Nonnatural Amino Acids

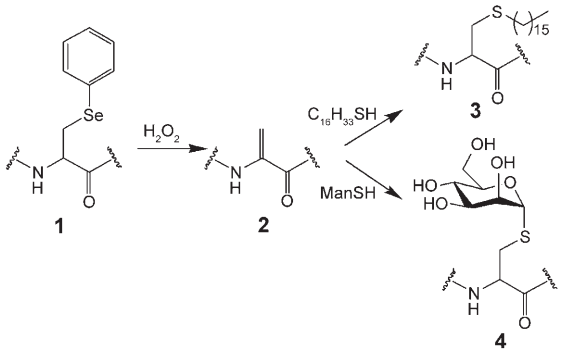
DOI: 10.1002/anie.200702305

A Biosynthetic Route to Dehydroalanine-Containing Proteins**

*Jiangyun Wang, Stefan M. Schiller, and Peter G. Schultz**

The amino acid dehydroalanine (Dha) is found in a number of proteins and nonribosomal natural products, and typically arises from post-translational modifications of serine or cysteine.[1–4] This α,β-unsaturated amino acid can alter the conformation, rigidity, and proteolytic susceptibility of the polypeptide backbone.[3,5] In addition, intra- or intermolecular electrophilic Michael addition reactions with this reactive amino acid can lead to the formation of Dha adducts, including intramolecular lanthionine crosslinks.[2,3] Recently, it has also been shown that Dha is a versatile electrophilic handle for the selective chemical modification of proteins with a wide variety of exogenous groups.[6,7] For example, Michael addition reactions with the corresponding thiols can be used to synthesize polypeptides that contain S-prenylated and S-glycosylated cysteine[8]. However, because free Dha is not stable, typically a precursor such as phenylselenocysteine[7] or selenolysine[9] is incorporated into polypeptides and efficiently converted into Dha under mild conditions.

The ability to genetically incorporate the Dha precursor, phenylselenocysteine **1**, directly into proteins in *Escherichia coli* would greatly facilitate the preparation of Dha-containing proteins, especially those that are not amenable to peptide synthesis or semisynthesis (Scheme 1). To selectively incorporate the amino acid **1** at defined sites in proteins in *E. coli,* a mutant *Methanococcus jannaschii* tyrosyl amber suppressor tRNA (MjRNA$^{Tyr}_{CUA}$)/tyrosyl-tRNA synthetase (MjTyrRS) pair was evolved that uniquely specifies **1** in response to the TAG codon.[10] To accommodate the phenylselenide side chain, a MjTyrRS library pBK-lib5[11] was generated in which His70 was first mutated to Gly, and Ala67 was fixed as either Ala or Gly to increase the size of the active site. Six residues (Tyr-32, Leu-65, Phe-108, Gln-109, Asp-158, and Leu-162) in proximity to bound tyrosine were then randomized.[11] 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 co-transformed with pBK-lib and MjRNA$^{Tyr}_{CUA}$ are grown in the presence of 1 mM **1** and chloramphenicol. Positively selected clones are then transformed into cells containing MjRNA$^{Tyr}_{CUA}$ and a gene encoding the toxic barnase protein with three amber mutations introduced at permissive sites. These cells are grown in the absence of **1** to remove any clones that utilize endogenous amino acids (negative selection). Three rounds of positive and two rounds of negative selection afforded three clones (Table S1 in the Supporting Information) that grow at more than 100 μg mL^{-1} of chloramphenicol in the presence of **1**, but less than 40 μg mL^{-1} chloramphenicol in the absence of **1**. Sequencing of clone PhSeRS-K4, which had the highest activity, revealed the following six mutations: Tyr32Trp, Leu65His, His70Gly, Phe108 Asn, Gln109Ser, Asp158Ser, and Leu162Glu; Ala67 is conserved.



Scheme 1. The genetic incorporation of phenylselenocysteine, followed by oxidative elimination, and Michael addition of the corresponding thiols results in the formation of (*S*)-hexadecylcysteine or (*S*)-mannosylcysteine.

To determine the efficiency and fidelity of the incorporation of **1** into proteins, an amber codon was substituted for Gly-242 in green fluorescent protein (GFP) containing a C-terminal His_6 tag. Protein expression was carried out in the presence of the selected synthetase (PhSeRS-K4) and MjRNA$^{Tyr}_{CUA}$ in *E. coli* grown in minimal media supplemented with 1 mM **1**. As a negative control, protein expression was carried out in the absence of **1**. Analysis of the purified protein by SDS-PAGE showed that full length GFP was expressed only in the presence of **1** (Figure 1 A). The yield of the mutant GFP was 5 mg L^{-1}. For comparison, the yield of wild-type GFP under similar conditions was 40 mg L^{-1}. Electrospray ionization mass spectrometry (ESI-MS) of the GFP Gly242→**1** mutant gave an observed average mass of 28254 Da, in close agreement with the calculated mass of 28253 Da (Figure S1 in the Supporting Information). These results confirm a high fidelity for incorporation of **1** into proteins in response to the amber codon TAG.

To demonstrate that mutant proteins that contain **1** can be selectively modified with thiol-containing moieties, the mutant GFP was first treated with 100 mM of H_2O_2 at room

[*] Dr. J. Wang, Dr. S. M. Schiller, Prof. P. G. Schultz
Department of Chemistry and the Skaggs Institute for Chemical Biology
The Scripps Research Institute
10550 North Torrey Pines Road, La Jolla, CA 92037 (USA)
Fax: (+1) 858-784-9440
E-mail: schultz@scripps.edu

[**] We thank Dr. M. Sever and Dr. J. Guo for helpful discussions. We thank Bill Webb for help in protein mass spectroscopy. We are grateful to the DOE (ER46051) and the Skaggs Institute for Chemical Biology for support.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

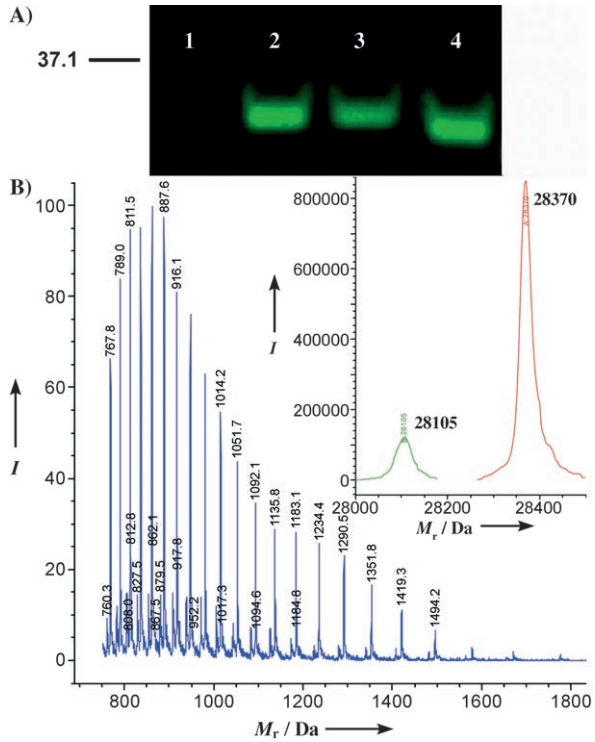


Figure 1. A) Fluorescence image of SDS-PAGE of the TAG242 mutant GFP expression in the presence (lane 2) and absence (lane 1) of 1 mM of **1**. Lanes 3 and 4 show the TAG242→(*S*)-mannosylcysteine and TAG242→(*S*)-hexadecylcysteine mutant GFPs, respectively. B) ESI-MS spectra of the TAG242→(*S*)-hexadecylcysteine mutant GFP. The insert shows the de-convoluted spectrum; expected mass: 28370 Da, found: 28370 Da. The signal at 28105 Da corresponds to TAG242→dehydroalanine mutant GFP that did not react with *n*-hexadecyl mercaptan.

temperature for one hour. Quantitative conversion of **1** into dehydroalanine was confirmed by ESI-MS: the predicted mass of GFP containing dehydroalanine at position 242 is 28112 Da; the observed mass was 28112 Da (see Figure S2 in the Supporting Information). The protein was then treated with thiol-containing nucleophiles (Scheme 1). To introduce S-glycosylated cysteine, the protein was allowed to react with 100 mM 1-thio-α-D-mannopyranose in pH 8.8, 50 mM Tris buffer in the presence of 50 mM tris(2-carboxyethyl)phosphine (TCEP) at room temperature for 2 hours. ESI-MS was used to monitor the reaction. As expected, reaction with 1-thio-α-D-mannopyranose resulted in a new signal at 28307 Da (see Figure S3 in the Supporting Information), in close agreement with the calculated mass of 28306 Da for the Gly242→**4** mutant, with a yield of 80%. Although it has been shown previously that Michael addition to Dha lacks diastereoselectivity in short peptides,[7] it is likely that in well-folded proteins, the diastereoselectivity will be determined by the local structure at the site of modification.

Next we attempted to generate mimics of palmitoylated proteins.[12] Since palmitoylation is reversible with a half-life of only a few hours in vivo,[12] substitution by its nonhydrolyzable analogue, (*S*)-hexadecylcysteine, can facilitate studies of its cellular activities.[13] To incorporate (*S*)-hexadecylcysteine into GFP, the Gly242→**2** mutant was added to pH 8.8, 50 mM Tris buffer containing 2% cetyltrimethylammonium bromide (CTAB), 1% *n*-hexadecyl mercaptan, 50 mM TCEP, and incubated at room temperature for 12 hours with gentle shaking. As shown in Figure 1B, a minor signal at 28105 corresponding to the Gly242→**2** mutant starting material, and a major signal at 28370 corresponding to the Gly242→**3** mutant product were observed. The relative signal intensity of the reacted and unreacted GFP indicate that 90% of GFP was transformed into the Gly242→**3** mutant.

In conclusion, we have demonstrated a novel chemoselective method to introduce (*S*)-hexadecylcysteine and (*S*)-mannosylcysteine, which are structural and functional analogues of important post-translational modifications, into proteins. An advantage of this approach is the large diversity of post-translational modifications and analogues that can be introduced in a single step starting from the same Dha-containing protein. In addition, all the reactions were carried out in water, with a near-neutral pH, at room temperature, without the requirement of enzymes or complex organic synthesis. Extension to other modifications should allow us to address a wide range of biochemical questions, including in the case of S-prenylated cysteine,[14] the role of lipidation patterns in the modulation of protein localization in living cells,[15,16]In addition, this method may allow the synthesis of proteins with novel backbone conformations,[5] as well as the generation of cyclic and crosslinked proteins with enhanced activity or stability.

Experimental Section

To express mutant green fluorescent protein, plasmid pEt-GFPx242 was co-transformed with pSup-vector-containing PhSeRS-K4 into BL21(DE3) *E. coli* cells. The cells were amplified in 2YT media (5 mL), supplemented with kanamycin (50 μg mL^{-1}) and chloramphenicol (50 μg mL^{-1}). The starter culture (1 mL) was used to inoculate 30 mL of liquid GMML supplemented with appropriate antibiotics and grown at 37 °C overnight. These cells were further diluted into liquid GMML (1 L) with appropriate antibiotics and grown at 37 °C for 6–10 h. When the cells had reached OD600 0.4, liquid GMML (100 mL) that contained **1** (10 mM) was added. After 1 hour, protein expression was induced by addition of IPTG (0.5 mM). After another 4–12 hours of growth at 30 °C, cells were harvested by centrifugation. The GFP mutant was then purified by Ni-NTA affinity chromatography under native conditions.

To prepare lipidated or glycosylated protein, green fluorescent protein (10 μM) was treated with H_2O_2 (100 mM) for 30 minutes at room temperature in Tris buffer (pH 8.5, 50 mM). The protein sample was then dialyzed to remove H_2O_2 and phenylselenic acid. After oxidative elimination, the protein (10 μM) was allowed to react with 1-thio-α-D-mannopyranose (100 mM) in Tris buffer (pH 8.8, 50 mM) in the presence of TCEP (50 mM), at room temperature for 2 hours. To incorporate (*S*)-hexadecylcysteine, the protein was incubated in 1% of *n*-hexadecylmercaptan, 2% of cetyltrimethylammonium bromide, Tris buffer (pH 8.8, 50 mM), in the presence of TCEP (50 mM), for 12 hours with gentle shaking. The modified protein was then purified by Ni-NTA affinity column chromatography.

Received: May 24, 2007
Published online: August 8, 2007

Keywords: glycosylation · lipidation · mutagenesis · phenylselenocysteine · protein modifications

[1] B. Langer, D. Rother, J. Retey, *Biochemistry* **1997**, *36*, 10867.
[2] C. Chatterjee, M. Paul, L. L. Xie, W. A. van der Donk, *Chem. Rev.* **2005**, *105*, 633.
[3] P. D. Cotter, C. Hill, R. P. Ross, *Nat. Rev. Microbiol.* **2005**, *3*, 777.
[4] P. D. Cotter, C. Hill, R. P. Ross, *Curr. Protein Pept. Sci.* **2005**, *6*, 61.
[5] D. E. Palmer, C. Pattaroni, K. Nunami, R. K. Chadha, M. Goodman, T. Wakamiya, K. Fukase, S. Horimoto, M. Kitazawa, H. Fujita, A. Kubo, T. Shiba, *J. Am. Chem. Soc.* **1992**, *114*, 5634.
[6] D. P. Galonic, W. A. van der Donk, D. Y. Gin, *Chem. Eur. J.* **2003**, *9*, 5997.
[7] Y. Zhu, W. A. van der Donk, *Org. Lett.* **2001**, *3*, 1189.
[8] L. A. Marcaurelle, C. R. Bertozzi, *Chem. Eur. J.* **1999**, *5*, 1384.
[9] F. P. Seebeck, J. W. Szostak, *J. Am. Chem. Soc.* **2006**, *128*, 7150.
[10] J. M. Xie, P. G. Schultz, *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775.
[11] J. Y. Wang, J. M. Xie, P. G. Schultz, *J. Am. Chem. Soc.* **2006**, *128*, 8738.
[12] J. E. Smotrys, M. E. Linder, *Annu. Rev. Biochem.* **2004**, *73*, 559.
[13] B. Bader, K. Kuhn, D. J. Owen, H. Waldmann, A. Wittinghofer, J. Kuhlmann, *Nature* **2000**, *403*, 223.
[14] F. L. Zhang, P. J. Casey, *Annu. Rev. Biochem.* **1996**, *65*, 241.
[15] O. Rocks, A. Peyker, M. Kahms, P. J. Verveer, C. Koerner, M. Lumbierres, J. Kuhlmann, H. Waldmann, A. Wittinghofer, P. I. H. Bastiaens, *Science* **2005**, *307*, 1746.
[16] D. A. Zacharias, J. D. Violin, A. C. Newton, R. Y. Tsien, *Science* **2002**, *296*, 913.

