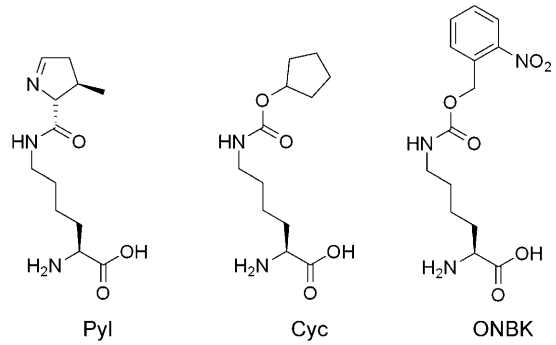
Expanded Genetic Code

DOI: 10.1002/anie.200900683

A Facile System for Encoding Unnatural Amino Acids in Mammalian Cells**

Peng R. Chen, Dan Groff, Jiantao Guo, Weijia Ou, Susan Cellitti, Bernhard H. Geierstanger, and Peter G. Schultz**

We have shown that additional amino acids, beyond the canonical twenty, can be added to the genetic codes of both prokaryotic and eukaryotic organisms.[1] This is accomplished by means of an orthogonal tRNA and aminoacyl-tRNA synthetase (aaRS) pair that incorporates the unnatural amino acid in response to a nonsense or four-base codon in the gene of interest. Directed evolution of the specificity of the aminoacyl-tRNA synthetase in either bacteria or yeast has been used to genetically encode approximately 50 unnatural amino acids with novel physical, chemical, or biological properties in these organisms.[2] One can also use an aaRS evolved in *S. cerevisiae* in conjunction with an amber suppressor tRNA from *B. stearothermophilus* (which is expressed at high levels) to incorporate unnatural amino acids in mammalian cells.[3] However, it is not currently possible to export the large number of aminoacyl-tRNA synthetases evolved in *E. coli* to mammalian cells because the *M. jannaschii* derived aminoacyl-tRNA synthetases typically used in *E. coli* are not orthogonal in mammalian cells. To overcome this limitation, we turned to a pyrrolysyl-tRNA synthetase (PylRS) and its cognate $tRNA^{Pyl}_{CUA}$, which naturally incorporates pyrrolysine (Pyl) (Scheme 1) in response to the amber nonsense codon in the archaea *Methanosarcina maize*.[4–6] Previous work has shown that $tRNA^{Pyl}_{CUA}$ is not recognized by endogenous aaRSs in *E. coli* and mammalian cells as a result of its unique structural features.[7,8] Moreover, the Yokoyama group has recently taken advantage of the known promiscuity of the natural *Methanosarcina maize* PylRS (MmPylRS)[9] to incorporate Pyl analogues into proteins in mammalian cells. In addition Chin and co-workers used a mutant *Methanosarcina barkeri* PylRS (MbPylRS), a close homologue of MmPylRS, to incorporate acetyl lysine in *E. coli*, demonstrating that the specificity of the PylRS can be altered by directed evolution methods.[10–12] Thus, this system offers the potential to evolve new PylRS specificities in *E. coli*, a host in which large libraries of mutant aminoacyl-tRNA synthetases can be generated and selected, and subsequently shuttle the evolved aaRSs directly into mammalian cells. Herein, we demonstrate the utility of such an *E. coli*–mammalian "shuttle" system by genetically encoding a photocaged lysine in both bacterial and mammalian cells.



Scheme 1. The structures of pyrrolysine (Pyl), the pyrrolysine analogue N^{ε}-cyclopentyloxycarbonyl-L-lysine (Cyc), and the photocaged lysine *o*-nitrobenzyl-oxycarbonyl-N^{ε}-L-lysine (ONBK).

First we confirmed the orthogonality of the *M. maize* pyrrolysyl-tRNA synthetase (MmPylRS)/$tRNA^{Pyl}_{CUA}$ pair in both *E. coli* and mammalian cells, which is the key requirement for establishing a robust system for shuttling tRNA/aaRS pairs between these two hosts. Northern blot analysis detected aminoacylated tRNAs only when *E. coli* cells harbored plasmids encoding both $tRNA^{Pyl}_{CUA}$ and MmPylRS and were supplemented with 5 mM of the Pyl analogue N^{ε}-cyclopentyloxycarbonyl-L-lysine (Cyc; Figure 1a). Aminoacylation of $tRNA^{Pyl}_{CUA}$ does not occur in the absence of Cyc or of the plasmid encoding MmPylRS, indicating that $tRNA^{Pyl}_{CUA}$ is not a substrate for endogenous aaRSs in *E. coli* and that MmPylRS does not recognize endogenous amino acids in *E. coli*. Western blot analysis of samples from CHO cells shows that a C-terminal His-tagged retinol binding protein 4 (RBP4) with an amber mutation at Phe36 (RBP4/Phe36TAG) was only expressed in the presence of MmPylRS, $tRNA^{Pyl}_{CUA}$, and 5 mM Cyc (Figure 1b). Again, these results verify that $tRNA^{Pyl}_{CUA}$ is not a substrate for endogenous aaRSs and that MmPylRS does not recognize endogenous amino acids in mammalian cells. These data confirm that MmPylRS/$tRNA^{Pyl}_{CUA}$ works as a functional amber suppressor pair in both *E. coli* and mammalian cells with the substrate Cyc, which is

[*] Dr. W. Ou, Dr. S. Cellitti, Dr. B. H. Geierstanger, Dr. P. G. Schultz
Genomics Institute of the Novartis Research Foundation
10675 John Jay Hopkins Drive, San Diego, CA 92121 (USA)
Fax: (+1) 858-812-1746
E-mail: bgeierstanger@gnf.org
Dr. P. R. Chen,[+] D. Groff,[+] Dr. J. Guo, Dr. P. G. Schultz
Department of Chemistry and the Skaggs Institute for Chemical Biology, The Scripps Research Institute
10550 Torry Pines Road, La Jolla, CA 92037 (USA)
Fax: (+1) 858-784-9440
E-mail: schultz@scripps.edu
[+] These authors contributed equally to this work.

[**] This work was supported by a grant from the U.S. National Institute of Health R01 GM062159 and the Skaggs Institute for Chemical Biology. This is manuscript 19983 of The Scripps Research Institute.
Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200900683.

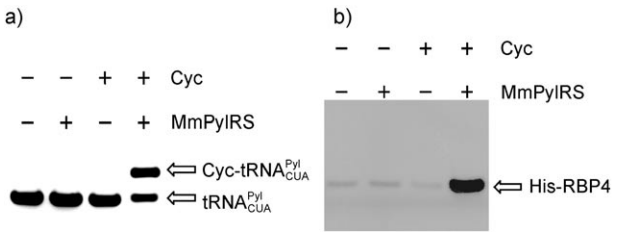


Figure 1. a) Northern blot analysis of tRNA charging in *E. coli*. The uncharged tRNA$^{Pyl}_{CUA}$ band and the charged tRNA$^{Pyl}_{CUA}$ band are indicated by arrows. tRNA$^{Pyl}_{CUA}$ is charged only in the presence of both PylRS and Cyc. b) Western blot analysis of protein expression in mammalian cells. The full-length mutant His-RBP4 is expressed only when CHO cells harboring both MmPylRS and tRNA$^{Pyl}_{CUA}$ plasmids were grown with 5 mM Cyc.

consistent with previous results obtained for MbPylRS and MmPylRS, respectively.[10,11]

We next created a library of MmPylRS active-site mutants in order to alter the amino acid specificity of this enzyme. On the basis of the crystal structure of MmPylRS bound to Pyl,[13] five residues (Leu305, Tyr306, Leu309, Cys348, and Tyr384) surrounding the methyl pyrroline ring of Pyl were randomized to expand the Pyl recognition pocket (Figure S1 in the Supporting Information). Overlap extension polymerase chain reaction was performed with synthetic oligonucleotide primers in which the randomized residues were encoded as NNK (N = A, C, T, or G, K = T or G) to generate a library with a diversity of 3×10^7, the quality of which was validated by sequencing.

We then evolved a mutant MmPylRS/tRNA$^{Pyl}_{CUA}$ pair specific for the N^{ε}-photocaged lysine analogue, *o*-nitrobenzyloxycarbonyl-N^{ε}-L-lysine (ONBK, Scheme 1 and Scheme S1 in the Supporting Information) in *E. coli*. Photocaging, in which a molecule is derivatized with a photoremovable inactivating group, is widely used as a noninvasive tool for spatial and temporal control of a variety of complex cellular processes.[14–19] We have previously genetically encoded photocaged Ser, Cys, and Tyr residues.[14,15,20] A photocaged lysine would, for example, allow photoactivation of ubiquitination, methylation, and acetylation in mammalian cells, and as a result could be used to activate protein degradation or modulate transcription. In order to identify MmPylRS mutants that can selectively aminoacylate tRNA$^{Pyl}_{CUA}$ with ONBK, a series of positive and negative selections were performed as previously described.[21,22] In brief, the positive selection is based on resistance to chloramphenicol (Cm), which is conferred by the suppression of an amber mutation at a permissive site (Asp112) in the type I chloramphenicol acetyltransferase gene (CAT112TAG) in the presence of the unnatural amino acid and the aaRS mutant. The negative selection uses the toxic barnase gene with amber mutations at permissive sites (Gln2TAG, Asp44TAG, and Gly65TAG) and was carried out in the absence of unnatural amino acid. Single MmPylRS mutant clones that passed through the selection (three positive and two negative rounds) and survived on Cm only in the presence of ONBK were obtained: 60% of the sequenced clones converged on a unique sequence (referred to as NBK-1) with the mutations Y306M, L309A, C348A, Y384F, while the other 40% converged to a second related sequence (referred to as NBK-2) with the mutations Y306I, L309A, C348A, Y384F. *E. coli* cotransformed with either NBK-1 or NBK-2, and CAT112TAG exhibited a significant difference in growth on Cm in the presence and absence of 1 mM ONBK (Figure 2a), suggesting that these evolved MmPylRS/tRNA$^{Pyl}_{CUA}$ pairs are selective for ONBK relative to endogenous host amino acids. NBK-1 exhibited enhanced amber suppression relative to NBK-2, and thus the NBK-1/tRNA$^{Pyl}_{CUA}$ pair was used for further studies.

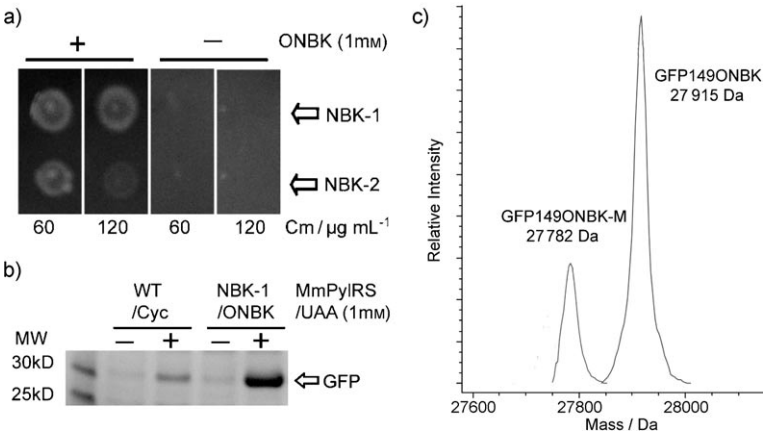


Figure 2. Evolution of a MmPylRS/tRNA$^{Pyl}_{CUA}$ pair that encodes ONBK in *E. coli*. a) Plate assay showing that NBK-1 and NBK-2 are able to survive up to 120 μg mL^{-1} Cm challenges when supplemented with 1 mM ONBK. b) Genetic incorporation of ONBK into GFP protein in *E. coli* analyzed by SDS-PAGE. The expressed full-length GFP proteins were purified by Ni^{2+}-NTA chromatography and stained with Coomassie blue. c) ESI-MS analysis of purified GFP149ONBK protein produced by NBK-1/tRNA$^{Pyl}_{CUA}$. The major peak (mass: 27915 Da) corresponds to the full-length GFP149ONBK; the minor peak (mass: 27782 Da) corresponds to the same protein with the N-terminal Met posttranslationally cleaved (GFP149ONBK-M).

To determine the efficiency and fidelity of ONBK incorporation into proteins in *E. coli*, an amber mutation (TAG) was introduced for Asp149 in a C-terminal His-tagged variant of GFP (GFP149TAG). A vector pSup-NBK-1 was constructed to encode the NBK-1/tRNA$^{Pyl}_{CUA}$ pair in which a single copy of the tRNA$^{Pyl}_{CUA}$ gene is expressed under control of the *proK* promoter and terminator, and the NBK-1 gene is expressed under control of a mutant *glnS* (*glnS'*) promoter.[23] This plasmid was cotransformed into BL21-DE3 *E. coli* cells with a plasmid carrying the GFP149TAG gene (pBAD-GFP149TAG). Protein expression was carried out in LB medium supplemented with and without 1 mM ONBK, followed by purification with Ni^{2+}-NTA affinity chromatography. SDS-PAGE analysis and subsequent Coomassie staining showed that full-length protein was produced only in the presence of ONBK (Figure 2b). Expression for 8 hours at

30 °C with NBK-1 and $tRNA^{Pyl}_{CUA}$ yielded around 10 mg L^{-1} protein in medium containing 1 mM ONBK. As a control, a plasmid containing the wild-type MmPylRS/$tRNA^{Pyl}_{CUA}$ pair was employed for expression of GFP149TAG in the presence of 1 mM Cyc (Figure 2b) and the protein yield was less than 1 mg L^{-1}.

Electrospray ionization mass spectrometry (ESI-MS) of purified GFP protein with ONBK at position 149 revealed two peaks (27915 Da and 27782 Da) corresponding to GFP protein containing the intact ONBK residue with and without the N-terminal Met (Figure 2c). This result confirms the high specificity of the NBK-1 mutant aminoacyl-tRNA synthetase for ONBK relative to endogenous amino acids, and for $tRNA^{Pyl}_{CUA}$ relative to endogenous tRNAs. At longer induction times, we also observed peaks for the intact protein with lysine at position 149. We suspect that the ONBK photocaging group is partially removed by degradative enzymes in *E. coli* (vide infra).

Next, the evolved NBK-1/$tRNA^{Pyl}_{CUA}$ pair from *E. coli* was shuttled into mammalian cells. A vector pCMV-NBK-1 was constructed containing the NBK-1 gene under control of a nonregulated CMV promoter, and a single $tRNA^{Pyl}_{CUA}$ gene under control of a human U6 promoter. Amber suppression was monitored using an enhanced GFP (EGFP) with an amber mutation at the permissive residue 37 (EGFP37TAG). The plasmid pCMV-NBK-1 was cotransfected with a plasmid encoding EGFP37TAG into HEK293 cells using an optimized transfection protocol. After induction, the cells were allowed to grow in the presence and absence of 1 mM ONBK for 36 h before being visualized under a fluorescence microscope (Figure 3a). Full-length EGFP was detected only in cells supplemented with 1 mM ONBK, while no EGFP was observed otherwise.

The incorporation of ONBK in mammalian cells in response to an amber codon was further confirmed by mass spectrometry. After purification by Ni^{2+}-NTA chromatography, 35 μg EGFP protein was isolated from 4×10^7 CHO cells and analyzed by ESI-MS (Figure 3b). Only one peak was observed corresponding to the full-length protein containing the intact ONBK residue (EGFP37ONBK), indicating that no loss of the photocaging group occurred in mammalian cells. In addition, this result shows that the mutant MmPylRS does not load endogenous tRNAs with ONBK to give heterogeneous protein product. To verify the presence of the intact photocaged Lys, purified EGFP37ONBK was irradiated with 365 nm light for 20 minutes. ESI-MS analysis of this protein sample revealed one peak with a change in mass corresponding to the loss of one *o*-nitrobenzyloxycarbonyl group (Figure 3c), indicating that EGFP37ONBK was cleanly converted into EGFP37K with near-visible light.

In summary, we have developed a straightforward strategy for the expansion of the amino acid repertoire of mammalian cells with the PylRS/$tRNA^{Pyl}_{CUA}$ pair from archaea. We demonstrated the utility of this approach by genetically encoding a photocaged lysine which is likely to be a useful probe of protein function in bacterial and mammalian cells. Moreover, the X-ray crystal structure[13] of the PylRS active site suggests that this "shuttle" system can also be used for the directed evolution of additional aaRSs specific for other unnatural amino acids for use in both prokaryotic and eukaryotic organisms.

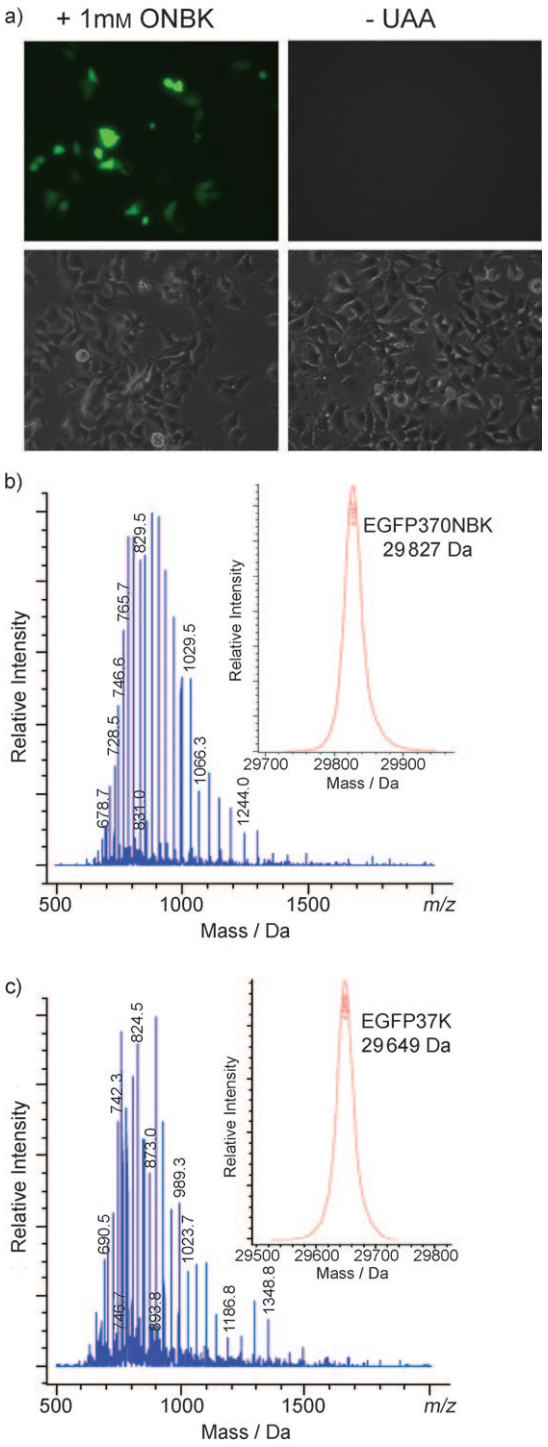


Figure 3. Shuttling the evolved synthetase into mammalian cells. a) Expression of EGFP37TAG protein using the NBK-1/$tRNA^{Pyl}_{CUA}$ pair in HEK293 cells in the presence of 1 mM ONBK. The top pictures show the fluorescence images of cells and the bottom pictures show cells illuminated with visible light. b) ESI-MS analysis of purified EGFP37ONBK protein from CHO cells. Inset shows the deconvoluted spectrum of EGFP37ONBK. c) ESI-MS analysis of EGFP37ONBK after photolysis. EGFP37ONBK protein at a final concentration of 100 μM was irradiated (365 nm) for 20 min.

Experimental Section

For Northern blot analysis, RNA samples isolated from *E. coli* cells were separated by acid–urea gel electrophoresis and electroblotted onto a Hybond N^+ membrane in 0.5 × TBE (Tris/borate/EDTA) running buffer at 30 V constant for 1 h using the Xcell II Blot Module (Invitrogen). The Chemiluminescent Nucleic Acid Detection Module (Pierce) was used with a 72-base oligonecleotide complementary to $tRNA^{Pyl}_{CUA}$ as the probe. For Western blot analysis, cells were detached and lysed in RIPA (RadioImmunoPrecepitaton Assay) buffer (Upstate) with protease inhibitor cocktail (Roche). The supernatant of cell lysate was fractionated by SDS-PAGE and transferred to 0.45 μm nitrocellulose membrane (Invitrogen). The proteins on the membrane were probed with anti-His-HRP followed by detection of the luminescence with the ECL Western blotting substrate (Pierce).

To acquire mass spectra of the intact proteins, the purified proteins were dialyzed against Tris buffer (20 mM, pH 7.3) and concentrated to $\approx 0.1\ mg\,mL^{-1}$. The mass spectra were acquired on an automated LC/MS system (Agilent). The dialyzed protein sample ($0.1\ mg\,mL^{-1}$) was loaded onto a C-8 (Agilent) column for desalting with 0.1% trifluoroacetic acid (TFA) in water and eluted with 80% acetonitrile/0.1% TFA into the ESI source of the mass spectrometer.

Photolysis of all purified proteins containing ONBK residues was carried in Tris buffer solution (40 mM Tris, pH 8.0, 100 mM NaCl, and 1 mM 1,4-dithiothreitol). Protein samples with a final concentration of 100 μM were irradiated with a high-pressure mercury lamp (500 W, Spectra Physics) equipped with 310 nm long-pass optical filter.

Other materials and methods can be found in the Supporting Information

Received: February 4, 2009
Published online: April 17, 2009

Keywords: gene expression · nonnatural amino acids · photocaged lysine · tRNA

[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] W. S. Liu, A. Brock, S. Chen, S. B. Chen, P. G. Schultz, *Nat. Methods* **2007**, *4*, 239.
[4] S. K. Blight, R. C. Larue, A. Mahapatra, D. G. Longstaff, E. Chang, G. Zhao, P. T. Kang, K. B. Church-Church, M. K. Chan, J. A. Krzycki, *Nature* **2004**, *431*, 333.
[5] G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, *296*, 1459.
[6] J. A. Krzycki, *Curr. Opin. Microbiol.* **2005**, *8*, 706.
[7] C. Polycarpo, A. Ambrogelly, A. Berube, S. A. M. Winbush, J. A. McCloskey, P. F. Crain, J. L. Wood, D. Soll, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12450.
[8] K. Nozawa, P. O'Donoghue, S. Gundllapalli, Y. Araiso, R. Ishitani, T. Umehara, D. Söll, O. Nureki, *Nature* **2008**, *457*, 1163.
[9] C. R. Polycarpo, S. Herring, A. Berube, J. L. Wood, D. Soll, A. Ambrogelly, *FEBS Lett.* **2006**, *580*, 6695.
[10] T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818.
[11] H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232.
[12] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Chem. Biol.* **2008**, *15*, 1187.
[13] J. M. Kavran, S. Gundliapalli, P. O'Donoghue, M. Englert, D. Soell, T. A. Steitz, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11268.
[14] A. Deiters, D. Groff, Y. H. Ryu, J. M. Xie, P. G. Schultz, *Angew. Chem.* **2006**, *118*, 2794; *Angew. Chem. Int. Ed.* **2006**, *45*, 2728.
[15] E. A. Lemke, D. Summerer, B. H. Geierstanger, S. M. Brittain, P. G. Schultz, *Nat. Chem. Biol.* **2007**, *3*, 769.
[16] D. S. Lawrence, *Curr. Opin. Chem. Biol.* **2005**, *9*, 570.
[17] G. Mayer, A. Heckel, *Angew. Chem.* **2006**, *118*, 5020; *Angew. Chem. Int. Ed.* **2006**, *45*, 4900.
[18] A. P. Pelliccioli, J. Wirz, *Photochem. Photobiol. Sci.* **2002**, *1*, 441.
[19] D. M. Rothman, M. D. Shults, B. Imperiali, *Trends Cell Biol.* **2005**, *15*, 502.
[20] N. Wu, A. Deiters, T. A. Cropp, D. King, P. G. Schultz, *J. Am. Chem. Soc.* **2004**, *126*, 14306.
[21] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498.
[22] J. M. Xie, W. S. Liu, P. G. Schultz, *Angew. Chem.* **2007**, *119*, 9399; *Angew. Chem. Int. Ed.* **2007**, *46*, 9239.
[23] Y. H. Ryu, P. G. Schultz, *Nat. Methods* **2006**, *3*, 263.

