Protein Modifications

DOI: 10.1002/anie.200704074

Addition of an α-Hydroxy Acid to the Genetic Code of Bacteria**

*Jiantao Guo, Jiangyun Wang, J. Christopher Anderson, and Peter G. Schultz**

Recently, a large number of amino acid building blocks with novel physical, biological, and chemical properties have been genetically encoded in both prokaryotic and eukaryotic organisms.[1,2] However, to date, substitutions in the polypeptide backbone itself are largely limited to in vitro protein translation systems[3–7] or semisynthetic chemical methods.[8–11] These methods generally place restrictions on the nature, size, or amounts of mutant protein that can be produced. Herein we show that the α-hydroxy acid *p*-hydroxy-L-phenyllactic acid can be directly incorporated into proteins in *E. coli* with good yields and high fidelity in response to an amber nonsense codon by means of an orthogonal amber suppressor tRNA/aminoacyl-tRNA synthetase pair[1,2] from archaebacteria (which does not cross-react with their endogenous counterparts in the host cell). The site-specific introduction of backbone ester mutations is a useful method to study the role of the polypeptide backbone in structure, folding, biomolecular recognition, and catalysis, as well as to cleave or derivatize the protein backbone selectively. The utility of this method was illustrated by the site-specific hydrolysis of a protein affinity tag, as well as determination of the energetic contribution of backbone hydrogen bonds to protein stability.

To determine whether α-hydroxy acids can be genetically introduced into proteins in bacteria, we initially focused on *p*-hydroxy-L-phenyllactic acid (**1**), an isostere of tyrosine. However, initial feeding studies revealed that this and other α-hydroxy acids are metabolized by *E. coli*.[12] When **1** was added to the growth media, the compound was undetectable in cell lysates and was depleted in the media. We then followed the metabolism of a related compound, *p*-methoxy-L-phenyllactic acid (**2**), whose products can be distinguished from endogenous metabolites. When **2** was added to growth media, LC–MS analysis of *E. coli* lysates showed that it was transformed into the corresponding amino acid.[12] The bacterial metabolism of α-hydroxy acids is likely to involve initial oxidation to the α-keto acid by an α-hydroxy acid dehydrogenase, followed by transamination to the corresponding amino acid by keto acid transaminases (Figure S1 in the Supporting Information).[12] Although there are many dehydrogenases in *E. coli* capable of oxidizing α-hydroxy acids, it is not clear which enzymes oxidize **1**. For example, deletion of lactate dehydrogenase (*lldD*) did not lead to accumulation of **1** in the cell lysate.[12] In contrast, transamination of *p*-hydroxyphenylpyruvate to tyrosine is known to involve tyrosine aminotransferase (*tyrB*) and aspartate aminotransferase[13] (*aspC*) in the last step of tyrosine biosynthesis (Figure S1). Therefore, we sequentially disrupted both of these genes in *E. coli* strain DH10B with a phage λ-red recombinase-based gene knockout method[14] resulting in the double knockout strain GWAP01. The growth of GWAP01 cells in minimal media required tyrosine and aspartic acid supplementation (Figure S1). Importantly, addition of 1 mM **1** to the growth medium resulted in greater than 100 μM levels of **1** in cell lysates.

To alter the substrate specificity of *M. jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS) to aminoacylate **1** and not tyrosine, two libraries of *Mj*TyrRS mutants were generated on the basis of an analysis of the X-ray crystal structures of the *M. jannaschii* $tRNA^{Tyr}$-TyrRS and L-tyrosine complex.[15,16] Five residues near the amino group of the tyrosine substrate were randomized in each library: Glu36, Ile137, Tyr151, Gln155, and Gln173 in plasmid pBK-lib-jw1a (3×10^7 in size), and residues Ile137, Tyr151, Gln155, Gln173, and Ile176 in plasmid pBK-lib-jw2a (3×10^7 in size). To identify *Mj*TyrRS mutants selective for **1**, these libraries were subjected to rounds of positive and negative selection as previously reported.[1] Surprisingly, only one round of negative and positive selection of the combined libraries generated a clone that grows at 100 $\mu g\,mL^{-1}$ chloramphenicol in the presence of **1**, but only at 20 $\mu g\,mL^{-1}$ chloramphenicol in the absence of **1**. This clone (PlaRS) has the mutations Gln155Arg, Gln173Gly, and Ile176Val. On the basis of the *Mj*TyrRS structure,[15,16] the Gln155Arg and Gln173Gly mutations likely render the synthetase inactive towards tyrosine by deleting two critical hydrogen-bond acceptors for the α-amino group of the tyrosine substrate. Arg155 may still serve as a hydrogen-bond donor to the α-hydroxy group of **1**.

To determine the efficiency and fidelity for the incorporation of **1** into proteins, an amber stop codon was substituted for Lys99 in sperm whale myoglobin containing a C-terminal His_6 tag. Protein expression was carried out in the presence of the selected synthetase (PlaRS) and $Mj tRNA^{Tyr}_{CUA}$ with 1 mM **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 full-length protein was expressed

[*] J. Guo,[+] J. Wang,[+] Prof. J. C. Anderson, 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
Prof. J. C. Anderson
Department of Bioengineering
University of California, Berkeley
306 Stanley Hall, Berkeley, CA 94720-1762 (USA)
[+] These authors contributed equally to this work.

[**] We thank Bill Webb for help in protein mass spectrometry, and we are grateful to the DOE (ER46051) and the Skaggs Institute for Chemical Biology for support. We thank Dr. M. Sever and Dr. W. Liu for helpful discussion.

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

only in the presence of **1** (Figure 1A), indicating that PlaRS does not utilize tyrosine or other endogenous amino acids to any significant degree. The yield of the mutant myoglobin was 2–3 mg L^{-1}. For comparison, the yield of the myoglobin Lys99Tyr mutant in the presence of the wild type *Mj*TyrRS and *Mj*$tRNA^{Tyr}_{CUA}$ under similar conditions was 5–10 mg L^{-1}. ESI-MS analyses of the mutant myoglobin gave an observed average mass of 18390 Da, which is in close agreement with the calculated mass of 18391 Da for the Lys99→**1** myoglobin mutant. In comparison, the Lys99Tyr mutant has observed and calculated masses of 18389 Da and 18390 Da, respectively. The Lys99→**1** mutant consistently and reproducibly produced an observed mass 1 Da higher than that of the Lys99Tyr mutant, again suggesting that **1** is selectively incorporated.

To confirm the selective incorporation of **1** and not tyrosine into protein, we attempted to hydrolyze the ester backbone linkage selectively under alkaline conditions. Five myoglobin mutant proteins (Ser4→**1**, Ala75→**1**, Lys99→**1**, Tyr104→**1**, and Lys99Tyr) were incubated in 0.67 M NaOH for 20 minutes at 4 °C.[17] As shown in Figure 1, the Lys99Tyr mutant myoglobin was completely intact under these conditions, whereas the Lys99→**1** mutant was selectively and efficiently cleaved into two fragments. ESI-MS analysis showed that the two fragments have molecular masses of 11048 Da and 7360 Da, respectively (Figure 1C). For comparison, the predicted masses of the two fragments after hydrolysis are 11049 Da and 7360 Da, respectively. The other three mutants Ser4→**1**, Ala75→**1**, and Tyr104→**1** were also cleaved selectively and efficiently into two fragments with the correct masses, as shown by both SDS-PAGE and LC-MS analysis (see the Supporting Information). Although some hydrolysis was observed in the SDS-PAGE for the mutants before base treatment, no hydrolysis was observed in LC–MS experiments (Figure 2B). This hydrolysis is therefore likely due to the basic SDS-PAGE buffer (pH 8.8).

Selective hydrolysis of the protein backbone can also be used to remove C-terminal fusion proteins and affinity tags to produce unmodified native proteins. Hydrolysis of an ester-linked C-terminal tag, such as the self-processing fusion tags derived from inteins,[19] should afford an unaltered protein. In contrast, selective cleavage of C-terminal tags with proteases typically requires the addition of extra amino acids at the C terminus and/or the absence of other cleavage sites in the protein.[19] To demonstrate the selective cleavage of a C-terminal His_6 tag, a Ser63→**1** mutant Z-domain protein containing a C-terminal His_6 tag immediately after the TAG63 site was expressed in the presence of the mutant synthetase (PlaRS), *Mj*$tRNA^{Tyr}_{CUA}$, and 1 mM **1** in GWAP01. After affinity purification with Ni–NTA (NTA = nitrilotriacetic acid), Z-domain protein was incubated at pH 9 at 4 °C for 12 hours, followed by dialysis against 20 mM phosphate buffer (pH 7.3). SDS-PAGE (Figure 2A) and ESI-MS (Figure 2B) indicated complete conversion to Z-domain protein without the His_6 tag. Circular dichroism measurements confirmed that the Z-domain protein was correctly folded (see the Supporting Information). The full-length Z-domain protein has an observed molecular mass of 6974 Da (without the N-terminal methionine residue, Figure 2B). For compar-

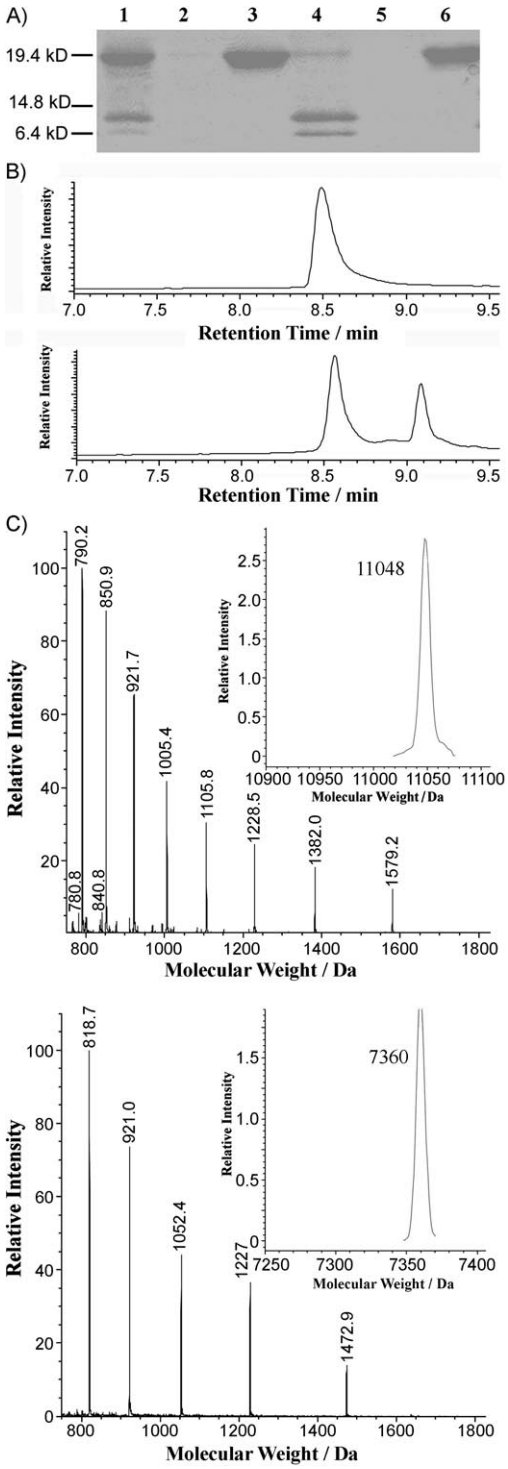

Figure 1. A) Coomassie-stained SDS-PAGE of Lys99→**1** mutant myoglobin expressed in the presence (lane 1, full-length protein and cleaved fragments owing to the basic SDS-PAGE buffer) and absence (lane 2) of 1 mM **1**. Lane 3 shows the myoglobin Lys99Tyr mutant. The same samples as in lanes 1, 2, and 3 were treated with 0.67 M NaOH for 20 min at 4 °C, neutralized to pH 7.0, and analyzed in lanes 4 (cleaved fragments), 5, and 6, respectively. B) LC-ESI spectra of the Lys99→**1** mutant myoglobin before (top panel) and after (bottom panel) base hydrolysis. C) ESI-MS spectra of the two fragments of Lys99→**1** mutant myoglobin after hydrolysis. The insert shows the deconvoluted spectra. Expected masses of the two fragments are 11049 and 7360 Da; observed masses are 11048 and 7360 Da (with N-terminal methionine).

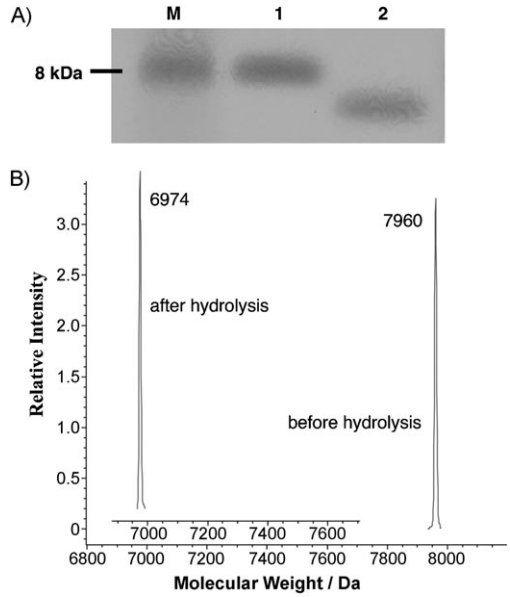

Figure 2. Efficient C-terminal His_6 tag cleavage of Ser63→**1** mutant Z-domain protein. A) SDS-PAGE gel of the Ser63→**1** mutant Z-domain protein. Molecular weight marker, lane M; Z-domain protein with C-terminal His_6 tag before base hydrolysis, lane 1; Z-domain protein with C-terminal His_6 tag after base hydrolysis, lane 2. B) ESI-MS spectra of Ser63→**1** mutant Z-domain protein. Calculated mass before hydrolysis: 7961 Da (without the N-terminal methionine); observed mass 7960 Da. Calculated mass after hydrolysis: 6975 Da (without the N-terminal methionine); observed mass: 6974 Da.

ison, the predicted mass for the full-length Z-domain protein is 6975 Da. It should also be possible to carry out acyl transfer reactions of ester-containing proteins to nucleophiles such as ammonia or alkoxyamine derivatives.

The amide-to-ester mutation can also be used to probe the role of backbone amide groups in catalysis, molecular recognition, and folding. In particular, this substitution is a useful probe of backbone hydrogen-bonding interactions in the formation of protein secondary structures.[9,10] The ester bond, like the amide bond, favors the *trans* conformation and has a significant *cis–trans* rotational barrier.[4,9] However, the ester substitution results in the loss of one hydrogen bond donor and a decrease in the basicity of the carbonyl oxygen atom. To this end, three mutants (Ala75→**1**, Tyr104→**1**, and Ala75Tyr) and wild-type myoglobin were used in unfolding studies to quantify the thermodynamic contribution of individual backbone hydrogen bonds to protein stability. The myoglobin structure consists of eight helices (A to H) connected by short loops and turns (Figure 3 A). The amide NH group of Ala75 is located at the C-terminus of helix E and is hydrogen bonded to the carbonyl group of Thr71 (helix E); the carbonyl group of Ala75 is hydrogen bonded to water. The amide NH group of Tyr104 is located in helix G and is hydrogen bonded to the carbonyl group of Pro101 (helix G); the carbonyl group of Tyr104 is hydrogen bonded to the amide NH group of Ile108 (helix G). To determine the difference in stabilities of the mutants, chemical denaturation of myoglobin mutants was measured by circular dichroism (CD) in the presence of different concentrations of guanidi-

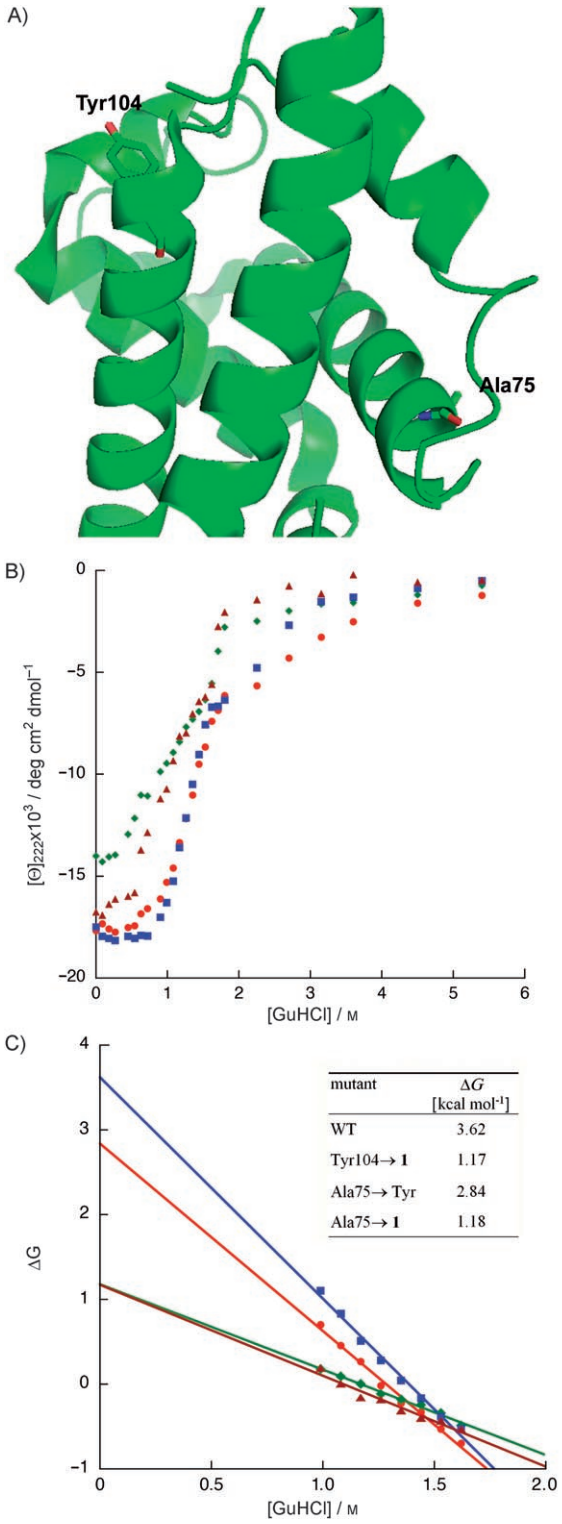

Figure 3. A) Structure of sperm whale myoglobin. Residues Ala75 and Tyr104 are shown as sticks. B) Guanidine hydrochloride (GuHCl) induced denaturation of wild-type and mutant apomyoglobin proteins monitored by circular dichroism. Wild type (■), Tyr104→**1** (▲), Ala75→Tyr (●), Ala75→**1** (◆) samples were assayed in 10 mM sodium phosphate (pH 7.3) with various concentrations of GuHCl as indicated. C) The free-energy change ΔG between folded (F) and unfolded (U) states in various concentrations of GuHCl. The inset shows the calculated unfolding energy of wild-type (WT) and mutant myoglobin at 0 M GuHCl.

nium hydrochloride (GuHCl). As shown in Figure 3, substitution of Ala75 or Tyr104 with **1** decreases the stability of the mutant protein by 1.66 or 2.45 kcal mol^{-1}, respectively, relative to the corresponding tyrosine mutants. In each case, the ester mutation is destabilizing, which is consistent with the fact that an ester is not a hydrogen-bond donor and is a weaker hydrogen-bond acceptor than an amide. The $\Delta\Delta G°(H_2O)$ values for the backbone amide-to-ester substitution determined here can be compared with those reported for T4 lysozyme.[6] For T4 lysozyme, the ester substitution in the C-terminal position of an α-helix perturbed only one hydrogen bond and was destabilizing by 0.9 kcal mol^{-1}. Introduction of the ester linkage in the middle of a helix, which alters two hydrogen-bonding interactions, destabilized the protein by 1.7 kcal mol^{-1}. Consistent with these results, ester substitution of Tyr104, which perturbs two hydrogen-bonding interactions, was more destabilizing than ester substitution of Ala75, which disrupts only one hydrogen-bonding interaction.

In summary, we have shown that α-hydroxy acid **1** can be directly incorporated into proteins in *E. coli* with high fidelity and good efficiency in response to the amber codon TAG. This is a first and significant step towards the expansion of the genetic code of living organisms beyond the realm of L-amino acids.[1] Because of the versatility, low cost, and high yield of recombinant protein expression techniques, it should now be possible to carry out systematic mutation of the protein backbone to investigate its role in protein folding,[4,9] enzyme catalysis,[11] ion-channel gating,[5] and molecular recognition both in vitro and in vivo.[9] Further development in this direction may allow us to incorporate other α-hydroxy acids into proteins to ultimately address the question of whether a "polyester protein" can fold.[17,18]

Experimental Section

p-Hydroxy-L-phenyllactic acid was purchased from Fluorochem. To generate the host strain, the *tyrB* and *aspC* genes were deleted by the methods described by Datsenko and Wanner[14] and the resulting strain (GWAP01) was confirmed by polymerase chain reaction and sequencing of genomic DNA. To characterize the biotransformation of α-hydroxy acids, DH10B or GWAP01 cells were grown in 2YT media with 1 mM *p*-methoxy-L-phenyllactic acid (**2**) overnight at 37 °C until the optical density (OD) was 1.8, and the cells were lysed by sonication. The cell lysate and clarified media were then analyzed by LC-MS for the presence **2** and its corresponding amino acid, *p*-methoxy-L-phenylalanine (**3**); compound **2** is detected in the negative ion mode (mass = 195 Da), and **3** is detected in the positive ion mode (mass = 196 Da).

To cleave selectively the C-terminal His_6 tag of Z-domain protein, strain GWAP01 was cotransformed with plasmids pLeiZ-TAG63 and pBK-PlaRS. Cells were grown in 2YT media (5 mL) supplemented with kanamycin (50 μg mL^{-1}) and chloramphenicol (50 μg mL^{-1}). The 5-mL culture was transferred to 100 mL 2YT with appropriate antibiotics and grown at 37 °C to give an OD_{600} value of 0.6. Gene expression was then induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 1 mM **1**. After 10 h, the cells were harvested by centrifugation and lysed by sonication. The Z-domain protein was purified by Ni–NTA affinity chromatography under native conditions. The purified protein was then incubated at pH 9 at 4 °C for 12 h to remove the His_6 tag. After dialysis against 20 mM phosphate buffer (pH 7.3), the wild-type Z-domain protein was obtained.

To determine the stability of mutant and wild-type myoglobin proteins, they were expressed with a C-terminal His_6 tag and purified to near homogeneity by Ni–NTA column chromatography under denaturing conditions. After dialysis against 20 mM phosphate buffer (pH 7.3), circular-dichroism measurements confirmed that the apo-myoglobin protein was correctly folded. Circular dichroism measurements were performed on an AVIV stopped-flow circular-dichroism spectrophotometer (Model 202SF) with 0.1-mm cuvettes for solutions containing 5 μM myoglobin protein. Samples were incubated with different concentrations of guanidinium hydrochloride (GuHCl) in 10 mM phosphate buffer (pH 7.3) at 25 °C for 30 min before measurements. The free-energy change ΔG between the fully folded (F) and unfolded (U) states was determined by using the equation: $\Delta G = -RT\ln K = -RT\ln[(y_F - y)/(y - y_U)]$, in which R is the universal gas constant, T is the absolute temperature, y_F is the molar ellipticity of myoglobin in the fully folded state, y_U is the molar ellipticity in the fully unfolded state, and y is the molar ellipticity at various concentrations of GuHCl. To determine the unfolding energy in the absence of GuHCl ($\Delta G°(H_2O)$), a linear least-squares analysis was performed by using the equation: $\Delta G = \Delta G°(H_2O) - m[D]$, in which m is a constant, and [D] is the concentration of GuHCl.

Received: September 4, 2007
Revised: November 11, 2007
Published online: December 11, 2007

Keywords: α-hydroxy acids · backbone mutations · biosynthesis · proteins

[1] L. Wang, J. Xie, P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225.
[2] J. M. Xie, P. G. Schultz, *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775.
[3] D. Mendel, V. W. Cornish, P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 435.
[4] J. A. Ellman, D. Mendel, P. G. Schultz, *Science* **1992**, *255*, 197.
[5] T. Lu, A. Y. Ting, J. Mainland, L. Y. Jan, P. G. Schultz, J. Yang, *Nat. Neurosci.* **2001**, *4*, 239.
[6] J. T. Koh, V. W. Cornish, P. G. Schultz, *Biochemistry* **1997**, *36*, 11314.
[7] H. H. Chung, D. R. Benson, P. G. Schultz, *Science* **1993**, *259*, 806.
[8] P. E. Dawson, T. W. Muir, I. Clarklewis, S. B. H. Kent, *Science* **1994**, *266*, 776.
[9] E. T. Powers, S. Deechongkit, J. W. Kelly, *Adv. Protein Chem.* **2006**, *72*, 39.
[10] S. Deechongkit, H. Nguyen, E. T. Powers, P. E. Dawson, M. Gruebele, J. W. Kelly, *Nature* **2004**, *430*, 101.
[11] X. Y. Yang, M. Wang, M. C. Fitzgerald, *Bioorg. Chem.* **2004**, *32*, 438.
[12] J. C. Anderson, Towards polyester proteins: biotransformation of alpha-hydroxy acids, Ph.D. thesis, The Scripps Research Institute, **2003**.
[13] I. G. Fotheringham, S. A. Dacey, P. P. Taylor, T. J. Smith, M. G. Hunter, M. E. Finlay, S. B. Primrose, D. M. Parker, R. M. Edwards, *Biochem. J.* **1986**, *234*, 593.
[14] K. A. Datsenko, B. L. Wanner, *Proc. Natl. Acad. Sci. USA Proc. Natl. Acad. Sci.* **2000**, *97*, 6640.
[15] T. Kobayashi, O. Nureki, R. Ishitani, A. Yaremchuk, M. Tukalo, S. Cusack, K. Sakamoto, S. Yokoyama, *Nat. Struct. Mol. Biol.* **2003**, *10*, 425.
[16] Y. Zhang, L. Wang, P. G. Schultz, I. A. Wilson, *Protein Sci.* **2005**, *14*, 1340.
[17] S. Fahnestock, A. Rich, *Science* **1971**, *173*, 340.
[18] A. L. Weber, S. L. Miller, *J. Mol. Evol.* **1981**, *17*, 273.
[19] D. S. Waugh, *Trends Biotechnol.* **2005**, *23*, 316.

