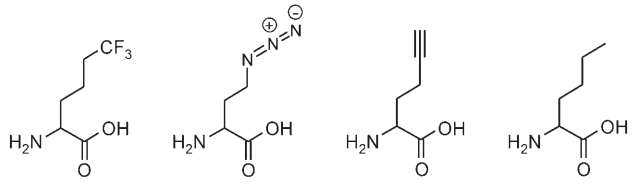
VIP Protein Engineering

DOI: 10.1002/anie.200700779

High-Throughput Screening for Methionyl-tRNA Synthetases That Enable Residue-Specific Incorporation of Noncanonical Amino Acids into Recombinant Proteins in Bacterial Cells**

*Tae Hyeon Yoo and David A. Tirrell**

Aminoacyl-tRNA synthetases (aaRS) with altered substrate specificities have been used to enable both site-specific and residue-specific incorporation of noncanonical amino acids into recombinant proteins.[1–4] Rational,[1a,2] computational,[1b] and combinatorial[1d,3] methods have been employed to engineer the amino acid binding pockets of several aaRS. Combinatorial strategies have been especially effective;[1d,3] Schultz and co-workers have developed powerful methods for selecting aaRS for site-specific incorporation,[3] and we have reported an efficient screening system for use in the global replacement of amino acids.[1d] However, because the latter method relies on bio-orthogonal derivatization of noncanonical amino acid side chains,[1d] a new approach is needed for the more general problem of activating noncanonical substrates that lack reactive functionality in the side chains. Here we describe a high-throughput method for screening aaRS libraries for the global incorporation of noncanonical amino acids. We demonstrate this strategy by identifying an *Escherichia coli* methionyl-tRNA synthetase (MetRS) variant that activates 6,6,6-trifluoronorleucine (Tfn, **1**; Scheme 1). Tfn does not support significant protein synthesis in conventional *E. coli* expression strains.[5a]



1 **2** **3** **4**

Scheme 1. Noncanonical amino acids used in this study. **1**: Tfn; **2**: Aha; **3**: Hpg; **4**: norleucine.

[*] T. H. Yoo, Prof. D. A. Tirrell
Division of Chemistry and Chemical Engineering
Joseph J. Jacobs Institute for Molecular Engineering for Medicine
California Institute of Technology
1200 E. California Blvd., Pasadena, CA 91125 (USA)
Fax: (+1) 626-568-8824
E-mail: tirrell@caltech.edu

[**] We thank A. James Link, Kimberly Beatty, and James Van Deventer for helpful discussions. Knocking out of the *metE* gene from *E. coli* strain DH10B was done with James Van Deventer. We thank Dr. Mona Shahgholi for assistance with the MALDI-MS and LC-MS analysis. This work was supported by NIH grant GM62523, ONR grant N00014-03-1-0793, and a Samsung Scholarship (to T.H.Y.).

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

Unless there are barriers to protein synthesis that lie downstream of the aminoacylation step, the activity of aaRS toward amino acid analogues can be monitored by translation of model proteins in media depleted of the corresponding canonical amino acids.[1d] The green fluorescent protein (GFP) is especially useful in experiments of this kind, in that its synthesis enables rapid screening of mutants by fluorescence-activated cell sorting (FACS).[6] However, global incorporation of noncanonical amino acids into recombinant proteins can cause misfolding and loss of function.[7] In particular, we found that global replacement of Met by various analogues led to substantial reductions in GFP fluorescence.[8] To render GFP fluorescence insensitive to Met replacement, we removed all of the Met codons in the β-barrel structure of GFP.

Starting from the previously described L024_3-3 variant of GFP[7d] (Table S1 in the Supporting Information), the codons for M78, M88, M218, and M233 were randomized by assembly PCR[9] using primers bearing NNH codons (N=A/T/G/C; H = A/T/C; NNH codons do not encode Met or Trp). The PCR fragment was inserted between the *Bam*HI and *Hind*III sites of pQE-80L and the resulting plasmid was transformed into electrocompetent *E. coli* DH10B cells, thus yielding more than 10^7 transformants. Following GFP expression in 2xYT media with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 h and isolation of highly fluorescent cells by FACS, we identified a GFP variant (GFPrm; M78L, M88F, M218A, M233I; Figure 1a) that exhibited fluorescence comparable to that of L024_3-3 (Figure 1b).

As GFPrm retains only two Met residues (Table S1 in the Supporting Information), its expression does not provide a sensitive measure of the MetRS activity of the host (data not shown). To introduce additional Met codons, seven positions located in GFP loops were randomized with NNK primers (K = G/T), and the fluorescence of cells expressing each library was measured by flow cytometry (Figure S1 in the Supporting information). Five positions (D117, K131, Q157, E172, and K214), where the effects of randomization were insignificant, were changed to Met (Figure 1a). Cells expressing the resulting GFP variant (GFPrm_AM) were as bright as those expressing GFPrm (Figure 1b). Three Met analogues known to be activated by wild-type (WT) *E. coli* MetRS, azidohomoalanine (Aha, **2**), homopropargylglycine (Hpg, **3**), and norleucine (**4**),[5] were incorporated into GFPrm_AM. In each case, the observed fluorescence was reduced only modestly, and was easily distinguished from the fluorescence of cells in which GFP expression was induced in Met-free media containing 19 amino acids (Figure 1c). GFPrm_AM

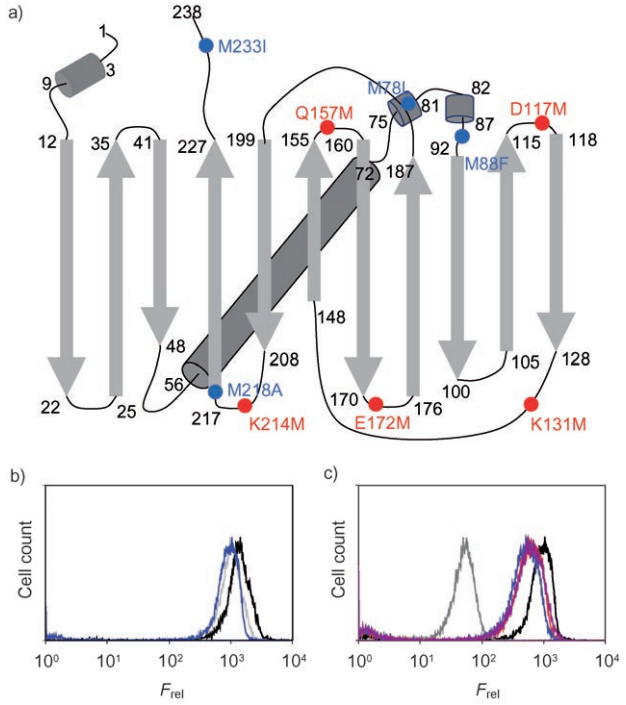


Figure 1. Engineering GFP as a translational reporter. a) Schematic representation of the GFP scaffold. Four Met positions (blue) were changed into other codons (GFPrm), and then five Met codons (red) were introduced into the loop regions of the protein (GFPrm_AM). b) Flow-cytometric analysis of cells expressing GFP (black, L024_3-3[6d]), GFPrm (gray), or GFPrm_AM (blue). c) Cytometric analysis of cells expressing GFPrm_AM with 19 amino acids (gray), with 20 amino acids (black), or with 19 amino acids plus Aha (blue), Hpg (purple), or norleucine (red).

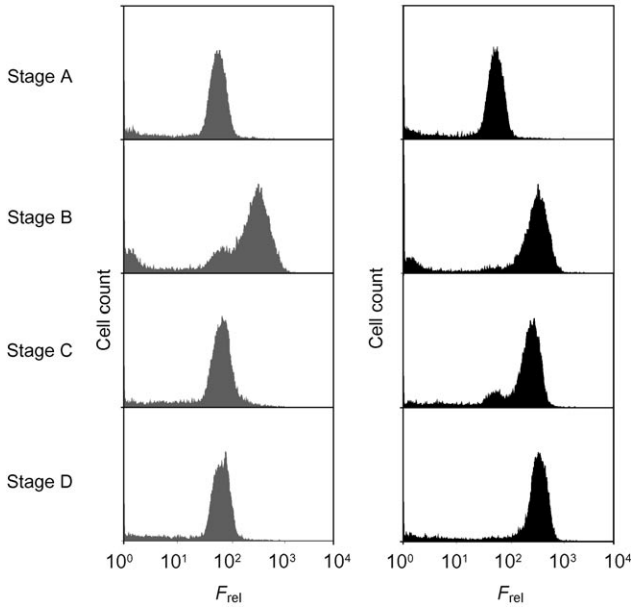


Figure 2. Screening of MetRS for activation of Tfn. Fluorescence distributions of cell populations at stages A, B, C, and D (see text) are shown. Gray plots (left) show populations of cells expressing GFPrm_AM with 19 amino acids; black plots (right) show cells expressing GFPrm_AM with Tfn.

was used as the reporter protein in a FACS-based screen for MetRS variants that activate Tfn.

Four positions (L13, P257, Y260, and H301) in the Met-binding site of *E. coli* MetRS were selected for saturation mutagenesis.[10] A256 was also randomized, because the fluorine atoms of 5,5-difluoromethionine and 5,5,5-trifluoromethionine have been reported to make unfavorable contact with main-chain atoms of A256 and P257.[10] A MetRS library was generated by assembly PCR with primers bearing NNK codons at the randomized positions. The PCR products were inserted into plasmid pMTY11 between the *Not*I and *BsrG*I sites, and the ligation mixture was transformed into electrocompetent *E. coli* DH10B cells, thereby yielding more than 5×10^8 transformants. Plasmid DNA from the library was retransformed into cells of *E. coli* strain DH10B(Met$^-$) bearing the reporter plasmid pQE-80L/GFPrm_AM.

After growth to the mid-log phase, cells harboring the library were suspended in M9 minimal media supplemented either with 19 amino acids or with 19 amino acids plus Tfn. Protein expression was induced with 1 mM IPTG. The fluorescence histograms of the resulting cells are shown in Figure 2 (stage A), and exhibit little sensitivity to the presence of Tfn in the expression medium. The gate in the fluorescence channel was set to recover 0.3% of the most highly fluorescent cells from the Tfn-supplemented medium, and approximately 10^9 cells were sorted to ensure full coverage of the MetRS library. This positive screening step was repeated twice. The majority of the resulting cells were fluorescent when induced in medium supplemented with 19 amino acids plus 4 mM Tfn, but more than 80% were fluorescent when induced with just 19 amino acids (Figure 2, stage B). To remove MetRS variants that activate any of the canonical amino acids (other than Met), the population of cells exhibiting reduced fluorescence with 19 amino acids was collected (negative screening). After three positive screening steps and one negative screening step, the resulting cells expressed with 19 amino acids plus Tfn exhibited higher fluorescence while the fluorescence of cells expressed with 19 amino acids was comparable to that of the initial library (Figure 2, stage C). One more cycle of positive and negative screening steps yielded library M02c_2 (Figure 2b, stage D).

Ten clones were selected at random from M02c_2 and sequenced (Table S2 in the Supplementary Information). Interestingly, nine of those clones had the same change in the amino acid sequence (L13S, Y260L, and H301L) even though they differed at the genetic level. In NNK randomization, there are three codons for Ser (TCT, TCG, and AGT), two for Ala (GCT and GCG), two for Pro (CCG and CCT), and three for Leu (CTT, CTG, and TTG). All of these codons were found in the sequenced clones. The results suggest that the screening method is effective in enriching clones active toward Tfn and removing clones that activate any of the canonical amino acids other than Met. Furthermore, the library is large enough to provide good coverage of the possible sequence variants.

The MetRS gene of clone 8 (M02c_2-8) was transferred into pQE-80/GFPrm_AM to yield expression plasmid

pMTY8. We used pMTY8 for testing protein expression with Tfn because the Col E1 origin of pMTY8 ensures a higher copy number than the P15A origin of pMTY11. In addition, the gene encoding the mutant MetRS gene was transferred into pAJL-61[1d] (which encodes the marker protein murine dihydrofolate reductase (mDHFR)), thus yielding pMTY10. The stop codon that truncates MetRS at position 548 was mutated back to Glu because the C-terminal domain of MetRS has been reported to enhance the tRNA-affinity of the synthetase.[11]

Recombinant marker proteins GFPrm AM and mDHFR (which contain 7 and 8 Met sites, respectively) were expressed in M9 minimal media supplemented with 19 amino acids, 20 amino acids, or 19 amino acids plus Tfn. For strains with the mutant MetRS, the yields of proteins made in Tfn-supplemented media were 20–30 mg L^{-1}, while negligible expression was observed in media supplemented with 19 amino acids (Figure 3 and Table 1). In contrast, for strains bearing WT MetRS, addition of Tfn to Met-depleted media had no effect on protein yield, which is consistent with our previous observation that Tfn is not a good substrate for the wild-type synthetase.[5a] (Met-depleted cultures of the strain bearing WT MetRS afforded trace amounts of GFPrm_AM, irrespective of the addition of Tfn (Figure 3a.) We believe this result to be a consequence of misincorporation of one or more canonical amino acids, as indicated by the reduced electrophoretic mobility of the induced protein band.)

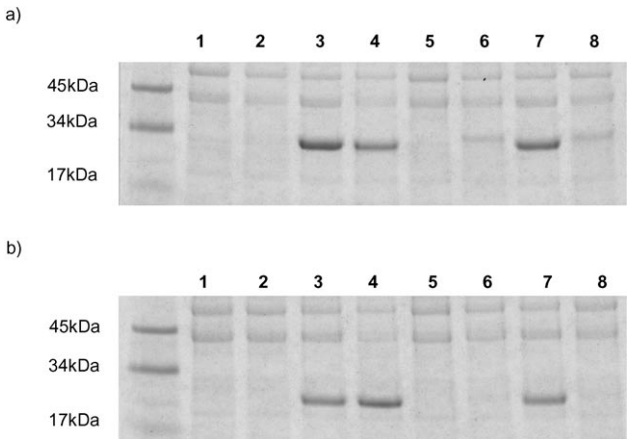


Figure 3. Expression of: a) GFPrm AM and b) mDHFR with mutant MetRS (lanes 1–4) or WT MetRS (lanes 5–8). Lanes 1 and 5: before induction; lanes 2 and 6: induction with 19 amino acids (-Met); lanes 3 and 7: induction with 20 amino acids; lanes 4 and 8: induction with 19 amino acids plus Tfn.

Table 1: Yields of purified GFPrm_AM and mDHFR produced by *E. coli* strains outfitted with the mutant MetRS

	GFPrm_AM		mDHFR	
	Met[a]	Tfn[b]	Met[a]	Tfn[b]
protein yield [mg l^{-1} culture]	150	21	20	31

[a] Protein expression in M9 minimal medium supplemented with 20 amino acids. [b] Protein expression in M9 minimal medium supplemented with 19 amino acids (-Met) and 2 mM Tfn.

Incorporation of Tfn into GFPrm_AM and mDHFR was confirmed by matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) and liquid chromatography/mass spectrometry (LC-MS). Tryptic peptides containing Met were barely detectable by either method (Figure 4; Figures S2 and S3 in the Supporting Information), which is consistent with near-quantitative replacement of Met by Tfn. The fact that GFPrm_AM containing Tfn is highly emissive indicates that the fluorinated amino acid is tolerated without loss of protein function. We have not evaluated the folding behavior or catalytic properties of fluorinated mDHFR.

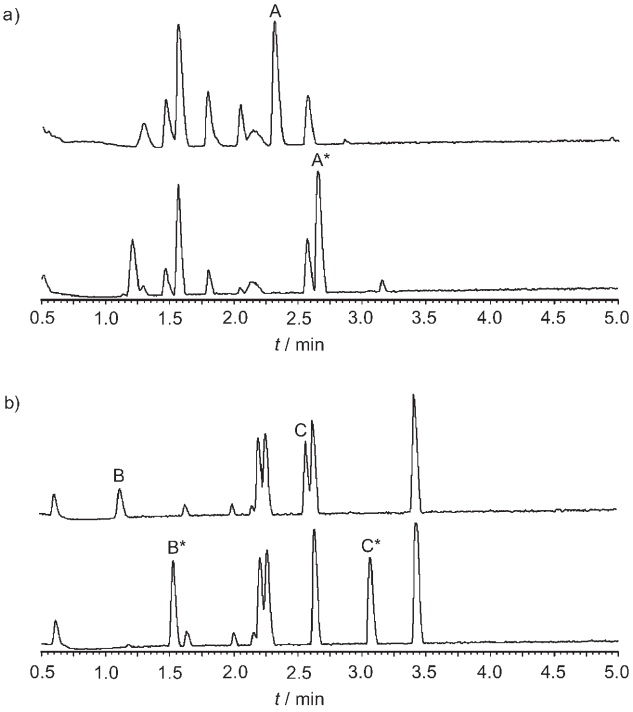


Figure 4. LC-MS chromatograms of tryptic digests of a) GFPrm_AM and b) mDHFR. Each protein was produced with 20 amino acids (top) and with 19 amino acids (-Met) plus Tfn (bottom). The signal corresponding to peptide A (t_{ret} = 2.31 min, sequence FEGMTIVNR) for GFPrm_AM was shifted to 2.66 min (peak A*) by replacement of Met by Tfn. For mDHFR, signals corresponding to peptides B (t_{ret} = 1.12 min, sequence MTTTSSVEGK) and C (t_{ret} = 2.57 min, sequence QNLVIMGR) were shifted to 1.53 min and 2.57 min, respectively, upon amino acid replacement.

The kinetics of activation of Tfn and Met by the mutant MetRS were analyzed by the ATP/inorganic pyrophosphate (PPi) exchange assay.[12] The specificity constant (k_{cat}/K_m) for Tfn was determined to be 2.9×10^{-4} $\mu M^{-1} S^{-1}$ (Table S3 in the Supporting Information), roughly sevenfold lower than that for Met and fully adequate to enable near-quantitative replacement of Met by Tfn in recombinant proteins expressed in Met-depleted media.

The screening method described here is simple, efficient, and directly applicable to a wide variety of Met analogues.

Experimental Section

Materials: All restriction enzymes were purchased from New England Biolabs (Beverly, MA). 6,6,6-Trifluoronorleucine (Tfn) was purchased from Oakwood Products (West Columbia, SC). DNA oligomers were synthesized at Qiagen (Valencia, CA).

Strain construction: The *metE* gene of *E. coli* DH10B was knocked out by the method of Datsenko and Wanner[13] to make the strain auxotrophic for Met.

Plasmid pQE-80L/GFP (pQE-80L/L024_3-3) was described previously.[7d] Plasmid pMTY11 was constructed by ligation between a *Nhe*I/*Sal*I fragment of pREP4 (containing the p15A origin and Kn^r gene; Qiagen) and a MetRS expression cassette (containing its promoter and terminator) amplified from plasmid pAJL-20[1d] using the PCR primers MetRS_NheI_F and MetRS_Sal_R described in the Supporting Information. Plasmid pMY8 was constructed by ligation between *Nhe*I-digested pQE-80/GFPrmAM and the MetRS gene amplified from clone M02c_2-8 using the PCR primers MetRS_NheI_F and MetRS_
NheI_R. The stop codon truncating MetRS at position 548 was mutated back to Glu by site-directed mutagenesis. Plasmid pMTY10 was constructed by ligation between an *Nhe*I fragment of pAJL-61[1d] (encoding a 6×His-tagged mDHFR) and the mutant MetRS gene (from clone M02c_2-8) excised with *Nhe*I from pMTY8. The MetRS genes of pMTY8 and pMTY10 were replaced with that of full-length WT MetRS to yield pMTY13 and pMTY14.

Screening of MetRS mutants: Cells harboring the MetRS library were grown at 37 °C to the mid-log phase (OD_{600} = 0.9–1.0) in M9 minimal medium (M9 salts, 0.2 % glucose, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 25 mg L^{-1} thiamine) supplemented with 40 mg L^{-1} of each of the 20 canonical amino acids, and then washed twice with cold 0.9 % NaCl. The cell pellet was resuspended in minimal media supplemented either with 19 amino acids (40 mg L^{-1} each) or with 19 amino acids plus 4 mM Tfn. After 15 min, protein expression was induced with 1 mM IPTG. After 3 h, the cells were washed and resuspended in 10 mM phosphate buffer (pH 7.4) containing 100 mM NaCl for cell sorting. All flow-cytometric analyses and cell sorting were carried out on a DakoCytomation MoFlo cell sorter (DakoCytomation, Ft. Collins, CO) equipped with an argon ion laser emitting at 488 nm and a 530/40 bandpass filter. The throughput rate of cells was adjusted to 20000–30000 events per second. Gates were set in the forward scatter and side scatter channels to exclude events arising from large particles. The sorted cells were incubated with SOC medium (2 mL) at 37 °C for 1 h and then diluted with LB medium. After overnight growth at 37 °C, the cells were stored in 25 % glycerol at −80 °C.

Protein expression, purification, and analysis: GFPrm_AM and mDHFR were produced in M9 minimal media supplemented with 19 amino acids, 20 amino acids, or 19 amino acids plus 2 mM Tfn in expression hosts with plasmid-borne copies of the mutant MetRS (pMTY 8 and pMTY 13) or WT MetRS (pMTY10 and pMTY 14). For the mutant MetRS, the marker proteins were purified using nickel-nitrilotriacetic acid (Ni-NTA, Qiagen) chromatography under denaturing conditions according to the manufacturer's instructions. The yield of protein produced per liter of culture was calculated by measuring the absorbance of solutions of the purified proteins at 280 nm, assuming extinction coefficients for GFPrm_AM and mDHFR of 20400 and 30940 $M^{-1}\,cm^{-1}$, respectively.[14] Purified GFPrm_AM and mDHFR were digested with trypsin at 37 °C overnight and the resulting mixtures were injected into an ACQUITY UPLC System equipped with an LCT Premier XE mass spectrometer (Waters, Milford, MA). The samples were separated on an ACQUITY UPLC_BEH C_{18} column (2.1×100 mm, 1.7 μm; Waters) with a gradient of 95 % to 50 % of solvent A (Milli-Q water with 0.1 formic acid) and solvent B (acetonitrile with 0.1 % formic acid) for 5 min (GFPrm_AM) or with a gradient of 95 % to 60 % solvent A for 5 min (mDHFR).

GFP library construction, MetRS library construction, and amino acid activation assays are described in the Supporting Information.

Received: February 20, 2007
Published online: June 14, 2007

Keywords: amino acids · directed evolution · enzymes · high-throughput screening · protein engineering

[1] a) K. Kirshenbaum, I. S. Carrico, D. A. Tirrell, *ChemBioChem* **2002**, *3*, 235–237; b) D. Datta, P. Wang, I. S. Carrico, S. L. Mayo, D. A. Tirrell, *J. Am. Chem. Soc.* **2002**, *124*, 5652–5653; c) Y. Tang, D. A. Tirrell, *Biochemistry* **2002**, *41*, 10635–10645; d) A. J. Link, M. K. S. Vink, N. J. Agard, J. A. Prescher, C. R. Bertozzi, D. A. Tirrell, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10180–10185.

[2] a) I. Kwon, K. Kirshenbaum, D. A. Tirrell, *J. Am. Chem. Soc.* **2003**, *125*, 7512–7513; b) I. Kwon, P. Wang, D. A. Tirrell, *J. Am. Chem. Soc.* **2006**, *128*, 11778–11783.

[3] a) L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498–500; b) S. W. Santoro, L. Wang, B. Herberich, D. S. King, P. G. Schultz, *Nat. Biotechnol.* **2002**, *20*, 1044–1048.

[4] a) A. J. Link, M. L. Mock, D. A. Tirrell, *Curr. Opin. Biotechnol.* **2003**, *14*, 603–609; b) L. Wang, P. G. Schultz, *Angew. Chem.* **2005**, *117*, 34–66; *Angew. Chem. Int. Ed.* **2005**, *44*, 34–66; c) N. Budisa, *Angew. Chem.* **2004**, *116*, 6586–6624; *Angew. Chem. Int. Ed.* **2004**, *43*, 6426–6463.

[5] a) K. L. Kiick, J. C. M. van Hest, D. A. Tirrell, *Angew. Chem.* **2000**, *112*, 2232–2236; *Angew. Chem. Int. Ed.* **2000**, *39*, 2148–2152; b) K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 19–24; c) A. J. Link, D. A. Tirrell, *J. Am. Chem. Soc.* **2003**, *125*, 11164–11165; d) K. E. Beatty, F. Xie, Q. Wang, D. A. Tirrell, *J. Am. Chem. Soc.* **2005**, *127*, 14150–14151.

[6] a) R. Y. Tsien, *Annu. Rev. Biochem.* **1998**, *67*, 509–544; b) M. P. DeLisa, P. Samuelson, T. Palmer, G. Georgiou, *J. Biol. Chem.* **2002**, *277*, 29825–29831; c) A. W. Nguyen, P. S. Daugherty, *Nat. Biotechnol.* **2005**, *23*, 355–360.

[7] a) P. C. Cirino, Y. Tang, K. Takahashi, D. A. Tirrell, F. H. Arnold, *Biotechnol. Bioeng.* **2003**, *83*, 729–734; b) J. K. Montclare, D. A. Tirrell, *Angew. Chem.* **2006**, *118*, 4630–4633; *Angew. Chem. Int. Ed.* **2006**, *45*, 4518–4521; c) M. Rubini, S. Lepthien, R. Golbik, N. Budisa, *Biochim. Biophys. Acta Proteins Proteomics* **2006**, *1764*, 1147–1158; d) T. H. Yoo, A. J. Link, D. A. Tirrell, submitted for publication.

[8] K. L. Kiick, PhD thesis, University of Massachusetts Amherst (USA), **2001**.

[9] a) W. P. C. Stemmer, A. Crameri, K. D. Ha, T. M. Brennan, H. L. Heyneker, *Gene* **1995**, *164*, 49–53; b) P. H. Bessette, M. A. Mena, A. W. Nguyen, P. S. Daugherty in *Directed Evolution: Library Creation Methods and Protocols, Vol. 231* (Eds.: F. H. Arnold, G. Georgiou), Humana Press, Totowa, **2003**, pp. 29–37.

[10] T. Crepin, E. Schmitt, Y. Mechulam, P. B. Sampson, M. D. Vaughan, J. F. Honek, S. Blanquet, *J. Mol. Biol.* **2003**, *332*, 59–72.

[11] T. Crepin, E. Schmitt, S. Blanquet, Y. Mechulam, *Biochemistry* **2002**, *41*, 13003–13011.

[12] K. L. Kiick, R. Weberskirch, D. A. Tirrell, *FEBS Lett.* **2001**, *502*, 25–30.

[13] K. A. Datsenko, B. L. Wanner, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6640–6645.

[14] ProtParam Software can be found under http://ca.expasy.org/tools/protparam.html.

