**Site-Specific Incorporation of 7-Fluoro- L-tryptophan into Proteins by Genetic Encoding to Monitor Ligand Binding by $^{19}$F NMR Spectroscopy**

Haocheng Qianzhu,[1] Elwy H. Abdelkader,[2] Iresha D. Herath,[1] Gottfried Otting,[2]* and Thomas Huber[1]*

[1] Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

[2] ARC Centre of Excellence for Innovations in Peptide & Protein Science, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

## Table of contents:

## Experimental procedures

### a) Materials

7-Fluoro-DL-tryptophan was purchased from Fluorochem Ltd, UK (catalogue code: 541432) and was used assuming a 7-fluoro-L-tryptophan (7FTrp) content of 50%. 4-Fluoro-L-tryptophan was purchased from Acrotein ChemBio Inc., USA (catalogue code: A-0816). 7-Fluoroindole (7FI) was purchased from AK Scientific Inc., USA (catalogue code: J5720).

### b) Screening of functional G1PylRS enzymes recognizing FTrp

To carry out the selection, the library plasmid pBK-G1RS was transformed into *E. coli* DH10B cells harboring the selection plasmid pBAD-H6RFP.[1] Following recovery from transformation, the culture was directly inoculated into a flask with 25 mL LB medium containing 100 mg/L carbenicillin and 50 mg/L kanamycin, supplied with 0.4% L-arabinose and 2.5 mM FTrp (7FTrp or 4FTrp), which served as the sample for the first round of positive selection (**1P+**). Overnight expression at 37 ºC led to a readily detectable level of RFP expression. Cells were resuspended in 7.5 mL PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 1.8 mM $KH_2PO_4$, pH 7.4) after harvesting. A 100-fold dilution yielded a concentration suitable for cell sorting by FACS on an Aria II high speed cell sorter (BD Biosciences, USA; Figures S1 and S2).

Cells with high RFP levels were selected from the **1P+** sample (0.4% for 7FTrp or 0.3% for 4FTrp, as indicated by rectangles in Figure S1), collecting $4.0\times10^5$ cells in 90 minutes. The cells collected were subjected to a following round of negative selection without the addition of FTrp and regrown as sample **2N–**, from where cells with low RFP expression levels (44.6% for 7FTrp; 38.8% for 4FTrp) were collected ($1.0\times10^6$ cells in 10 minutes). These cells were aliquoted to inoculate media with positive (**3P+**) and negative (**3P–**) conditions. The RFP-positive cells (11.6% for 7FTrp; 7.1% for 4FTrp) from the **3P+** round were collected ($4.0\times10^5$ cells in 25 minutes) and recovered under negative condition to obtain **4N–**. Following sorting ($1.0\times10^6$ cells in 10 minutes), cells showing the lowest level of RFP fluorescence (37.2% for 7FTrp; 39.7% for 4FTrp) were selected from the **4N–** sample. They were aliquoted to be recovered as **5P+** and **5P–,** respectively. The cell population with high RFP fluorescence in the **5P** round was 10.1% with 7FTrp provided compared to 0.2% without 7FTrp, indicating the successful accumulation of active G1PylRS variants specific for 7FTrp. In contrast, the selection process targeting 4FTrp did not produce significant enrichment, as indicated by similar numbers of high RFP fluorescent cells in the **5P+** and **5P–** samples. Given the estimated

diversity of the remaining library pool was lower than ten, this suggests the absence of a specific synthetase for 4FTrp in this G1PylRS library. $1.0\times10^5$ cells collected in 10 minutes from the top 10.1% RFP fluorescent cells of the **5P+** sample for 7FTrp were recovered for storage. An aliquot of 2,000 cells were allowed to recover on LB agar plates containing 100 mg/L carbenicillin and 50 mg/L kanamycin, and individual clones were analyzed using 96-well plates. 60 candidates were inoculated into both positive (with 2.5 mM 7FTrp) and negative (without 7FTrp) growth conditions. The fluorescence level was measured after expression overnight, using a TECAN Infinite 200 Pro M Plex plate reader (Tecan, Switzerland) and normalized by the $OD_{600}$ of the cell culture. The average $OD_{600}$ of the overnight cell cultures supplied with 2.5 mM 7FTrp was 0.386, which is not different from the same cell cultures without 7FTrp (0.364), indicating that cell growth is not affected by 2.5 mM 7FTrp. Seven candidates with the highest RFP level in the positive condition were chosen for sequencing. Five individually different sequences were found (Table S1), with one mutation set identified three times. The clone producing the highest fluorescence (G17F02) was found to carry the mutations L124G, Y125F, N165G, V167F, Y204W, A221G, and W237Y. This G1PylRS mutant was used in the subsequent experiments and referred to as G1(7FTrp)RS.

**c) *In vivo* protein expression and purification**

Site-specific incorporation of 7FTrp was achieved in *E. coli* B-95.ΔA cells[2] co-transformed with pRSF-G1(7FTrp)RS and the pCDF plasmid containing the amber codon interrupted gene of the protein of interest. The transformed cells were grown at 37 °C in LB medium containing 25 mg/L kanamycin and 25 mg/L spectinomycin. An aliquot (0.5 mL) of an overnight culture was used to inoculate 50 mL LB medium supplemented with 25 mg/L kanamycin, 25 mg/L spectinomycin, 2 mM 7FTrp and 4 mM tryptophan. The cells were grown at 37 °C to an $OD_{600}$ of 0.6–1. At this point, the temperature was reduced to 25 °C and protein expression was induced by the addition of 1 mM IPTG.

When using 7FI as precursor of 7FTrp, a 10 mL aliquot of an overnight culture was used to inoculate 1 L LB medium supplemented with 25 mg/L kanamycin and 25 mg/L spectinomycin. The cells were grown at 37 °C to an $OD_{600}$ of 0.6, at which point 7FI was added to the culture at a final concentration of 1 mM. The cells were grown at 37 °C for an additional 30 minutes. Subsequently, the temperature was reduced to 25 °C and protein expression was induced by the addition of 1 mM IPTG.

Following expression for 16 h, the cells were harvested by centrifugation. Following resuspension in buffer A (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% glycerol, 10 mM

imidazole), the cells were lysed using an Avestin Emulsiflex C5 system (Avestin, Canada) using two passes with a pressure of 10,000–15,000 psi. The cell lysates were centrifuged for 1 h at 30,000 *g*. The supernatant was loaded onto a 1 mL His GraviTrap column (Cytiva, USA). The column was washed with 20 column volumes buffer A and the protein was eluted with 5 column volumes buffer B (same as buffer A but with 500 mM imidazole). Afterwards, the buffer was exchanged to NMR buffer (PBS for RFP13-TAG; 20 mM MES, pH 6.5, and 150 mM NaCl for ZiPro) using an Amicon ultrafiltration centrifugal tube (Merck Millipore, USA) with a molecular weight cut-off of 10 kDa. Prior to NMR measurements, 1 mM trifluoroacetic acid (TFA) was added for $^{19}$F NMR referencing. Yields ranged between 3.7 and 8.5 mg of purified protein per liter of cell culture, yielding NMR sample concentrations ranging between 0.3 and 0.7 mM.

**d) Intact protein mass spectrometry**

Intact protein analysis was performed on an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, USA) connected to a Thermo Fisher Scientific UltiMate 3000 HPLC system equipped with ZORBAX 300SB-C3, 3.5 µm, 4.6 x 50 mm HPLC column (Agilent Technologies, USA). Approximately 50 pmol of sample was injected using a 500 µL/min linear gradient of solvent A (0.1% (v/v) formic acid in water) and solvent B (0.1% (v/v) formic acid in acetonitrile), ramping solvent B from 5% solvent B at the start to 80% after 12 min. Data were collected using an electrospray ionization (ESI) source in positive ion mode. Protein intact mass was determined by deconvolution using the program Xcalibur 3.0.63 (Thermo Fisher Scientific, USA).

**e) NMR spectroscopy**

$^{19}$F NMR spectra were recorded on a Bruker 400 MHz NMR spectrometer equipped with a room temperature broadband probe. Parameters used: 200 ms acquisition time, $^{1}$H decoupling during acquisition, recovery delay 1 s, exponential window multiplication with 50 Hz linebroading prior to Fourier transformation. Each spectrum of ZiPro was recorded at 25 °C in about 5–25 minutes depending on the protein concentration. Chemical shifts were calibrated relative to internal TFA (-75.25 ppm). The $^{19}$F-$^{19}$F NOESY spectrum of ZiPro with 7FTrp in position 83 was recorded with a mixing time of 200 ms, $t_{1max}$ = 7.1 ms, $t_{2max}$ = 35 ms, and a total recording time of 14 h. The data were processed with exponential window multiplication in both dimensions, using a line broadening value of 80 Hz.

## Supplementary figures

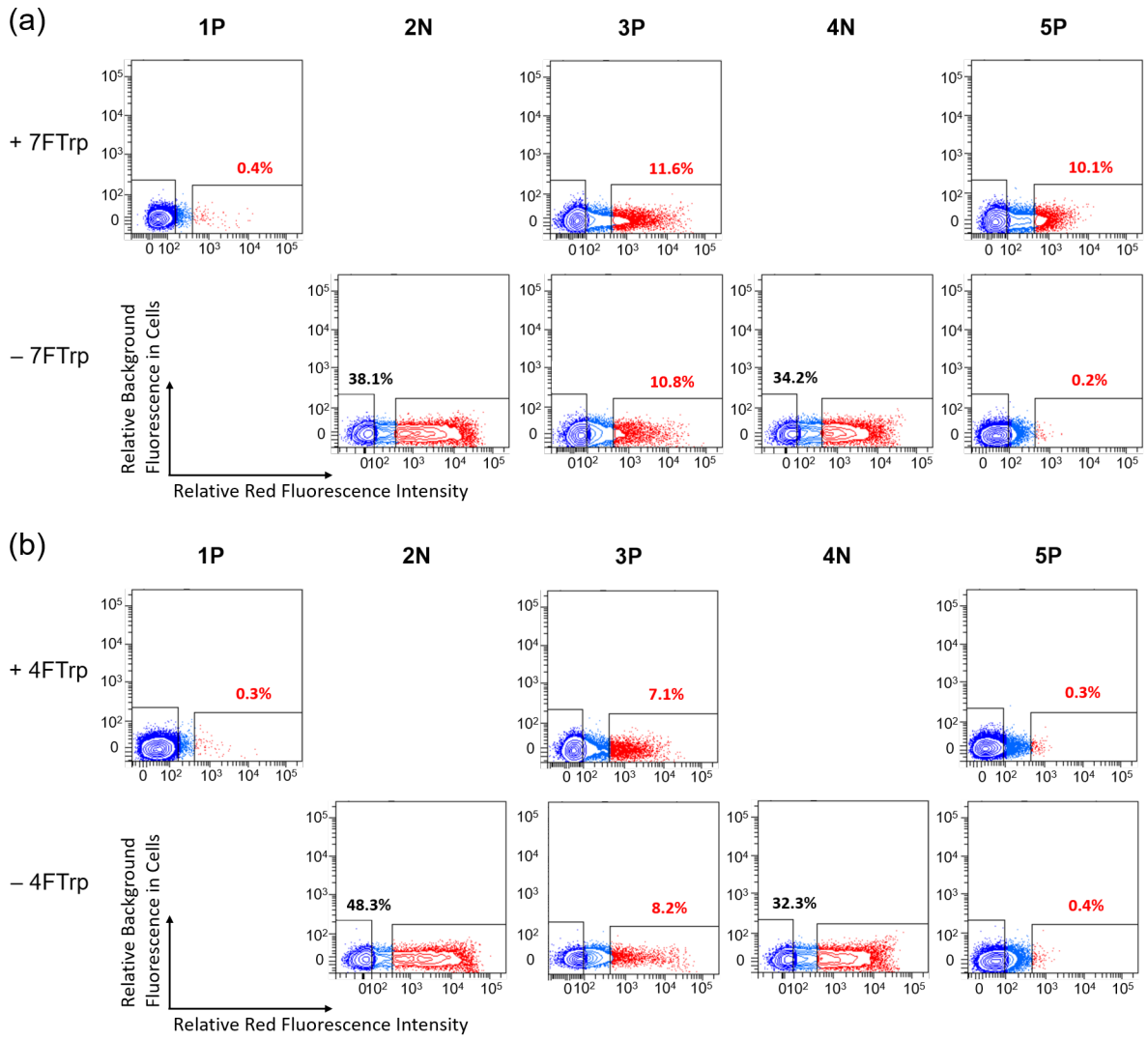


**Figure S1.** FACS experiments for selection of active and specific G1PylRS enzymes that recognize (a) 7FTrp or (b) 4FTrp. The horizontal axis indicates the relative intensity of red fluorescence. The vertical axis plots the level of background fluorescence in cells excited at 488 nm. **P** stands for positive selection rounds where cells were cultured in the presence of FTrp; **N** indicates negative rounds without ncAA supply. Cells with high level of RFP fluorescence in the **P** rounds were collected (indicated by boxes drawn) and re-cultured. Following **N** rounds, cells with low RFP expression were selected for use in subsequent rounds. Identical cell cultures omitting FTrp (– FTrp) in **3P** and **5P** rounds serve as control.

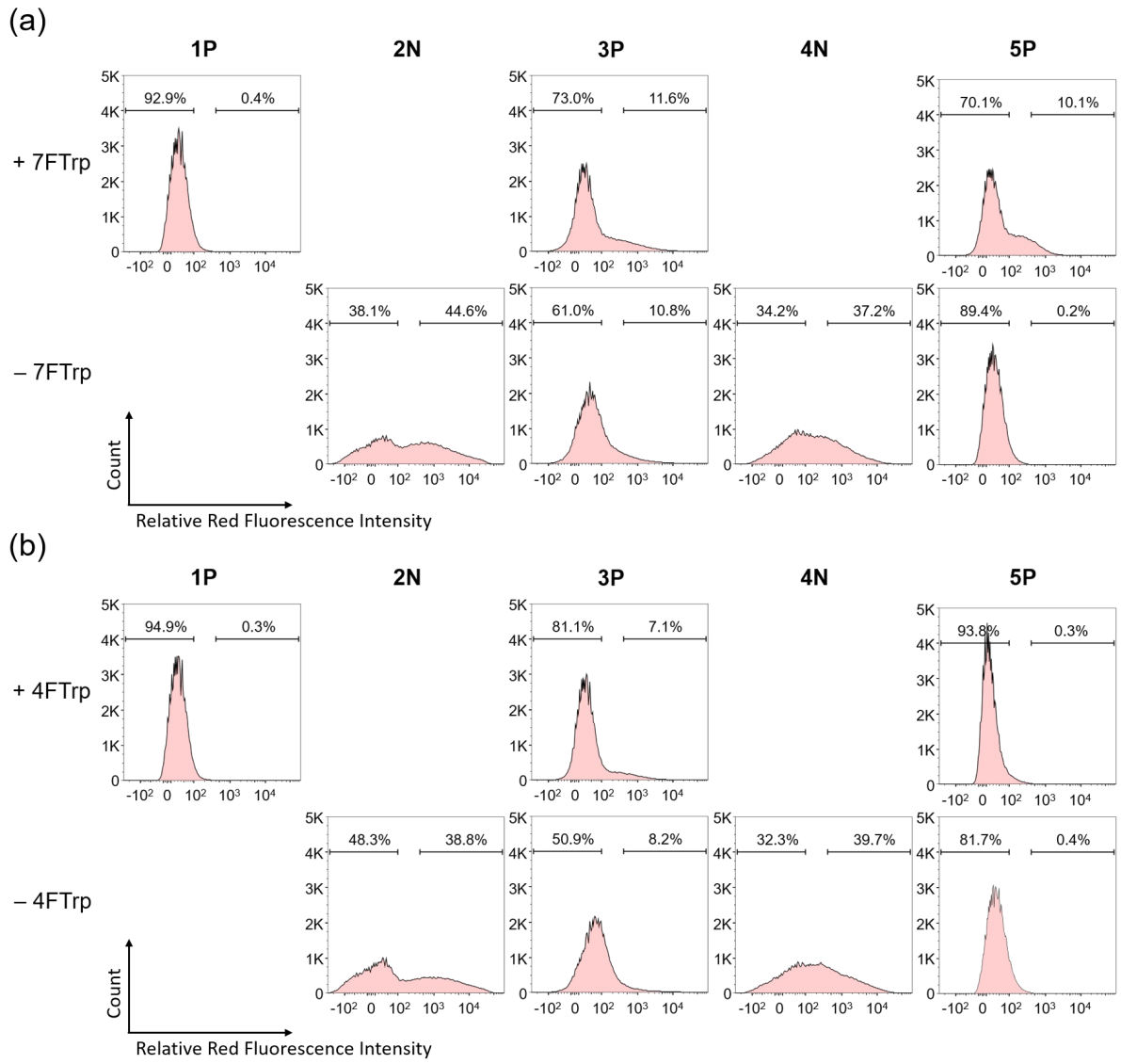


**Figure S2.** Histograms of five selection rounds to identify active G1PylRS enzymes for (a) 7FTrp or (b) 4FTrp via FACS screening. The difference in RFP fluorescence intensity of cells grown with or without FTrp serves as an indicator of the presence of FTrp-specific RS enzymes in the gene pool. The plots show the cell count versus relative intensity of red fluorescence. 10.1% of the cells in the **5P+** sample for 7FTrp with the brightest RFP fluorescence level were collected and characterized further.

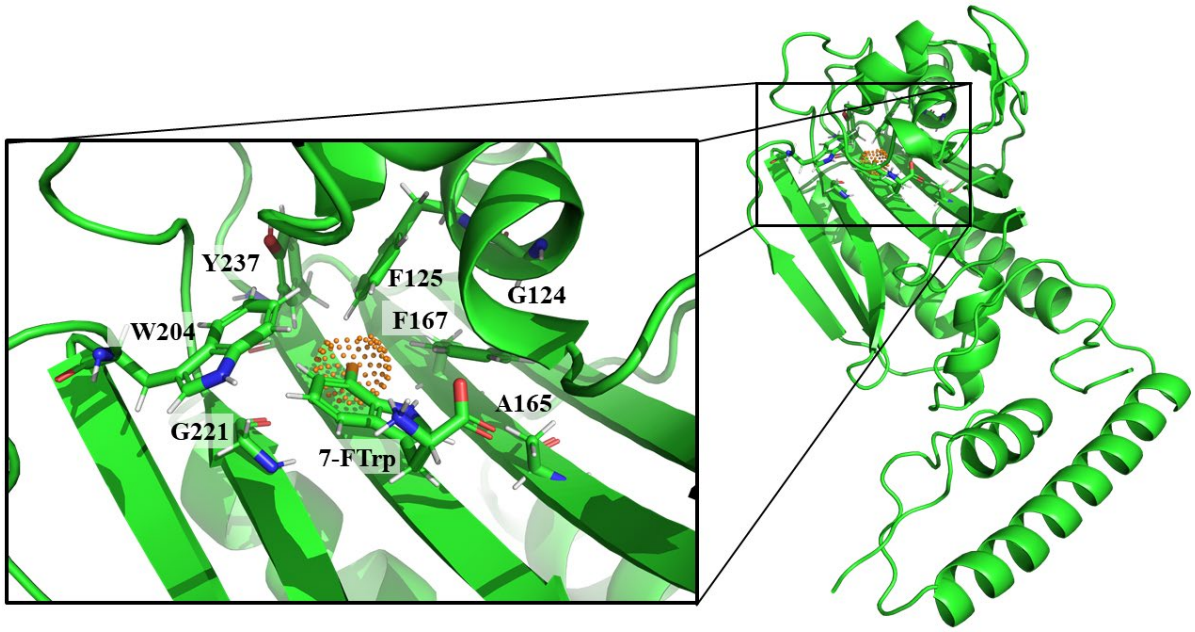


**Figure S3.** Computationally derived model of G1(7FTrp)RS. The structure of *m*-cyanopyridylalanyl-tRNA synthetase (PDB ID: 7R6O)[1] was mutated to G1(7FTrp)RS and the ligand location was transferred from the crystal structure of *M. mazei* PylRS in complex with 1-methyl-L-tryptophan and AMPNP (PDB ID: 6LYA)[3] before the structure was energy minimized with restraints to the start structure in Rosetta. The aromatic residues F125, F167, W204, and Y237 form a strongly hydrophobic cavity around the fluorine in the 7 position of the indole moiety (orange; with dots visualizing the approximate van der Waals radius). The 4 position in the indole has fewer possibilities to interact with the residues mutated in the G1PylRS library, limiting the likelihood of finding a tRNA synthetase that effectively can discriminate 4FTrp from tryptophan.

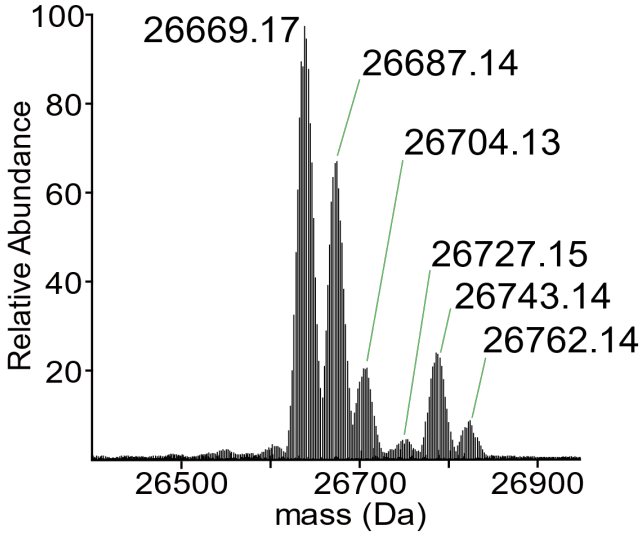


**Figure S4.** Intact protein mass spectrometric analysis of RFP-13TAG expressed in LB medium containing 3 mM 4FTrp co-expressed with G1(7FTrp)RS. The experiment was conducted to check the specificity of G1(7FTrp)RS. Less than 0.1 mg protein was obtained from 50 mL of cell culture. The peak at 26669.17 Da corresponds to amber stop codon suppression by glutamyl-tRNA. The peaks at 26687.14 Da and 26704.13 Da correspond to, respectively, single and double replacement of natural tryptophan residues by 4FTrp. The peak at 26727.15 Da corresponds to the misincorporation of tryptophan in response to the amber codon in position 13. The peak at 26743.14 Da corresponds to single 4FTrp incorporation in position 13 (the calculated mass is 26,745.09 Da). The peak at 26762.14 Da corresponds to the additional installation of 4FTrp in place of a natural tryptophan residue besides the 4FTrp replacement at position 13.

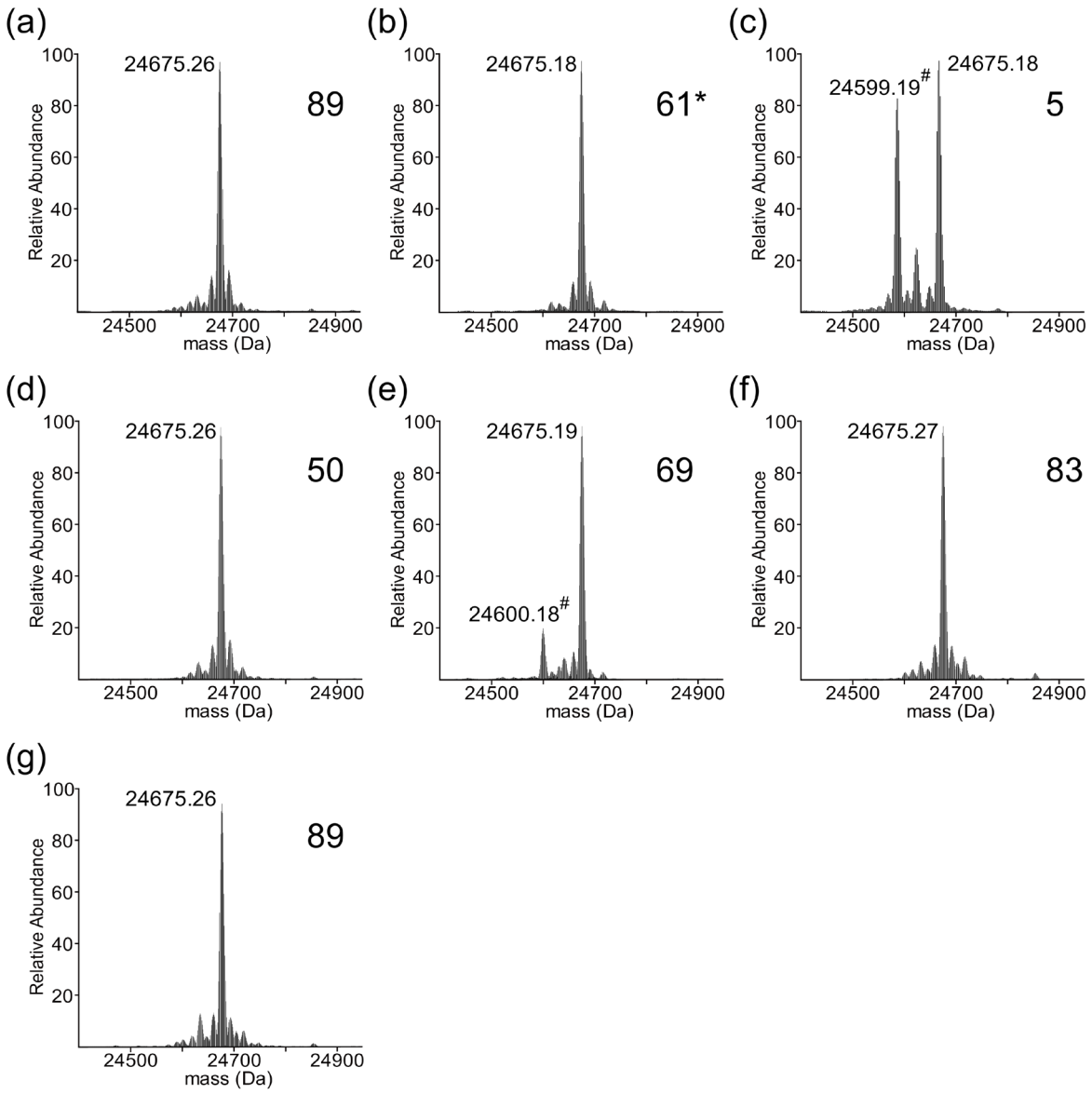


**Figure S5.** Intact protein mass spectrometric analysis of ZiPro mutants with 7FTrp. (a) ZiPro with 7FTrp in position 89 produced in LB medium supplemented with 2 mM 7FTrp and 4 mM tryptophan. The calculated mass (following loss of the N-terminal methionine) is 24676.45 Da. (b–g) ZiPro mutants with 7FTrp produced in LB medium containing 1 mM 7FI. The spectra are annotated with the sequence number of the respective mutation sites. The masses of the peaks marked with a "#" symbol in (c) and (e) correspond to amber stop codon suppression by glutamyl-tRNA.

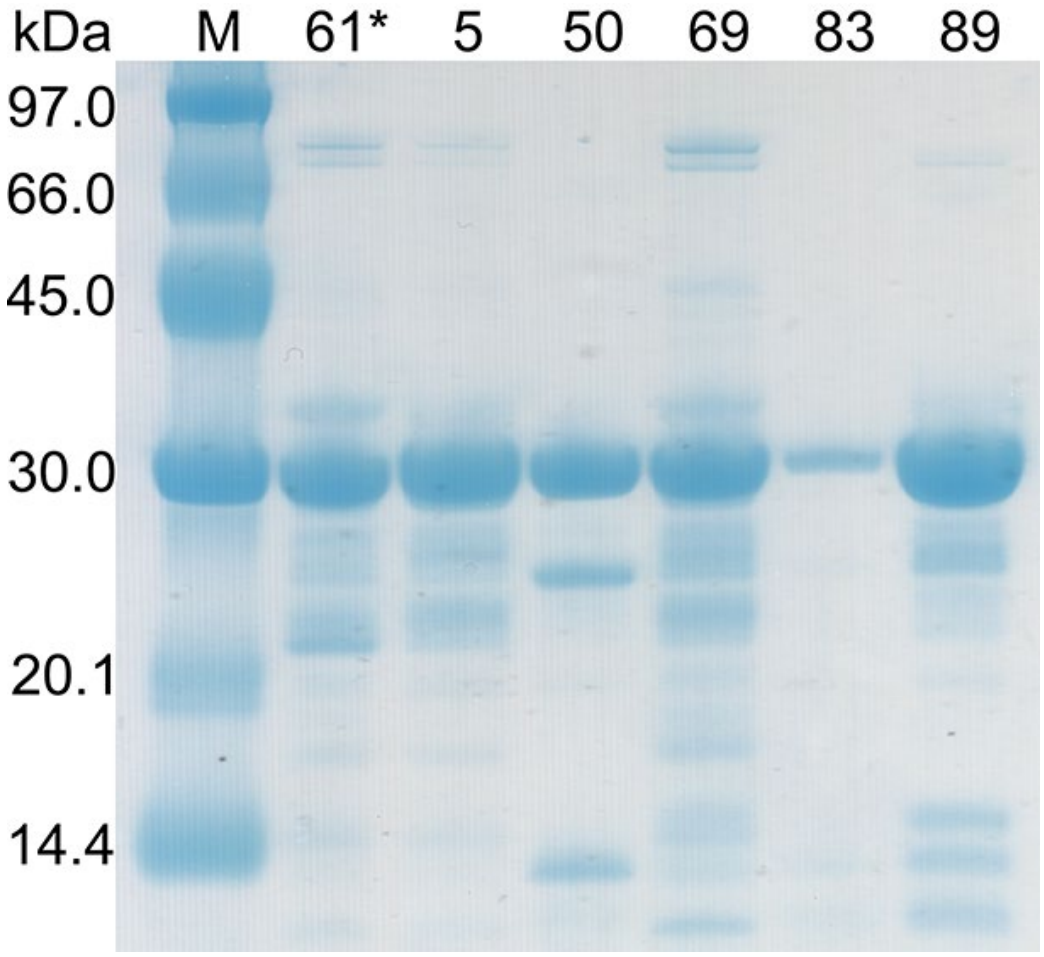


**Figure S6.** SDS-PAGE analysis of ZiPro mutants with 7FTrp: lane 1, M, protein molecular weight marker (the molecular weight of each band is indicated on the left); lanes 2-7, mutant proteins annotated with the sequence number of the respective mutation sites. Mutant proteins were produced in LB medium supplemented with 1 mM 7FI, purified in a single step on a His GraviTrap column and analyzed on a Bolt 10% Bis-Tris gel (Thermo Fisher Scientific, USA), stained with Bio-Safe Coomassie blue (Bio-Rad Laboratories, USA). The expected molecular weight (following loss of the N-terminal methionine) is 24676.45 Da.

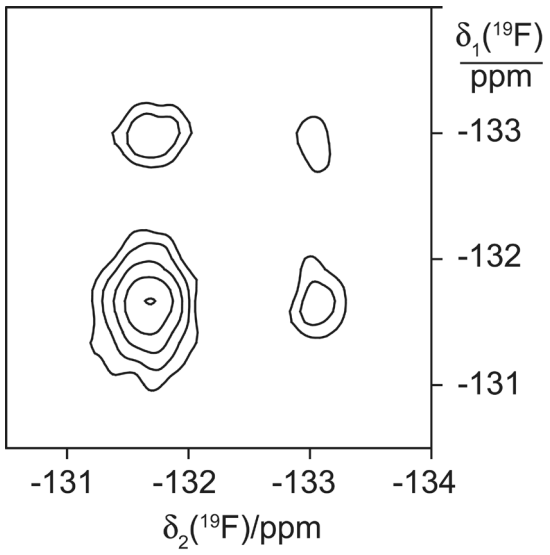


**Figure S7.** Exchange cross-peaks in the 2D $^{19}F–^{19}F$ NOESY spectrum of ZiPro with 7FTrp in position 83. The spectrum was recorded with a mixing time of 200 ms.

## Supplementary tables

**Table S1.** Mutations found in seven selected colonies with G1PylRS variants that recognize 7FTrp.[a]

| Mutant | Site | | | | | | |
|---|---|---|---|---|---|---|---|
| G1PylRS wt | L124 | Y125 | N165 | V167 | Y204 | A221 | W237 |
| G17F02 | G | F | G | F | W | G | Y |
| G17F15 | S | F | A | F | W | G | Y |
| G17F29 | H | F | A | A | W | G | Y |
| G17F30 | H | F | A | A | W | G | Y |
| G17F36 | Q | F | A | F | W | G | Y |
| G17F52 | H | F | A | A | W | G | Y |
| G17F53 | L | Y | A | F | W | G | Y |

[a] Mutation sets highlighted in gray are of the same variant. The colony with mutation set G17F02 (highlighted in red) produced the greatest fluorescence and the G1PylRS isolated from this colony is referred to as G1(7FTrp)RS, which was used for all following applications.

**Table S2.** DNA and corresponding amino acid sequences of the proteins used in the current study.

| Protein[a] | DNA sequence | Amino acid sequence[b] |
|---|---|---|
| RFP 13TAG | ATGGCTTCTATGACCGGTCATCACCATCACCATCACTAGGCCA<br>GTAGTGAAGACGTTATCAAGGAGTTTATGCGTTTCAAAGTACG<br>TATGGAGGGTAGTGTTAACGGACACGAATTTGAGATCGAGGGA<br>GAGGGGGAAGGTCGTCCTTACGAGGGAACTCAAACGGCCAAAT<br>TAAAGGTGACCAAAGGTGGGCCCTTGCCATTCGCGTGGGACAT<br>CTTGTCACCCCAGTTCCAGTACGGGTCGAAGGCATACGTAAAA<br>CACCCAGCGGACATTCCTGACTATCTTAAGTTATCTTTCCCGG<br>AAGGTTTTAAATGGGAACGCGTGATGAACTTTGAGGATGGGGG<br>GGTTGTTACGGTGACACAAGACTCCTCATTGCAAGATGGAGAG<br>TTTATCTATAAAGTCAAACTTCGCGGCACCAATTTTCCATCTG<br>ACGGTCCTGTAATGCAGAAAAAAACAATGGGCTGGGAAGCCTC<br>CACAGAACGTATGTACCCCGAAGATGGAGCTTTAAAGGGCGAA<br>ATTAAAATGCGCTTAAAACTTAAAGACGGCGGCCATTACGACG<br>CCGAAGTGAAAACGACGTATATGGCTAAGAAACCCGTCCAGCT<br>TCCGGGAGCCTATAAAACTGACATCAAACTGGATATTACATCA<br>CACAACGAAGATTATACTATTGTCGAACAGTACGAACGCGCCG<br>AAGGCCGCCATTCAACGGGAGCATAA | MASMTGHHHHHHXASSEDVIKEFMR<br>FKVRMEGSVNGHEFEIEGEGEGRPY<br>EGTQTAKLKVTKGGPLPFAWDILSP<br>QFQYGSKAYVKHPADIPDYLKLSFP<br>EGFKWERVMNFEDGGVVTVTQDSSL<br>QDGEFIYKVKLRGTNFPSDGPVMQK<br>KTMGWEASTERMYPEDGALKGEIKM<br>RLKLKDGGHYDAEVKTTYMAKKPVQ<br>LPGAYKTDIKLDITSHNEDYTIVEQ<br>YERAEGRHSTGA |
| ZiPro Trp61*TAG | ATGTCGGTAGATATGTACATCGAACGCGCAGGAGACATTACCT<br>AGGAGAAGGACGCTGAAGTCACTGGCAATTCACCCCGTTTAGA<br>CGTTGCTCTGGATGAATCGGGTGACTTTAGTCTTGTAGAAGAC<br>GATGGTCCTCCAATGGCGGGGGGAGGTGGCTCAGGAGGGGGTG<br>GGTCTGGTGCATTATGGGACGTACCAGCACCTAAGGAAGTAAA<br>AAACGGCGAAACTACAGACGGTGTGTATCGCGTGATGACACGT<br>GGTTTGTTGGGTAGCACACAAGTGGGCGTCGGAGTAATGCAAG<br>AGGGCGTATTTCACACGATGTGGCATGTAACCAAAGGTTCTGC<br>GTTACGCTCCGGCGAAGGACGCCTTGACCCGTATTGGGGCGAC<br>GTCAAGCAAGATTTGGTTAGCTATAGCGGACCTTGGAAACTGG<br>ACGCGGCATGGGACGGCCACTCGGAGGTACAGTTATTGGCCGT<br>TCCGCCAGGAGAGCGTGCTCGTAACATTCAAACGCTGCCTGGT<br>ATTTTCAAAACTAAGGATGGTGATATCGGGGCGGTGGCCTTAG<br>ATTATCCGGCAGGAACATCCGGGTCTCCCATTTTAGACAAGAG<br>TGGTCGCGTGATTGGGTTATATGGGAATGGGGTTGTAATCAAG<br>AATGGATCTTACGTCTCGGCGATCACACAGGGTCGCCGCCATC<br>ACCACCATCATCACTAA | MSVDMYIERAGDITXEKDAEVTGNS<br>PRLDVALDESGDFSLVEDDGPPMAG<br>GGGSGGGGSGALWDVPAPKEVKNGE<br>TTDGVYRVMTRGLLGSTQVGVGVMQ<br>EGVFHTMWHVTKGSALRSGEGRLDP<br>YWGDVKQDLVSYSGPWKLDAAWDGH<br>SEVQLLAVPPGERARNIQTLPGIFK<br>TKDGDIGAVALDYPAGTSGSPILDK<br>SGRVIGLYGNGVVIKNGSYVSAITQ<br>GRRHHHHHH |
| ZiPro Trp5TAG | ATGTCGGTAGATATGTACATCGAACGCGCAGGAGACATTACCT<br>GGGAGAAGGACGCTGAAGTCACTGGCAATTCACCCCGTTTAGA<br>CGTTGCTCTGGATGAATCGGGTGACTTTAGTCTTGTAGAAGAC<br>GATGGTCCTCCAATGGCGGGGGGAGGTGGCTCAGGAGGGGGTG<br>GGTCTGGTGCATTATAGGACGTACCAGCACCTAAGGAAGTAAA<br>AAACGGCGAAACTACAGACGGTGTGTATCGCGTGATGACACGT<br>GGTTTGTTGGGTAGCACACAAGTGGGCGTCGGAGTAATGCAAG<br>AGGGCGTATTTCACACGATGTGGCATGTAACCAAAGGTTCTGC<br>GTTACGCTCCGGCGAAGGACGCCTTGACCCGTATTGGGGCGAC<br>GTCAAGCAAGATTTGGTTAGCTATAGCGGACCTTGGAAACTGG<br>ACGCGGCATGGGACGGCCACTCGGAGGTACAGTTATTGGCCGT<br>TCCGCCAGGAGAGCGTGCTCGTAACATTCAAACGCTGCCTGGT<br>ATTTTCAAAACTAAGGATGGTGATATCGGGGCGGTGGCCTTAG<br>ATTATCCGGCAGGAACATCCGGGTCTCCCATTTTAGACAAGAG<br>TGGTCGCGTGATTGGGTTATATGGGAATGGGGTTGTAATCAAG<br>AATGGATCTTACGTCTCGGCGATCACACAGGGTCGCCGCCATC<br>ACCACCATCATCACTAA | MSVDMYIERAGDITWEKDAEVTGNS<br>PRLDVALDESGDFSLVEDDGPPMAG<br>GGGSGGGGSGALXDVPAPKEVKNGE<br>TTDGVYRVMTRGLLGSTQVGVGVMQ<br>EGVFHTMWHVTKGSALRSGEGRLDP<br>YWGDVKQDLVSYSGPWKLDAAWDGH<br>SEVQLLAVPPGERARNIQTLPGIFK<br>TKDGDIGAVALDYPAGTSGSPILDK<br>SGRVIGLYGNGVVIKNGSYVSAITQ<br>GRRHHHHHH |

| | | |
|---|---|---|
| ZiPro Trp50TAG | ATGTCGGTAGATATGTACATCGAACGCGCAGGAGACATTACCT<br>GGGAGAAGGACGCTGAAGTCACTGGCAATTCACCCCGTTTAGA<br>CGTTGCTCTGGATGAATCGGGTGACTTTAGTCTTGTAGAAGAC<br>GATGGTCCTCCAATGGCGGGGGGAGGTGGCTCAGGAGGGGGTG<br>GGTCTGGTGCATTATGGGACGTACCAGCACCTAAGGAAGTAAA<br>AAACGGCGAAACTACAGACGGTGTGTATCGCGTGATGACACGT<br>GGTTTGTTGGGTAGCACACAAGTGGGCGTCGGAGTAATGCAAG<br>AGGGCGTATTTCACACGATGTAGCATGTAACCAAAGGTTCTGC<br>GTTACGCTCCGGCGAAGGACGCCTTGACCCGTATTGGGGCGAC<br>GTCAAGCAAGATTTGGTTAGCTATAGCGGACCTTGGAAACTGG<br>ACGCGGCATGGGACGGCCACTCGGAGGTACAGTTATTGGCCGT<br>TCCGCCAGGAGAGCGTGCTCGTAACATTCAAACGCTGCCTGGT<br>ATTTTCAAAACTAAGGATGGTGATATCGGGGCGGTGGCCTTAG<br>ATTATCCGGCAGGAACATCCGGGTCTCCCATTTTAGACAAGAG<br>TGGTCGCGTGATTGGGTTATATGGGAATGGGGTTGTAATCAAG<br>AATGGATCTTACGTCTCGGCGATCACACAGGGTCGCCGCCATC<br>ACCACCATCATCACTAA | MSVDMYIERAGDITWEKDAEVTGNS<br>PRLDVALDESGDFSLVEDDGPPMAG<br>GGGSGGGGSGALWDVPAPKEVKNGE<br>TTDGVYRVMTRGLLGSTQVGVGVMQ<br>EGVFHTMXHVTKGSALRSGEGRLDP<br>YWGDVKQDLVSYSGPWKLDAAWDGH<br>SEVQLLAVPPGERARNIQTLPGIFK<br>TKDGDIGAVALDYPAGTSGSPILDK<br>SGRVIGLYGNGVVIKNGSYVSAITQ<br>GRRHHHHHH |
| ZiPro Trp69TAG | ATGTCGGTAGATATGTACATCGAACGCGCAGGAGACATTACCT<br>GGGAGAAGGACGCTGAAGTCACTGGCAATTCACCCCGTTTAGA<br>CGTTGCTCTGGATGAATCGGGTGACTTTAGTCTTGTAGAAGAC<br>GATGGTCCTCCAATGGCGGGGGGAGGTGGCTCAGGAGGGGGTG<br>GGTCTGGTGCATTATGGGACGTACCAGCACCTAAGGAAGTAAA<br>AAACGGCGAAACTACAGACGGTGTGTATCGCGTGATGACACGT<br>GGTTTGTTGGGTAGCACACAAGTGGGCGTCGGAGTAATGCAAG<br>AGGGCGTATTTCACACGATGTGGCATGTAACCAAAGGTTCTGC<br>GTTACGCTCCGGCGAAGGACGCCTTGACCCGTATTAGGGCGAC<br>GTCAAGCAAGATTTGGTTAGCTATAGCGGACCTTGGAAACTGG<br>ACGCGGCATGGGACGGCCACTCGGAGGTACAGTTATTGGCCGT<br>TCCGCCAGGAGAGCGTGCTCGTAACATTCAAACGCTGCCTGGT<br>ATTTTCAAAACTAAGGATGGTGATATCGGGGCGGTGGCCTTAG<br>ATTATCCGGCAGGAACATCCGGGTCTCCCATTTTAGACAAGAG<br>TGGTCGCGTGATTGGGTTATATGGGAATGGGGTTGTAATCAAG<br>AATGGATCTTACGTCTCGGCGATCACACAGGGTCGCCGCCATC<br>ACCACCATCATCACTAA | MSVDMYIERAGDITWEKDAEVTGNS<br>PRLDVALDESGDFSLVEDDGPPMAG<br>GGGSGGGGSGALWDVPAPKEVKNGE<br>TTDGVYRVMTRGLLGSTQVGVGVMQ<br>EGVFHTMWHVTKGSALRSGEGRLDP<br>YXGDVKQDLVSYSGPWKLDAAWDGH<br>SEVQLLAVPPGERARNIQTLPGIFK<br>TKDGDIGAVALDYPAGTSGSPILDK<br>SGRVIGLYGNGVVIKNGSYVSAITQ<br>GRRHHHHHH |
| ZiPro Trp83TAG | ATGTCGGTAGATATGTACATCGAACGCGCAGGAGACATTACCT<br>GGGAGAAGGACGCTGAAGTCACTGGCAATTCACCCCGTTTAGA<br>CGTTGCTCTGGATGAATCGGGTGACTTTAGTCTTGTAGAAGAC<br>GATGGTCCTCCAATGGCGGGGGGAGGTGGCTCAGGAGGGGGTG<br>GGTCTGGTGCATTATGGGACGTACCAGCACCTAAGGAAGTAAA<br>AAACGGCGAAACTACAGACGGTGTGTATCGCGTGATGACACGT<br>GGTTTGTTGGGTAGCACACAAGTGGGCGTCGGAGTAATGCAAG<br>AGGGCGTATTTCACACGATGTGGCATGTAACCAAAGGTTCTGC<br>GTTACGCTCCGGCGAAGGACGCCTTGACCCGTATTGGGGCGAC<br>GTCAAGCAAGATTTGGTTAGCTATAGCGGACCTTAGAAACTGG<br>ACGCGGCATGGGACGGCCACTCGGAGGTACAGTTATTGGCCGT<br>TCCGCCAGGAGAGCGTGCTCGTAACATTCAAACGCTGCCTGGT<br>ATTTTCAAAACTAAGGATGGTGATATCGGGGCGGTGGCCTTAG<br>ATTATCCGGCAGGAACATCCGGGTCTCCCATTTTAGACAAGAG<br>TGGTCGCGTGATTGGGTTATATGGGAATGGGGTTGTAATCAAG<br>AATGGATCTTACGTCTCGGCGATCACACAGGGTCGCCGCCATC<br>ACCACCATCATCACTAA | MSVDMYIERAGDITWEKDAEVTGNS<br>PRLDVALDESGDFSLVEDDGPPMAG<br>GGGSGGGGSGALWDVPAPKEVKNGE<br>TTDGVYRVMTRGLLGSTQVGVGVMQ<br>EGVFHTMWHVTKGSALRSGEGRLDP<br>YWGDVKQDLVSYSGPXKLDAAWDGH<br>SEVQLLAVPPGERARNIQTLPGIFK<br>TKDGDIGAVALDYPAGTSGSPILDK<br>SGRVIGLYGNGVVIKNGSYVSAITQ<br>GRRHHHHHH |

| | | |
|---|---|---|
| ZiPro Trp89TAG | ATGTCGGTAGATATGTACATCGAACGCGCAGGAGACATTACCT<br>GGGAGAAGGACGCTGAAGTCACTGGCAATTCACCCCGTTTAGA<br>CGTTGCTCTGGATGAATCGGGTGACTTTAGTCTTGTAGAAGAC<br>GATGGTCCTCCAATGGCGGGGGGAGGTGGCTCAGGAGGGGGTG<br>GGTCTGGTGCATTATGGGACGTACCAGCACCTAAGGAAGTAAA<br>AAACGGCGAAACTACAGACGGTGTGTATCGCGTGATGACACGT<br>GGTTTGTTGGGTAGCACACAAGTGGGCGTCGGAGTAATGCAAG<br>AGGGCGTATTTCACACGATGTGGCATGTAACCAAAGGTTCTGC<br>GTTACGCTCCGGCGAAGGACGCCTTGACCCGTATTGGGGCGAC<br>GTCAAGCAAGATTTGGTTAGCTATAGCGGACCTTGGAAACTGG<br>ACGCGGCATAGGACGGCCACTCGGAGGTACAGTTATTGGCCGT<br>TCCGCCAGGAGAGCGTGCTCGTAACATTCAAACGCTGCCTGGT<br>ATTTTCAAAACTAAGGATGGTGATATCGGGGCGGTGGCCTTAG<br>ATTATCCGGCAGGAACATCCGGGTCTCCCATTTTAGACAAGAG<br>TGGTCGCGTGATTGGGTTATATGGGAATGGGGTTGTAATCAAG<br>AATGGATCTTACGTCTCGGCGATCACACAGGGTCGCCGCCATC<br>ACCACCATCATCACTAA | MSVDMYIERAGDITWEKDAEVTGNS<br>PRLDVALDESGDFSLVEDDGPPMAG<br>GGGSGGGGSGALWDVPAPKEVKNGE<br>TTDGVYRVMTRGLLGSTQVGVGVMQ<br>EGVFHTMWHVTKGSALRSGEGRLDP<br>YWGDVKQDLVSYSGPWKLDAAXDGH<br>SEVQLLAVPPGERARNIQTLPGIFK<br>TKDGDIGAVALDYPAGTSGSPILDK<br>SGRVIGLYGNGVVIKNGSYVSAITQ<br>GRRHHHHHH |

[a] A star marks mutation sites in NS2B.

[b] X indicates the positions of 7FTrp.

**Table S3.** Yields of different ZiPro amber mutants co-expressed with G1(7FTrp)RS in LB media supplemented with 1 mM 7FI.

| Amber site | Yield (mg per 1 L cell culture) |
|---|---|
| wild-type | 70.0 |
| Trp61* | 8.9 |
| Trp5 | 9.5 |
| Trp50 | 3.8 |
| Trp69 | 6.3 |
| Trp83 | 5.1 |
| Trp89 | 6.0 |

## References


1. Abdelkader, E. H.; Qianzhu, H.; George, J.; Frkic, R. L.; Jackson, C. J.; Nitsche, C.; Otting, G.; Huber, T., Genetic encoding of cyanopyridylalanine for in-cell protein macrocyclization by the nitrile-aminothiol click reaction. *ChemRxiv*, October 27, 2021, ver. 1. DOI: 10.33774/chemrxiv-2021-r5dn6.

2. Mukai, T.; Hoshi, H.; Ohtake, K.; Takahashi, M.; Yamaguchi, A.; Hayashi, A.; Yokoyama, S.; Sakamoto, K., Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon. *Sci. Rep.* **2015,** *5*, 9699.

3. Jiang, H.-K.; Wang, Y.-H.; Weng, J.-H.; Kurkute, P.; Li, C.-L.; Lee, M.-N.; Chen, P.-J.; Tseng, H.-W.; Tsai, M.-D.; Wang, Y.-S., Probing the active site of deubiquitinase USP30 with noncanonical tryptophan analogues. *Biochemistry* **2020,** *59*, 2205–2209.