residues on isolated proteins (e.g., the modification of cysteine with maleimide derivatives).[1] Biosynthetic labeling methods using chemically misacylated tRNAs[4] afford limited yields of protein and are typically carried out in vitro. Methodology that would allow one to genetically encode fluorescent amino acids at defined sites in proteins directly in living organisms would significantly extend the scope of this technique.[5] Here we report the generation of an orthogonal tRNA/aminoacyl-tRNA synthetase pair that allows the selective introduction of the fluorescent amino acid L-(7-hydroxycoumarin-4-yl) ethylglycine **1** (Figure 1A) into proteins in *E. coli* in response to the amber stop codon, TAG.

The 7-hydroxycoumarin moiety was initially investigated due to its high fluorescence quantum yield, relatively large Stoke's shift (Figure 1B), small size, and sensitivity to pH (Figure S1) and solvent polarity.[6] The coumarinyl amino acid **1** was synthesized by first converting *N*-α-Cbz-L-glutamic acid α-benzyl ester into the side-chain *β*-keto ester, which was then reacted with resorcinol in methanesulfonic acid (von Pechmann reaction)[7] to afford amino acid **1** (Figure 1A). To selectively incorporate amino acid **1** at defined sites in proteins in *E. coli*, a mutant *Methanococcus jannaschii* tyrosyl amber suppressor tRNA (MjtRNA$^{Tyr}_{CUA}$)/tyrosyl-tRNA synthetase (MjTyrRS) pair was evolved that uniquely specifies **1** in response to the TAG codon.[5]

To accommodate the large coumarin side chain, an MjTyrRS library pBK-lib5 was generated in which His70 was mutated to Gly and Ala67 was fixed as either Ala or Gly to increase the active site size. Six residues, Tyr-32, Leu-65, Phe-108, Gln-109, Asp-158, and Leu-162, in close proximity to bound tyrosine[8,9] were then randomized. Subsequently, this library was subjected to rounds of both positive and negative selections. In the positive selection, cell survival is dependent on the suppression of an amber codon introduced at a permissive site in the chloramphenicol acetyl transferase (CAT) gene when cells cotransformed with pBK-lib and MjtRNA$^{Tyr}_{CUA}$ are grown in the presence of 1 mM unnatural amino acid (UAA) and chloramphenicol.[10] Positively selected clones are then transformed into cells containing MjtRNA$^{Tyr}_{CUA}$ and a gene encoding the toxic barnase protein with three amber mutations introduced at permissive sites. These cells are grown in the absence of UAA to remove any clones that utilize endogenous amino acids (negative selection).

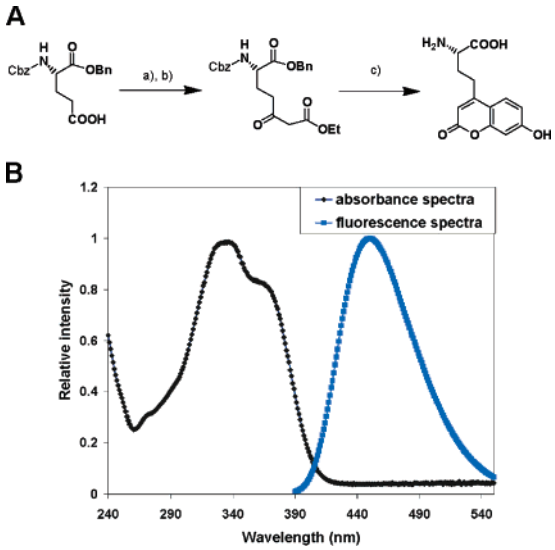

Figure 1. (A) Synthesis of L-(7-hydroxycoumarin-4-yl) ethylglycine **1**. (a) *N*,*N*′-Carbonyldiimidazole, rt, 2 h; (b) ethyl magnesium malonate, rt, overnight; (c) resorcinol, methanesulfonic acid, rt, 2 h. (B) Absorption and emission spectra of **1**. Both spectra were recorded at pH 7.4, 100 mM sodium phosphate buffer. At this pH, **1** is present in both phenol and phenolate forms (in approximately a 2:1 ratio). The phenolate form has an extinction coefficient of 17 000 at 360 nm and a quantum yield of 0.63.

Three rounds of positive and two rounds of negative selections of this library generated a clone that grows at 100 μg/mL chloramphenicol in the presence of **1**, but only at 20 μg/mL chloramphenicol in the absence of **1**. This clone CouRS-D8 has the following eight mutations: Tyr32Glu, Leu65His, Ala67Gly, His70Gly, Phe108Tyr, Gln109His, Asp158Gly, and Leu162Gly. Four residues are mutated to glycine, most likely to create enough space to accommodate **1**. Tyr32 and Asp158 which hydrogen bond to bound tyrosine in the wild-type enzyme are also mutated, consistent with the loss of activity toward tyrosine.[8,9]

To determine the efficiency and fidelity for the incorporation of **1** into proteins, an amber codon was substituted for Ser-4 in sperm whale myoglobin containing a C-terminal His_6 tag. Protein expression was carried out in the presence of the selected synthetase (CouRS-D8) and MjtRNA$^{Tyr}_{CUA}$ in *E. coli* grown in either minimal media or 2YT supplemented with 1 mM of **1**. As a negative control, protein expression was also carried out in the absence of **1**. Analysis of the purified protein by SDS-PAGE showed that the full length protein was expressed only in the presence of **1** (Figure 2). The yield of the mutant myoglobin in either media is 2 mg/L. For comparison, the yield of wild-type myoglobin under similar conditions is 5 mg/L. ESI-mass spectrometric analyses of the mutant myoglobin gave an observed average mass of 18 511 Da, in close agreement with the calculated mass of 18 512 Da (Figure S2, Supporting Information).

10.1021/ja062666k CCC: $33.50

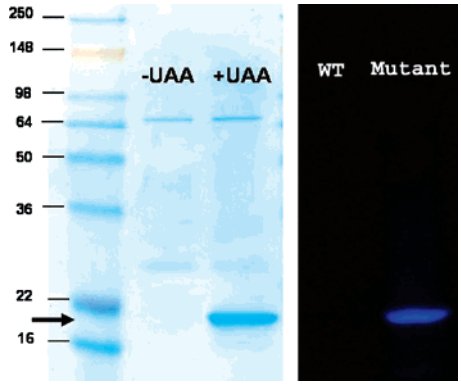

Figure 2. Coomassie-stained SDS-PAGE (left) of TAG4 mutant myoglobin (indicated by black arrow) expression in the presence and absense of 1 mM **1**. The right panel shows the fluorescence image of wild-type and TAG4 mutant myoglobin.

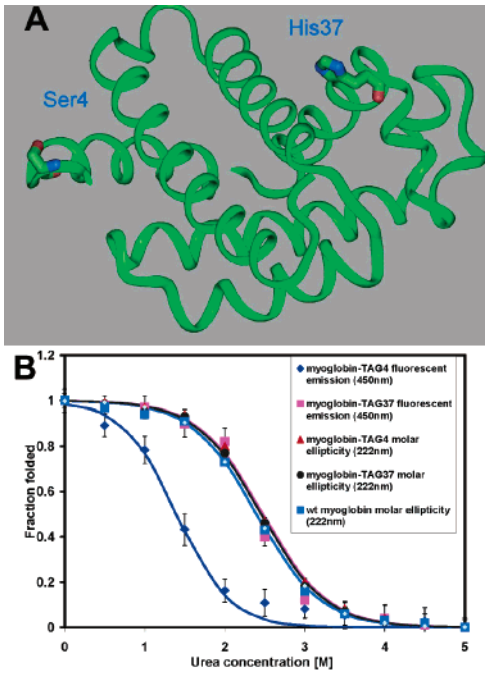

Figure 3. (A) Structure of sperm whale myoglobin (pdb code 105M). Residues Ser4 and His37 are shown as sticks. (B) Urea-induced unfolding of the Ser4 → **1** and His37 → **1** mutant holomyoglobins, monitored by both CD and the fluorescent intensity of **1**. Both experiments were carried out in 100 mM sodium phosphate, 300 mM NaCl, pH 7.4, with various concentrations of urea, as indicated. In the prescence of 5 M urea, both myoglobin mutants show a 30% increase in fluorescence signal compared to the signal in the presence of 0 M urea. The "fraction folded" is calculated by normalizing the fluorescence signal or molar ellipticity at 5 M urea to the fully unfolded state.

To demonstrate the utility of **1**, it was used as a probe of the urea dependent denaturation of holomyoglobin. Because coumarin fluorescence is sensitive to solvent polarity, its fluorescence intensity should correlate with local unfolding of the protein in close proximity to **1**. The myoglobin structure consists of eight helices (A to H) connected by short loops and turns.[11] Ser4 in helix A and His37 in helix C (both residues are largely solvent exposed and do not significantly interact with other nearby residues) were each mutated to **1** (Figure 3A). As indicated in Figure 3B, the urea induced unfolding curves of the wild type, Ser4 → **1** and His37 → **1** holomyoglobins, as monitored by circular dichroism, are virtually identical, suggesting that introduction of coumarin **1** into either helix A or helix C does not significantly perturb protein stability. At 2 M urea concentration, the fluorescence intensity of **1** at position 4 in holomyogloin increases 30% (and remains at this level from 2 to 5 M urea), suggesting that this region of the protein is disordered.[9] In contrast, mutant myoglobin with **1** at residue 37 shows little change in its fluorescence intensity at 2 M urea, but undergoes a similar fluorescence increase at 3 M urea. Consistent with this result, a previous NMR study has shown that when the urea concentration is higher than 2.2 M, helix A and B are largely disordered, as shown by the disappearance of short and medium range NOEs in this region, whereas helices C, D, and F unfold later, when the urea concentration is higher than 3.0 M.[11] Thus, it appears that amino acid **1** is a site-specific probe of protein conformational changes, in contrast to circular dichroism, which reports global changes averaged over the entire structure.

The sensitivity of **1** to solvent polarity and pH should make it a useful probe for many biological studies, both in vitro and in vivo.[7] For example, amino acid **1** can be used to monitor bimolecular interactions or conformational changes in proteins or the topology of membrane-bound proteins. In addition, because **1** has a pK_a of 7.8, which can be systematically altered by substitution of the coumarin ring,[12] it should be a useful probe of organellar pH and pH-dependent cellular processes. Moreover, in its excited state, 7-hydroxycoumarin is both a strong photoacid[13] and may facilitate the study of proton-transfer processes in proteins.

Acknowledgment. This manuscript is dedicated to Peter B. Dervan on the occasion of his 60th birthday. We thank Dr. Q. Zhang and Dr. J. C. Anderson for helpful discussions. We are grateful to the DOE (ER46051) and the Skaggs Institute for Chemical Biology.

Supporting Information Available: Materials and methods (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hermanson, G. T. *Bioconjugate Techniques*; Academic Press: San Diego, CA, 1996.
(2) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509−44.
(3) (a) Martin, B. R.; Giepmans, B. N. G.; Adams, S. R.; Tsien, R. Y. *Nat. Biotechnol.* **2005**, *23*, 1308−14. (b) Keppler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K. **2003**, *Nat. Biotechnol. 21*, 86−9. (c) Lin, C.-W.; Ting, A. Y. *J. Am. Chem. Sci.* **2006**, *128*, 4542−3.
(4) (a) Mendel, D.; Cornish, V. W.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 435−62. (b) Hohsaka, T.; Muranaka, N.; Komiyama, C.; Matsui, K.; Takaura, S.; Abe, R.; Murakami, H.; Sisido, M. *FEBS Lett.* **2004**, *560*, 173−7.
(5) (a) Wang, L.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2005**, *44*, 34−66. (b) Wang, L.; Xie, J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225−49.
(6) Zinsli, P. E. *J. Photochem.* **1974**, *3*, 55−69.
(7) Brun, M. P.; Bischoff, L.; Garbay, C. *Angew. Chem., Int. Ed.* **2004**, *43*, 3432−6.
(8) Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S. *Nat. Struct. Biol.* **2003**, *10*, 425−32.
(9) Zhang, Y.; Wang, L.; Schultz, P. G.; Wilson, I. A. *Protein Sci.* **2005**, *14*, 1340−9.
(10) Xie, J. M.; Schultz, P. G. *Methods* **2005**, *36*, 227−38.
(11) Delli Castelli, D.; Lovera, E.; Ascenzi, P.; Fasano, M. *Protein Sci.* **2002**, *11*, 2273−6.
(12) Sun, W. C.; Gee, K. R.; Haugland, R. P. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3107−10.
(13) Cohen, B.; Huppert, D. *J. Phys. Chem. A* **2001**, *105*, 7157−64.

JA062666K