The Genetic Incorporation of Seven *Ortho*-substituted Phenylalanine Derivatives

Jeffery M. Tharp, Yane-Shih Wang, Yan-Jiun Lee, Yanyan Yang, and Wenshe R. Liu*

Department of Chemistry, Texas A&M University, College Station, TX 77843

Contents

1. DNA Sequences

MPP8-sfGFP:

atgggcgaagacgtgttcgaagtagagaagatcctggacatgaagaccgaaggcggtaaagttctgtacaaagtgcgctggaaaggctat
acttctgacgatgacacctgggaaccagagatccatctggaagattgtaaagaagttctgctggaatttcgcaagaagattgccgagaacaa
agctaaagagaacctgtatttccaaggtgccatggttagcaaaggtgaagaactgtttaccggcgttgtgccgattctggtggaactggatgg
tgatgtgaatggccataaatttagcgttcgtggcgaaggcgaaggtgatgcgaccaacggtaaactgaccctgaaatttatttgcaccaccg
gtaaactgccggttccgtggccgaccctggtgaccaccctgacctatggcgttcagtgctttagccgctatccggatcatatgaaacgccatg
atttctttaaaagcgcgatgccggaaggctatgtgcaggaacgtaccattagcttcaaagatgatggcacctataaaacccgtgcggaagtta
aatttgaaggcgataccctggtgaaccgcattgaactgaaaggtattgattttaaagaatagggcaacattctgggtcataaactggaatataa
tttcaacagccataatgtgtatattaccgccgataaacagaaaaatggcatcaaagcgaactttaaaatccgtcacaacgtggaagatggtag
cgtgcagctggcggatcattatcagcagaataccccgattggtgatggcccggtgctgctgccggataatcattatctgagcacccagagc
gttctgagcaaagatccgaatgaaaaacgtgatcatatggtgctgctggaatttgttaccgccgcgggcattacccacggtatggatgaactg
tataaaggcagccaccatcatcatcaccattaa

2. Protein Sequences

sfGFP-S2→X:

MASKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWP
TLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEG
DTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSV
QLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMD
ELYKGSHHHHHH

sfGFP:

MVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWP
TLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEG
DTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSV
QLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMD
ELYKGSHHHHHH

*Colored letters indicate residues mutated for sfGFP-F27→**X** and sfGFP-N135→X.

Methanosarcina mazei PylRS:

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARA
LRHHKYRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAP
KPLENTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSISTGATASALVKGNTNPITS
MSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQQIYAEE
RENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRPM
LAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLE
SIITDFLNHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGA
GFGLERLLKVKHDFKNIKRAARSESYYNGISTNL

*Colored residues indicate those mutated to alanine for PylRS(N346A/C348A)

MPP8-sfGFP:

MGEDVFEVEKILDMKTEGGKVLYKVRWKGYTSDDDTWEPEIHLEDCKEVLLEFRKKIA
ENKAKENLYFQGAMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTL
KFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDD
GTYKTRAEVKFEGDTLVNRIELKGIDFKEGNILGHKLEYNFNSHNVYITADKQKNGIKA
NFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLL
EFVTAAGITHGMDELYKGSHHHHHH

*Colored residue indicates site of mutation to NCAA

3. Plasmid Construction

The construction of pET-PylT-sfGFP-S2TAG, pBAD-sfGFPwt and mutants (F27TAG and N135TAG), and pEVOL-pylT-PylRS(N346A/C348A) has been described by us elsewhere[1, 2].

pBAD-MPP8-sfGFP: The codon optimized MPP8 chromo domain gene was synthesized by Epoch Life Science Inc. (Houson, TX) and was amplified using a 5'-end primer containing an ScaI site: 5'-AGTCGAGCTCGGCGAAGACGTGTTCGAAGTAGAGAAGATC-3' and a 3'-end primer with an SalI site: 5'-AGTCGTCGACTTATTTAGCTTTGTTCTCGGCAATCTTCTTGCGAAATTC-3'. The PCR product was cloned into the SacI and SalI site of a previously constructed pBAD-sfGFP to afford pBAD-MPP8-sfGFP. The MPP8-F59TAG containing construct for *ortho*-cyanophenylalanine incorporation was generated by Phusion high fidelity DNA polymerase-based (NEB, MA) site-directed mutagenesis with primer pairs 5'-AGTC GAGCTC GGCGAAGACGTGTAGGAAGTAGAGAAGATC-3' and 5'-AGTC GTCGAC

TTATTTAGCTTTGTTCTCGGCAATCTTCTTGCGAAATTC-3' to afford pBAD-MPP8-F59TAG-sfGFP.

pEGFP-N1-EGFR: The pEGFP-N1-EGFR plasmid was purchased from Addgene (Plasmid 32751). The EGFR-N128TAG construct was generated by Phusion high fidelity DNA polymerase-based (NEB, MA) site-directed mutagenesis with the primer pairs 5'-TAGAAAACCGGACTGAAGGAGCTG-3'and 5'-TGCATCATAGTTAGATAAGACTGC-3'.

pCMV-U6-PylT-PylRS(N346A/C348A): PylRS(N346A/C348A) was amplified from the pEVOL construct using the primers 5'-CTAGTCTAGAATGGATAAAAAACCACTAAACA-3' and 5'-CTAGCTAGCTCACAGGTTGGTAGAAATCCCGTTAT-3'. The PCR product was cloned into Xbal I and Nhe I site of pCMV- U6-PylT.

4. Protein Expression and Purification

sfGFP-S2X:

E. coli BL21(DE3) cells were cotransformed with pET-PylT-sfGFPS2TAG and pEVOL-PylT-PylRS(N346A/C348A) via heatshock at 42°C followed by recovery in 1 mL LB at 37°C for a period of 1 h. The cells were then plated on LB agar for 16 h. The following day a single colony was selected and grown in 5 mL LB at 37°C overnight. The overnight culture was used to inoculate 500 mL LB media which was grown at 37°C (250 rpm) until the OD 600 reached 0.8 at which time the media was switched to a liquid glycerol minimal media (33.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 9.4 mM NH_4Cl, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 1% glycerol, 0.3 mM leucine). To switch the media, cells were pelleted by centrifugation (4,000 rpm, 15 min), washed twice with phosphate buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.5), and resuspended in the minimal media. All media was supplemented with 34 μg/mL chloramphenicol (Cm) and 100 μg/mL ampicillin (Amp). Protein expression was induced by the addition of 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and 0.2 % arabinose along with 2 mM of one of the NCAAs (**1-7**). All NCAAs were purchased from Chem-Impex International, Inc. (Wood Dale, IL) and used without further purification. After expression for 16 h, cells were harvested by centrifugation (4,000 rpm, 20 min), resuspended in ice cold lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), and sonicated. The soluble fraction was recovered by centrifugation (10,000 rpm, 50 min) and the clear supernatant was incubated with 2 mL Ni^{2+}-NTA resin with end-over-end rotation for 1 h at 4°C. The resin was then washed with 40 mL of the lysis buffer and the protein was eluted from the column with an elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0). The eluted protein was dialyzed against PBS and concentrated using an Amicon Ultra-15 Centrifugal Filter (10,000 MWCO). The purified proteins were analyzed via 12% SDS-PAGE and electrospray ionization mass spectrometry.

sfGFP-N135X, sfGFP-F27X, and MPP8-F59X-sfGFP:

E. coli Top10 cells were cotransformed with pBAD plasmid containing the gene encoding sfGFP-N135X, sfGFP-F27X, or MPP8-F59X-sfGFP, along with pEVOL-pylT-PylRS(N346A/C348A) via heat shock at 42°C followed by recovery in LB for 1 h at 37°C. The cells were then plated on LB agar. The following day a single colony was grown in 5 mL LB overnight. The overnight culture was used to inoculate 200 mL of a synthetic auto-induction media (described elsewhere)[3] supplemented with 2 mM NCAA **7**. To avoid background, when preparing the auto-induction media, L-phenylalanine was not included. All media was supplemented with 34 μg/mL Cm and 100 μg/mL Amp. The culture was grown for 16 h at 37°C (250 rpm) and cells were harvested by centrifugation (4,000 rpm, 20 min). The proteins were then purified as described above.

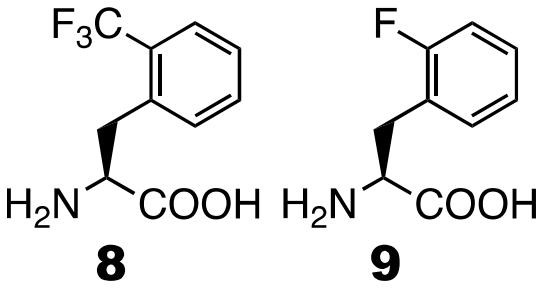
5. Mammalian Cell Cultures and Transfection

HEK293T cells purchased from American Type Culture Collection (ATCC, Manassas, VA) were grown in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen) containing 10% fetal bovine serum (Invitrogen). Cells were seeded and grown on a Becton Dickinson 24-well plate. At approximately 70-80% confluence, cells were transfected with two plasmids, pCMV-U6-PylT-PylRS (N346A/C348A) and pEGFP-N1-EGFR (N128TAG). Transient transfection was carried out using Lipofectamine® 2000 Transfection Reagent from Invitrogen, according to the manufacturer's protocol. After transfection, the growth medium was DMEM with 10% FBS which contained NCAAs **1-7** or no unnatural amino acids. The images were recorded by an Olympus FluoView FV1000 Confocal Microscope (Japan) after 48 hours and analyzed using ImageJ (NIH).

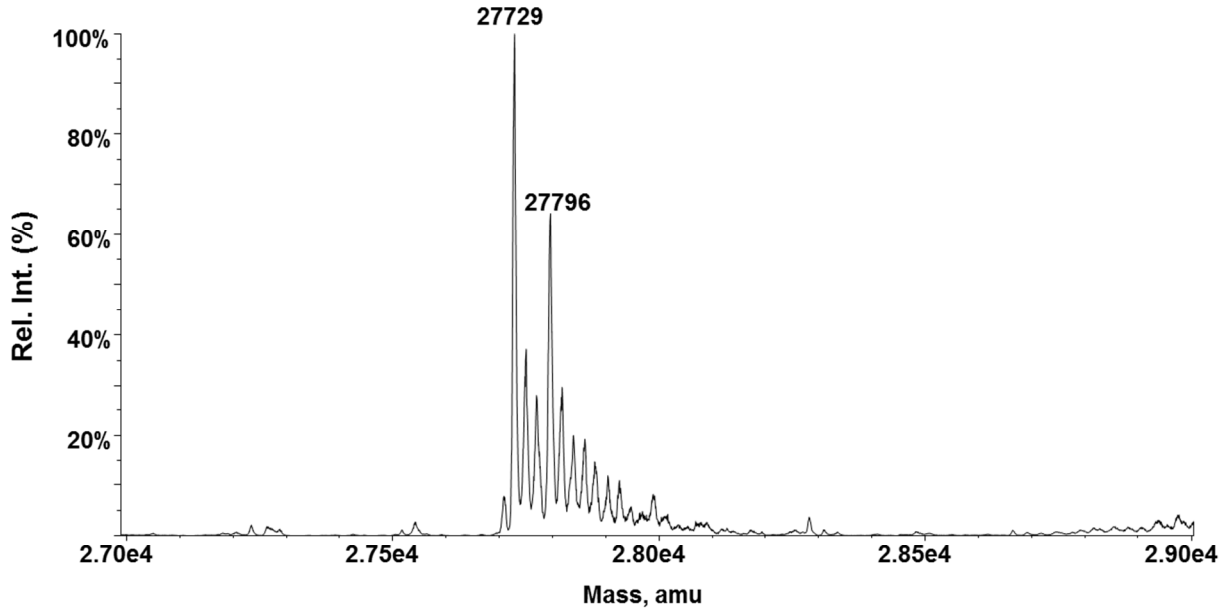
6. Fluorescence Measurements

For all samples the fluorescence was measured using a Quanta Master 40 fluorometer (Photon Technology International) equipped with a Xenon lamp. Spectra were collected at room temperature using a 1 cm quartz sample holder, 0.5 mm slit width, 1 nm resolution, and 1 s/nm integration time. The samples were excited at 240 nm and the fluorescence was measured from 250-450 nm. For the protein folding measurements, stock solutions of the wildtype and mutant proteins were diluted to an appropriate volume in PBS, pH 7.5 and varying concentrations of a 10 M urea solution were added to give a final urea concentration ranging from 0 to 8. After the addition of the denaturant, 2 h were allotted for the transitions to occur before recording the fluorescence spectrum.

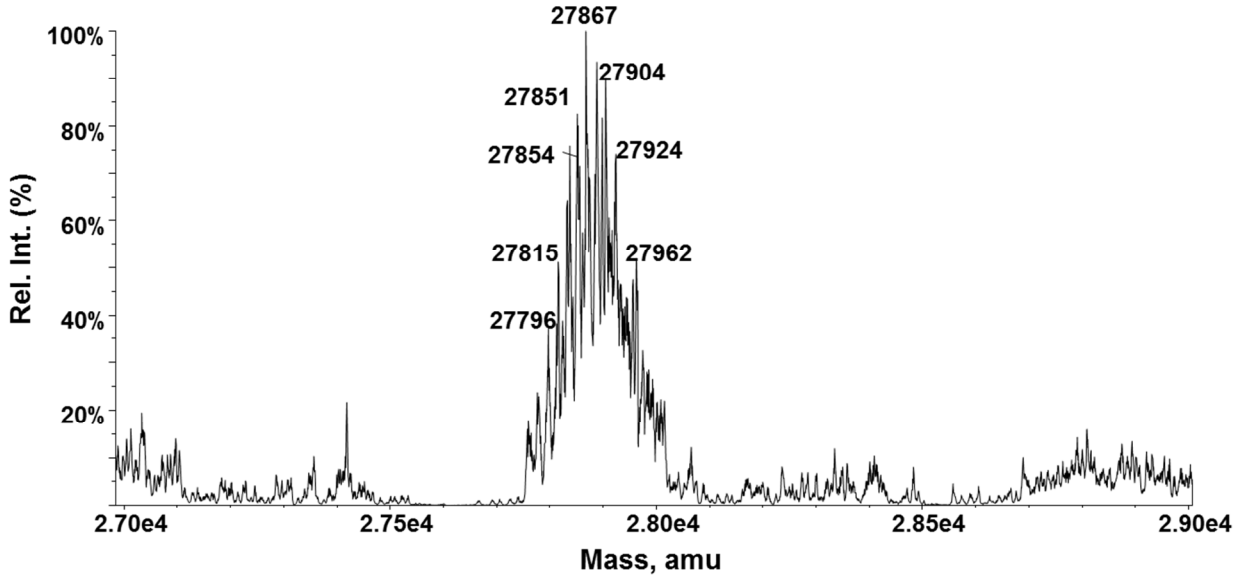
7. Supplementary Figures



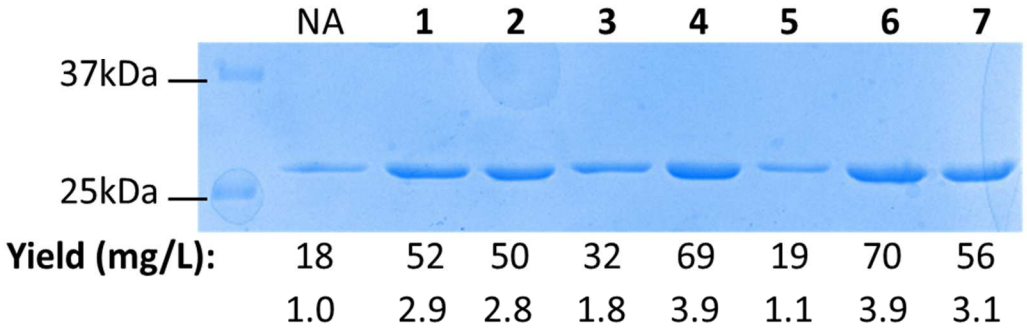
Supplementary Figure 1. Structures of NCAAs **8** and **9**.



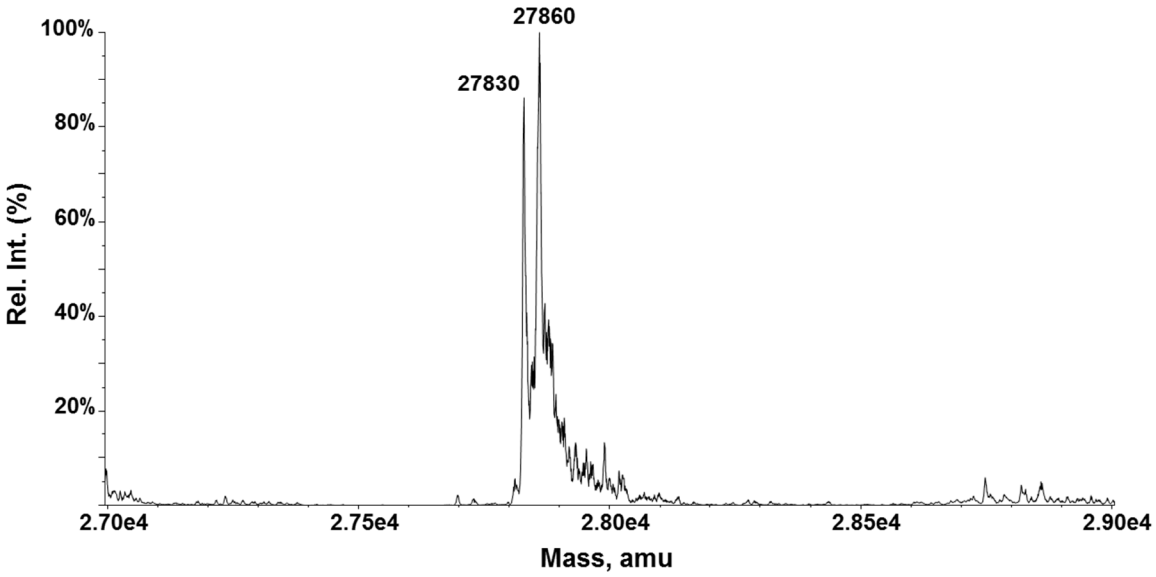
Supplementary Figure 2. Deconvoluted electrospray ionization mass spectrum of sfGFP with **8** incorporated at the S2 position. The major peak in the spectrum is from the protein with L-phenylalanine incorporated at the amber mutation (27729 Da) while the minor peak is from the protein with **8** incorporated (27796 Da).



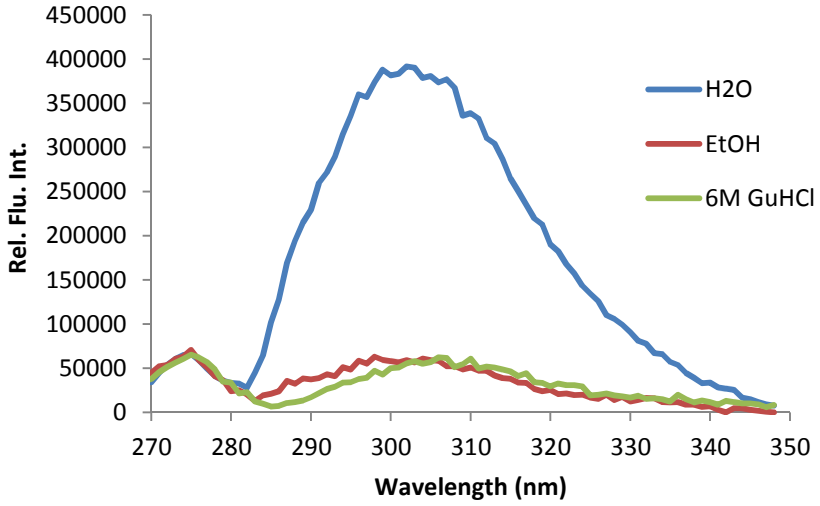
Supplementary Figure 3. Deconvoluted electrospray ionization mass spectrum of sfGFP with **9** incorporated at the S2 position. The major peaks with larger masses than the predicted mass (27747 Da) indicate **9** may also be incorporated at phenylalanine sites.



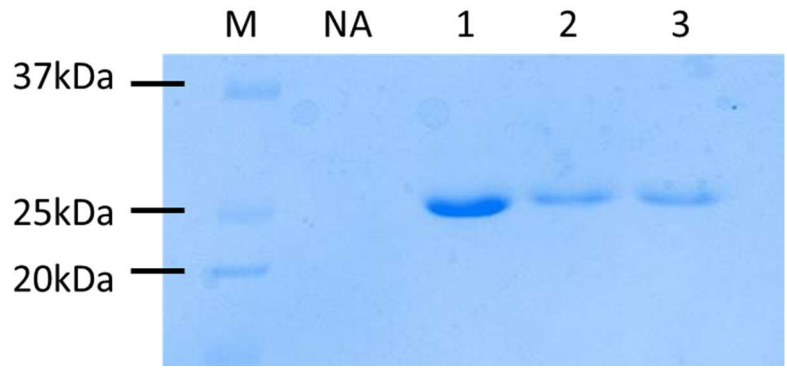
Supplementary Figure 4. Expression of sfGFP-F27→X (where X is one of the NCAAs **1-7**) in LB media. NA indicates no NCAA was present in the growth media. Bottom numbers indicate the expression level relative to the control. Proteins were expressed for 12 hours in LB media supplemented with 2 mM NCAAs and then purified as detailed above.



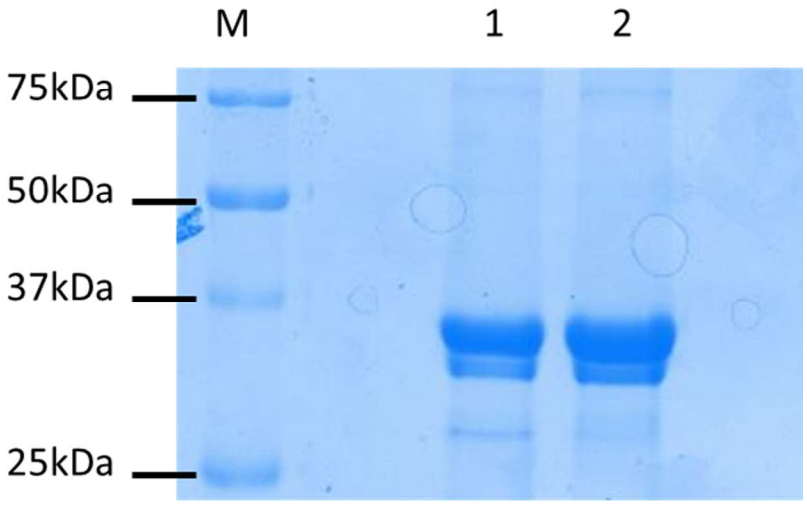
Supplementary Figure 5. Deconvoluted electrospray ionization mass spectrum of sfGFP-F27→**5.** The major peak in the spectrum is from the incorporation of **5** (27859 Da) while the minor peak is from the incorporation of phenylalanine (27830 Da).



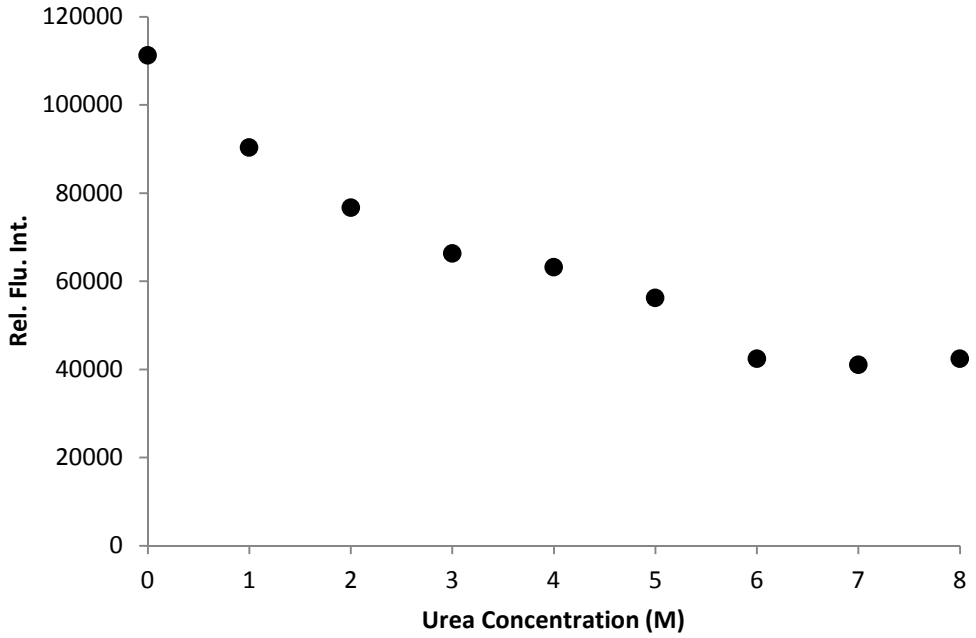
Supplementary Figure 6. Fluorescence spectrum of 20 μM, *ortho*-cyano-L-phenylalanine dissolved in water, ethanol, and 6 M guanidine hydrochloride. Fluorescence intensity increases when going from organic to aqueous solvents. As observed with *p*CNF, chloride is an effective quencher of fluorescence. Spectra were recorded in a 1 cm quartz cuvette at room temperature with a 1 nm resolution and 1 s/nm integration time, λex = 240 nm.



Supplementary Figure 7. Expression of sfGFP-F27→**7** (1), sfGFP-N135→**7** (2), and wildtype sfGFP (3). Mutant proteins were expressed for 16 hours in auto-induction media supplemented with 2 mM *ortho*-cyano-L-phenylalanine. Wild-type protein was expressed in LB for 16 hours. sfGFP was purified via nickel affinity chromatography. Protein purity was confirmed by 12% SDS-PAGE with Coomassie blue staining. NA indicates no NCAA was available in the growth medium.



Supplementary Figure 8. Expression of MPP8-F59→**7**-sfGFP (1) and MPP8wt-sfGFP (2). Mutant protein was expressed for 16 hours in auto-induction media supplemented with 2 mM *ortho*-cyano-L-phenylalanine. The wild-type protein was expressed in LB for 16 hours. Both proteins were purified via nickel affinity chromatography and the purified proteins were analyzed by 12% SDS-PAGE with Coomassie blue staining.



Supplementary Figure 9. Fluorescence intensity of wild-type MPP8 measured at 297 nm as a function of the urea concentration. The sample was excited at 240 nm. No increase in fluorescence at 297 was observed for the wild-type protein. Protein concentration was 2.5 μM in PBS, pH 7.5.

References

1. Wang, Y. S., Fang, X., Wallace, A. L., Wu, B., and Liu, W. R. (2012) A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum, *J. Am. Chem. Soc. 134*, 2950-2953.
2. Wang, Y. S., Fang, X., Chen, H. Y., Wu, B., Wang, Z. U., Hilty, C., and Liu, W. R. (2013) Genetic incorporation of twelve *meta*-substituted phenylalanine derivatives using a singly pyrrolysyl-tRNA synthetase mutant, *ACS Chem. Biol. 8*, 405-415.
3. Hammill, J. T., Miyake-Stoner, S., Hazen, J. L., Jackson, J. C., and Mehl, R. A. (2007) Preparation of site-specifically labeled fluorinated proteins for ^{19}F-NMR structural characterization, *Nat. Protocols 2*.