CHAPTER 4

High-throughput Screening for Methionyl-tRNA Synthetases that Enable Residue-specific Incorporation of Noncanonical Amino Acids into Recombinant Proteins in Bacterial Cells

This chapter appeared as a communication in Tae Hyeon Yoo and David A. Tirrell, *Angew. Chem. Int. Ed.* **2007**, *46*, 5340-5343.

Abstract

We report a high-throughput method of screening aminoacyl-tRNA synthetase libraries for global incorporation of noncanonical amino acids. A variant of the green fluorescent protein was engineered to permit incorporation of methionine analogs without loss of fluorescence. Using the engineered variant as a translational reporter, we screened a saturation mutagenesis library of *E. coli* methionyl-tRNA synthetases (MetRS) for activity toward 6,6,6-trifluoronorleucine (Tfn), and identified a MetRS mutant that enabled high-yield expression of recombinant proteins containing Tfn. The screening method described here is simple, efficient, and directly applicable to Met analogs other than Tfn.

Introduction

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 coworkers have developed powerful methods of selecting aaRS for site-

specific incorporation,[3] and our laboratory has reported an efficient screening system for use in 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 of screening aaRS libraries for 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 4.1). Tfn does not support significant protein synthesis in conventional *E. coli* expression strains.[5a]

Results and Discussion

Unless there are barriers to protein synthesis that lie downstream of the aminoacylation step, the activity of aaRS toward amino acid analogs 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 via 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 analogs 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 8.3.3 variant of GFP[7d] (Table 4.1), 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 BamHI and HindIII sites of pQE-80L; the resulting plasmid was transformed into electrocompetent *E. coli* DH10B cells, yielding more than 10^7 transformants. Following GFP expression in 2xYT media with 1 mM IPTG for 3 h and FACS isolation of highly fluorescent cells, we identified a GFP variant (GFPrm; M78L, M88F, M218A, M233I; Figure 4.1a) that exhibited fluorescence comparable to that of 8.3.3 (Figure 4.1b).

Because GFPrm retains only two Met residues (Table 4.1), 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 4.2). Five positions, D117, K131, Q157, E172, and

K214, where the effects of randomization were insignificant, were changed to Met (Figure 4.1a). Cells expressing the resulting GFP variant (GFPrm_AM) were as bright as those expressing GFPrm (Figure 4.1b). Three Met analogs 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 florescence 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 4.1c). GFPrm_AM 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 NotI and BsrGI sites, and the ligation mixture was transformed into electrocompetent *E. coli* DH10B cells, 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 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 4.3 (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 4.3, 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 exhibited higher fluorescence with 19 amino acids plus Tfn, while the fluorescence of cells expressed with 19 amino acids was comparable to

that of the initial library (Figure 4.3, stage C). One more cycle of positive and negative screening steps yielded library M02c_2 (Figure 4.3, stage D).

Ten clones were selected at random from M02c_2 and sequenced (Table 4.2). Interestingly, nine of those clones had the same amino acid sequence changes (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, while 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)), 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, while negligible expression was observed in media supplemented with 19 amino acids (Figure 4.4 and Table 4.3). In contrast, for strains bearing WT MetRS, addition of Tfn to Met-depleted media had no effect on protein yield, 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 addition of Tfn (Figure 4.4a). 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.)

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 4.5-4.7), 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.

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^{-5} $\mu M^{-1}S^{-1}$ (Table 4.4), roughly seven-fold 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 Met analogs.

Materials and Methods

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/8.3.3) was described previously.[7d] Plasmid pMTY11 was constructed by ligation between a NheI/SalI 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 PCR primers, MetRS_NheI_F and MetRS_Sal_R. Plasmid pMY8 was constructed by ligation between NheI-digested pQE-80/GFPrmAM and the MetRS gene amplified from clone M02c_2-8 using 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 a NheI fragment of pAJL-61[1d] (encoding a 6xHis-tagged mDHFR) and the mutant MetRS gene (from clone M02c_2-8) excised with NheI from pMTY8. The MetRS genes of pMTY8 and pMTY10 were replaced with that of full-length WT MetRS, yielding pMTY13 and pMTY14.

GFP library construction. Three PCR reactions were performed with PfuUltra™ High-Fidelity DNA Polymerase (Strategene) using pQE-80L/L024_3-3 as a template and the following pairs of primers: Out_F and GFP_M78_R, GFP_M78_F and GFP_M218_R,

and GFR_M218_F and Out_R. The DNA fragments obtained from these PCR steps were purified on a 1.5 % agarose gel (QIAquick gel extraction kit, Qiagen). Equimolar quantities of the fragments were mixed and assembled by PCR, and the reaction mixture was purified using Zymo-spin columns (Zymo Research, Orange, CA). The PCR product was amplified using primers, Out_F and Out_R. In order to randomize positions 88 and 233, three PCR reactions were performed using the GFP gene (previously randomized at positions 78 and 218 with NNH codons) as a template and the following pairs of primers: Out_F and GFP_M88_R, GFP_M88_F and GFP_M233_R, and GFP_M233_F and Out_R. The DNA fragments were purified and assembled by PCR. The PCR fragment was digested with BamHI and HindIII and ligated into pQE-80L digested with the same enzymes. A 5 μg portion of digested pQE-80L was used for the ligation; the molar ratio of insert to vector was 3. The ligation mixture was purified using Zymo-spin columns and transformed into electrocompetent *E. coli* DH10B cells, yielding more than 10^7 transformants.

MetRS library construction. Four PCR reactions were performed using pMTY11 as template and the following pairs of primers: Out_OutF and MetRS_L13_R, MetRS_L13_F and MetRS_Mid_R, MetRS_Mid_F and MetRS_H301_R, and

MetRS_Mid_F and Lib_R. The DNA fragments obtained from these PCR steps were purified on a 1.5 % agarose gel and assembled by PCR. The resulting PCR fragment was digested with NotI and BsrGI and ligated into pMTY11 digested with the same enzymes. A 7 μg sample of digested pQE-80L was used for the ligation; the molar ratio of insert to vector was 3. The ligation mixture was purified using Zymo-spin columns and transformed into electrocompetent *E. coli* DH10B cells, yielding more than 5×10^8 transformants. Plasmid DNA library was retransformed into cells of *E. coli* strain DH10B(Met^-) bearing the reporter plasmid pQE-80L/GFPrm_AM.

Screening of MetRS mutants. Cells harboring the MetRS library were grown at 37°C to 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 thiamine) supplemented with 40 mg/L 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 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 20,000 – 30,000 events/sec. 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^{o}C for 1 h and then diluted with LB medium. After overnight growth at 37^{o}C, the cells were stored in 25% glycerol at -80^{o}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 outfitted 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 Ni-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 absorbance at 280 nm of solutions of the purified proteins, assuming extinction coefficients of 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^{o}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).

Amino acid activation assays. The expression plasmid for the mutant MetRS (pMTY21) was constructed by ligation between a BamHI/SalI fragment of pQE-80L and the mutant MetRS gene amplified from clone M02c_2-8 using primers MetRS_N-terminal_BamHI_F and MetRS_N-terminal_SalI_R. The form of MetRS truncated at position 548 was used in these experiments. *E. coli* strain DH10B transformed with pMTY21 was grown at 37°C to OD_{600} = 1 in 1 L of 2xYT medium. After induction with 1 mM IPTG, cells were further grown at 25°C overnight. The cells were harvested by centrifugation (6,000 x *g* for 15 min). The mutant MetRS was purified using Ni-NTA chromatography under native conditions according to the manufacturer's instructions. The column eluent was buffer-exchanged with 100 mM Tris buffer (pH 7.5) containing 2 mM DTT by using PD-10 columns (GE Heathcare). The protein solution was mixed with an equal volume of glycerol and frozen at -80°C until needed. The protein concentration was

determined by measuring absorbance at 280 nm, assuming an extinction coefficient of the mutant MetRS of 93280 $M^{-1}cm^{-1}$. Activation assays were carried out as described[12]. The analog concentrations tested ranged from 312.5 μM to 10 mM, and the mutant MetRS was added to reactions at a concentration of 1 μM for Met and 4 μM for Tfn. Data were fit to the Michaelis–Menten model by using Origin software (Origin Lab).

Oligonucleotides used in this study.

Out_F: 5'-CTTTCGTCTTCACCTCGAG-3'

Out_R: 5'-CTCCATTTTAGCTTCCTTAGCTC-3'

GFR_M78_F: 5'-

CTTTGCGCGTTATCCGGATCATnnhAAACGGCATGACTTTTTCAAG-3'

GFP_M78_R: 5'-

CTTGAAAAAGTCATGCCGTTTdnnATGATCCGGATAACGCGCAAAG-3'

GFP_M88_F: 5'-GACTTTTTCAAGAGTGCCnnhCCCGAAGGTTATGTACAG-3'

GFP_M88_R: 5'-CTGTACATAACCTTCGGGdnnGGCACTCTTGAAAAAGTC-3'

GFP_M218_F: 5'-CAACGAAATGCGTGACCACnnhGTCCTTCATGAGTTTGTAAC-

3'

GFP_M218_R: 5'-GTTACAAACTCATGAAGGACdnnGTGGTCACGCATTTCGTTG-

3’

GFP_M233_F: 5’-

CTGCTGGGATTACACATGGCnnhGATGAGCTCTACAAATAGAAG-3’

GFP_M233_R: 5’-

CTTCTATTTGTAGAGCTCATCdnnGCCATGTGTAATCCCAGCAG-3’

Lib_OutF: 5’-CTTCCTGGCATCTTCCAGGAAATCTC-3’

Lib_R: 5’-AGCGCCTTGAACGGATTCACTTTGTG-3’

MetRS_L13_F: 5’-CTGGTGACGTGCGCAnnkCCGTACGCTAACGGCTC-3’

MetRS_L13_R: 5’-GAGCCGTTAGCGTACGGmnnTGCGCACGTCACCAG-3’

MetRS_Mid_F: 5’-

CTACGTCTGGCTGGACnnknnkATTGGCnnkATGGGTTCTTTCAAG-3’

MetRS_Mid_R: 5’-

CTTGAAAGAACCCATmnnGCCAATmnnmnnGTCCAGCCAGACGTAG-3’

MetRS_H301_F: 5’-GATATTGTTTACTTCnnkAGCCTGTTCTGGCCTG-3’

MetRS_H301_R: 5’-CAGGCCAGAACAGGCTmnnGAAGTAAACAATATC-3’

MetRS_NheI_F: 5`-TTCCGCgctagcTCTAGAGACGTCCGGCCGGAGCTC-3’

MetRS_NheI_R: 5`-TTTGGGgctagcTCTAGAGACGTCCGGCCGGGTAC-3’

MetRS_SalI_R: 5’- TTTGGGgtcgacTCTAGAGACGTCCGGCCGGGTAC-3’

MetRS_N-terminal_BamHI_F: 5'-

TTCCGCggatccATGACTCAAGTCGCGAAGAAAATTC-3'

MetRS_N-terminal_SalI_R: 5'-TTTGGGgtcgacTCATTTAGAGGCTTCCACCAGTG-3'

Acknowledgements

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

References

[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*, 10635-45, d) A. J. Link, M. K. S. Vink, N. J. Agard, J. A. Prescher, C. R. Bertozzi, D. A. Tirrell, *Pro. 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*, 11778-83.

[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. Biotechol.* **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. Biotech.* **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* **2006**, *1764*, 1147-1158, d) T. H. Yoo, A. J. Link, D. A. Tirrell, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 13887 - 13890.

[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.

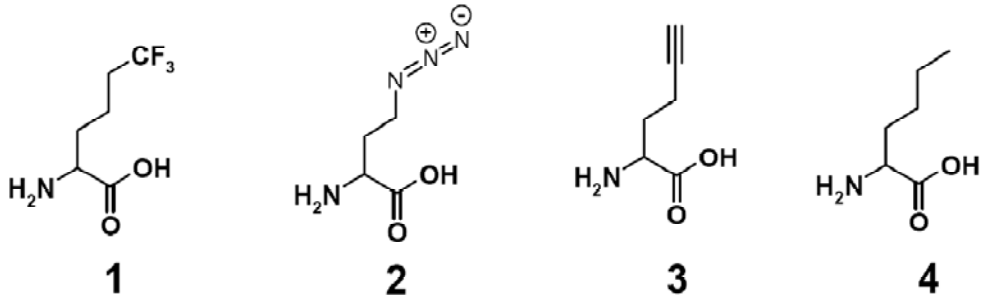
[10] T. Crepin, E. Schmitt, Y. Mechulam, P. B. Sampson, M. D. Vaughan, J. F. Honek, S. Blanquet, *J. Mol. Biol.* **2003**, *332*, 59-72.

[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



Scheme 4.1. Noncanonical amino acids. **1**, 6,6,6-trifluoronorleucine (Tfn); **2**,azidohomoalanine (Aha); **3**, homopropargylglycine (Hpg); **4**, norleucine.

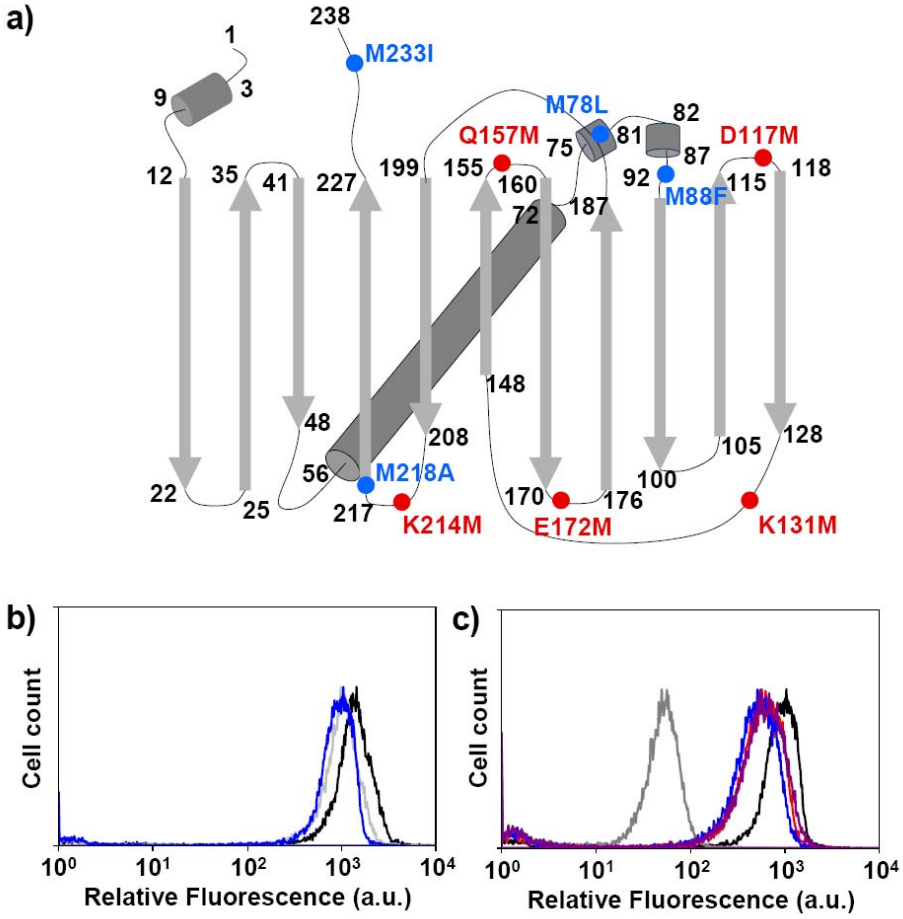


Figure 4.1. Engineering GFP as a translational reporter. a) Schematic representation of the GFP scaffold. Four Met positions (blue) were changed into other codons (GFPrm), and then five Met codons (red) were introduced into the loop regions of the protein (GFPrm_AM). b) Flow cytometric analysis of cells 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).

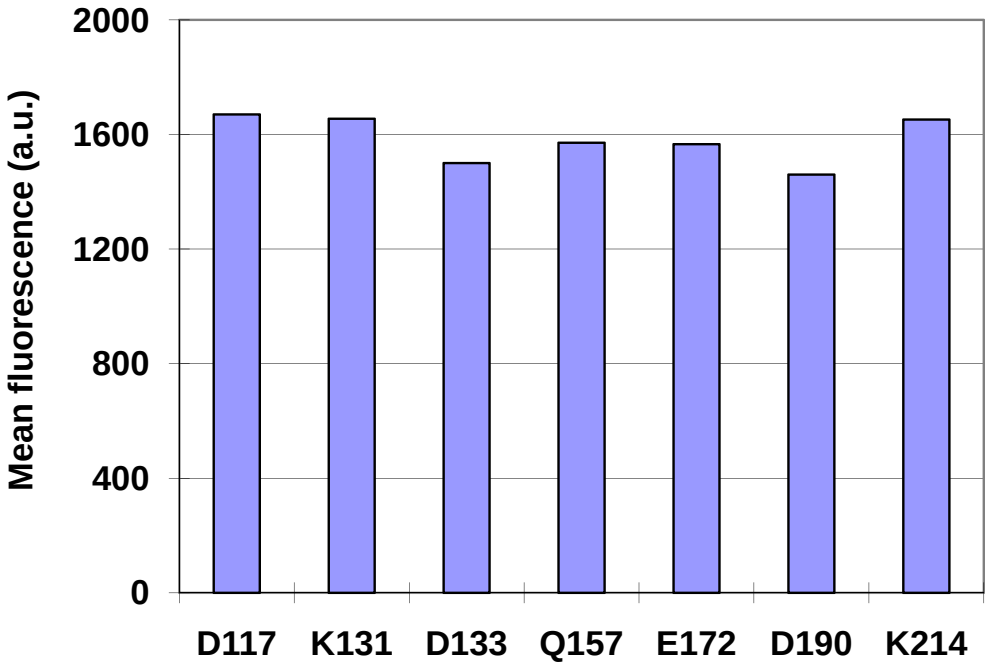


Figure 4.2. Fluorescence of cells expressing GFPrm libraries in which the indicated position was randomized using an NNK codon. Mean cell fluorescence was measured on a cell sorter equipped with an argon ion laser emitting at 488 nm and a 530/40 bandpass filter.

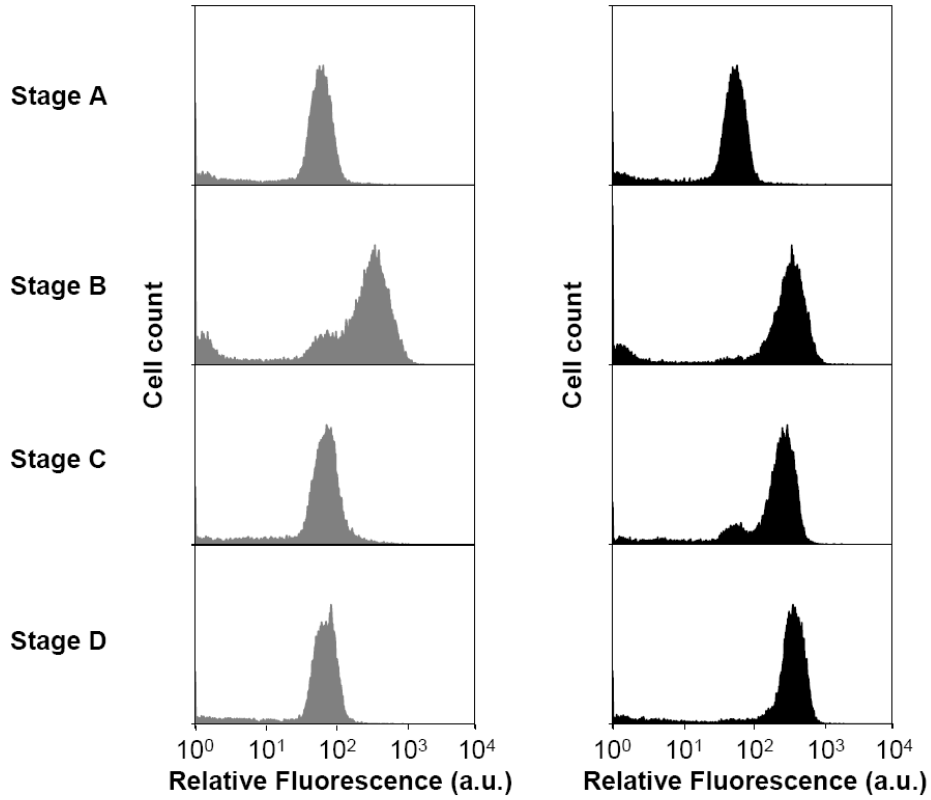


Figure 4.3. Screening of MetRS for activation of Tfn. Fluorescence distributions of cell populations at stages A, B, C, and D (see text) are shown. Gray plots (left) show populations of cells expressing GFPrm_AM with 19 amino acids; black plots (right) show cells expressing GFPrm_AM with Tfn.

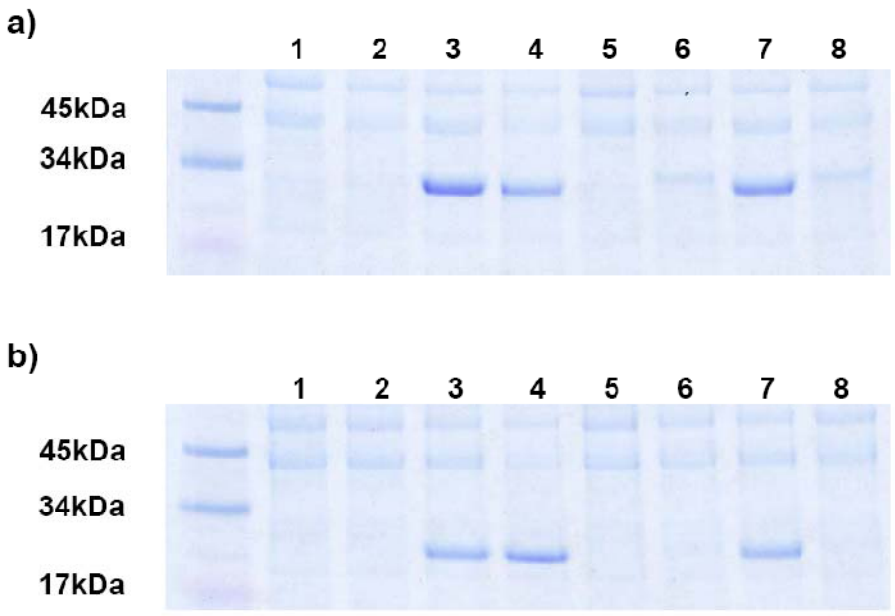


Figure 4.4. Expression of: a) GFPrm_AM and b) mDHFR with mutant MetRS (lanes 1-4) or WT MetRS (lanes 5-8). Lanes 1 and 5: before induction; lanes 2 and 6: induction with 19 amino acids (-Met); lanes 3 and 7: induction with 20 amino acids; lanes 4 and 8: induction with 19 amino acids plus Tfn.

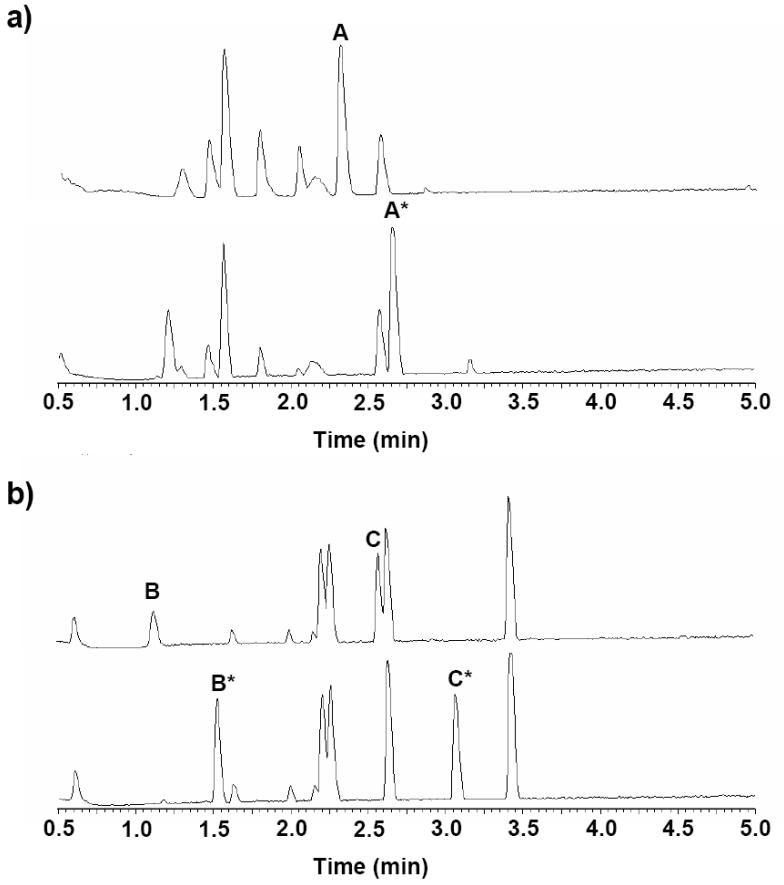


Figure 4.5. LC-MS chromatograms of tryptic digests of a) GFPrm_AM and b) mDHFR. Each protein was produced with 20 amino acids (top) and with 19 amino acids (-Met) plus Tfn (bottom). The signal corresponding to peptide A (retention time 2.31 min, sequence FEGMTIVNR) for GFPrm_AM was shifted to 2.66 min (peak A*) by replacement of Met by Tfn. For mDHFR, signals corresponding to peptides B (retention time 1.12 min, sequence MTTTSSVEGK) and C (retention time 2.57 min, sequence QNLVIMGR) were shifted to 1.53 min and 2.57 min, respectively, upon amino acid replacement.

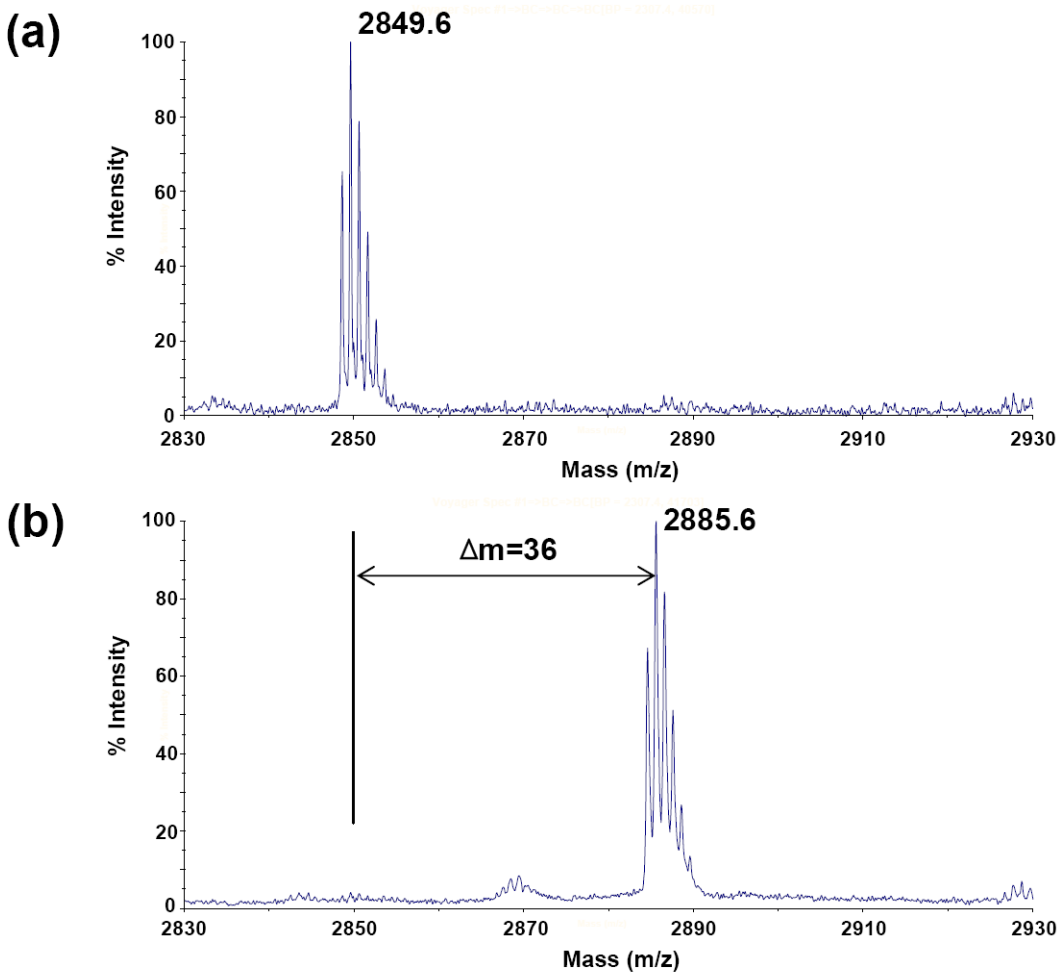


Figure 4.6. MALDI-MS analysis of GFPrm_AM after trypsin digestion. The protein was expressed with a) 20 amino acids or b) 19 amino acids plus 2 mM Tfn, and purified under denaturing conditions. A peptide fragment of sequence HNVMDGSVQLADHYQQNTPIGDGPVR (2849.087 Da) yields the spectra shown. Replacement of Met by Tfn results in a 36 amu mass increase.

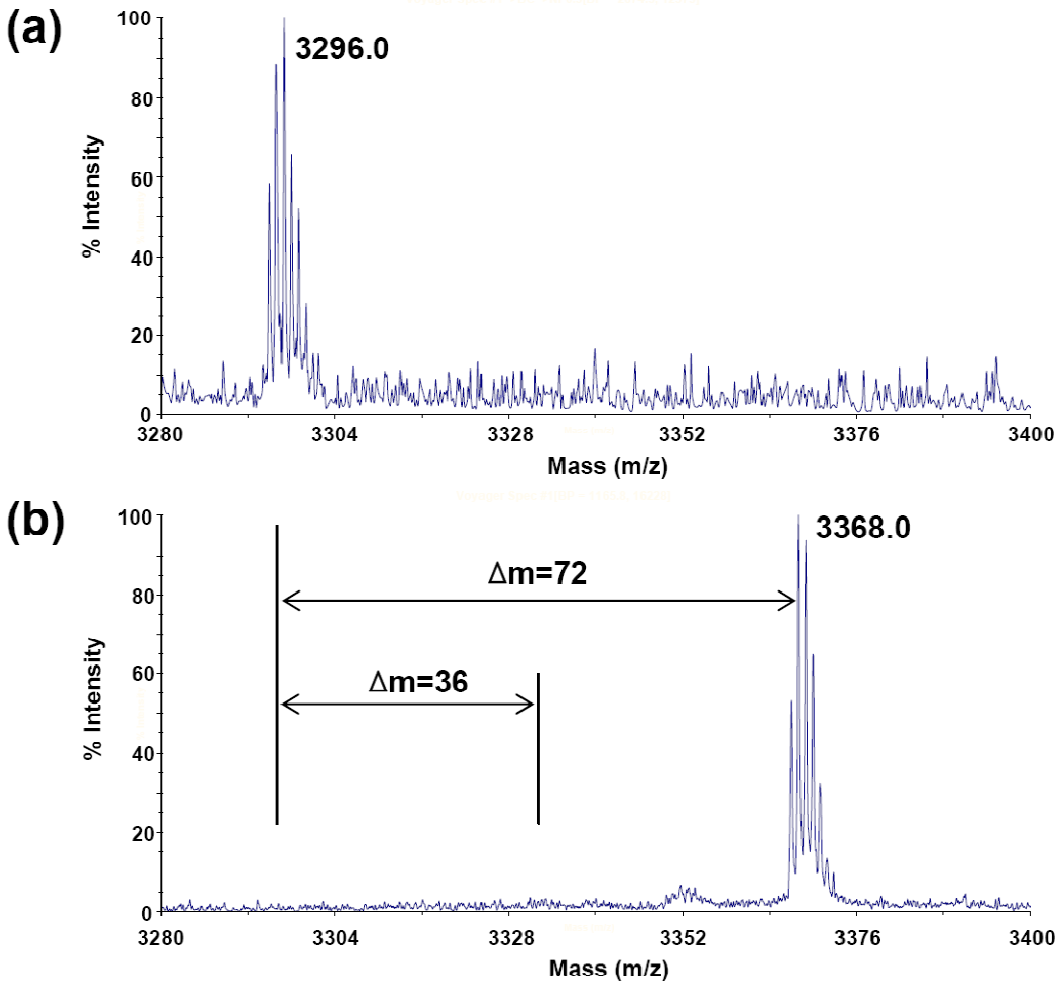


Figure 4.7. MALDI-MS analysis of mDHFR after trypsin digestion. The protein was expressed with a) 20 amino acids or b) 19 amino acids plus 2 mM Tfn, and purified under denaturing conditions. A peptide fragment of sequence GSHHHHHHGSGIMVRPLNSIVAVSQNMGIGK (3295.750 Da) yields the spectra shown. Replacement of Met by Tfn results in a 36 amu mass increase.

Table 4.1. Amino acid sequences of wild-type GFP (WT GFP), L024_3-3, GFPrm, and GFPrm_AM.[a]

	5	10	15	20	25	30
WT GFP	M S K G E E L F T G V V P I L V E L D G D V N G H K F S V S					
8.3.3						R
GFPrm						R
GFRrm_AM						R

	35	40	45	50	55	60
WT GFP	G E G E G D A T Y G K L T L K F I C T T G K L P V P W P T L					
8.3.3		I	L			
GFPrm		I	L			
GFPrm_AM		I	L			

	65	70	75	80	85	90
WT GFP	V T T F S Y G V Q C F S R Y P D H M K Q H D F F K S A M P E					
8.3.3	C G	A		R		
GFPrm	C G	A	L	R		F
GFPrm_AM	C G	A	L	R		F

	95	100	105	110	115	120
WT GFP	G Y V Q E R T I F F K D D G N Y K T R A E V K F E G D T L V					
8.3.3		S	K F			I
GFPrm		S	K F			I
GFPrm_AM		S	K F		M	I

	125	130	135	140	145	150
WT GFP	N R I E L K G I D F K E D G N I L G H K L E Y N Y N S H N V					
8.3.3	K					D
GFPrm	K					D
GFPrm_AM	K	M				D

	155								160								165								170								175								180
WT GFP	Y	I	M	A	D	K	Q	K	N	G	I	K	V	N	F	K	I	R	H	N	I	E	D	G	S	V	Q	L	A	D
8.3.3			T						T				A								V										
GFPrm			T						T				A								V										
GFPrm_AM			T				M		T				A								V	M									

	185								190								195								200								205								210
WT GFP	H	Y	Q	Q	N	T	P	I	G	D	G	P	V	L	L	P	D	N	H	Y	L	S	T	Q	S	A	L	S	K	D
8.3.3														R								L				V	I			
GFPrm														R								L				V	I			
GFPrm_AM														R								L				V	I			

	215								220								225								230								235				238
WT GFP	P	N	E	K	R	D	H	M	V	L	L	E	F	V	T	A	A	G	I	T	H	G	M	D	E	L	Y	K		
8.3.3											H																				
GFPrm								A			H												I								
GFPrm_AM				M				A			H												I								

[a] A hexahistidine tag of sequence MRGSHHHHHHGS was appended to the N-terminus of each protein to enable purification by Ni-NTA chromatography.

Table 4.2. Sequence changes of 10 randomly selected clones from M02c_2. The single letters in parentheses represent the encoded amino acid.

WT	L13	A256	P257	Y260	H301
	CTG	GCA	CCG	TAC	CAC
1	TCT (S)	GCG (A)	CCG (P)	TTG (L)	CTG (L)
2	AGT (S)	GCG (A)	CCG (P)	CTG (L)	CTT (L)
3	TCG (S)	GCG (A)	CCT (P)	CTT (L)	CTG (L)
4	AGT (S)	GCT (A)	CCT (P)	CTT (L)	CTG (L)
5	TCG (S)	GCT (A)	CCG (P)	CTT (L)	CTT (L)
6	TCT (S)	GCG (A)	CCT (P)	TTG (L)	CTG (L)
7	GCG (A)	GCG (A)	CCG (P)	CTT (L)	CTG (L)
8	AGT (S)	GCG (A)	CCG (P)	CTG (L)	CTT (L)
9	AGT (S)	GCA (A)[a]	CCG (P)	TTG (L)	TTG (L)
10	AGT (S)	GCT (A)	CCT (P)	CTG (L)	TTG (L)

[a] GCA codon for A256 of clone 9 might come from errors in synthesis of oligonucleotides.

Table 4.3. 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/liter 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.

Table 4.4. Kinetic parameters for activation of Met and Tfn by the M02c_2-8 MetRS variant

Amino acid	k_{cat}, s^{-1}	K_m, μM	k_{cat}/K_m, $\mu M^{-1}s^{-1}$
Met	1.090 ± 0.024	5313 ± 109	2.1×10^{-4}
Tfn	0.385 ± 0.092	13233 ± 1789	2.9×10^{-5}