Supporting Information

Genetic Encoding of 7-Aza-L-Tryptophan: Isoelectronic Substitution of a Single CH-Group in a Protein for a Nitrogen Atom for Site-Selective Isotope Labeling

Elwy H. Abdelkader, Haocheng Qianzhu, Thomas Huber,* Gottfried Otting*

Table of contents:

Expression and purification of *Tm*9D8* TrpB

*Tm*9D8* TrpB[1] was produced in *E. coli* BL21(DE3) cells transformed with the plasmid pET-21(+)*Tm*9D8* TrpB (Twist Bioscience, USA). The transformed cells were grown at 37 °C in LB medium containing 100 mg/L carbenicillin. An aliquot (10 mL) of the overnight culture was used to inoculate 1 L LB medium supplemented with 100 mg/L carbenicillin. The cells were grown at 37 °C to an OD_{600} of 0.6. At this point, the temperature was reduced to 25 °C and protein expression induced by the addition of 1 mM IPTG.

The cells were harvested after 16 h by centrifugation. Following resuspension in buffer A, the cells were lysed as described above. The cell lysate was then heated at 75 °C for 30 min in a water bath. The heat-treated lysate was centrifuged for 1 h at 30,000 *g*. The supernatant was loaded onto a 5 mL Ni-NTA HisTrap column connected to an ÄKTA pure 25 chromatography system (Cytiva, USA). The column was washed with 20 column volumes buffer A and the protein was eluted with 5 column volumes buffer B. Afterwards, the buffer of the eluted protein was exchanged to 50 mM potassium phosphate buffer, pH 8.0, using an Amicon ultrafiltration centrifugal tube with a molecular weight cut-off of 10 kDa. The yield of purified *Tm*9D8* TrpB was 10 mg/L cell culture.

Enzymatic synthesis of isotope-labeled 7AW (Scheme 1)

In a 50 mL centrifuge tube, 99 mg (0.92 mmol) $^{13}C/^{15}N$-labeled serine (^{13}C-labeled only in the C1 and C2 positions; Martek, USA) and 4.8 mg (0.02 mmol) pyridoxal 5′-phosphate (Merck, Germany) were dissolved in 10 mL 50 mM potassium phosphate buffer, pH 8.0. The solution was then heated to 55 ºC. Next, 97 mg (0.82 mmol) 7-azaindole (Ambeed, USA) dissolved in 0.5 mL ethanol were added and the solution kept at 55 ºC. The enzymatic reaction was initiated by the addition of *Tm*9D8* TrpB (20 μM final concentration) and the reaction volume adjusted to 15 mL with 50 mM potassium phosphate buffer. The reaction mixture was shaken during incubation at 55 ºC. To monitor the reaction progress, 20 μL samples of the reaction mixture were removed, diluted to 0.5 mL with PBS buffer, and analyzed by ^{1}H NMR. After 72 h, the reaction mixture was cooled down to room temperature, and the pH was adjusted to 1 using concentrated HCl, to ensure complete dissolution of 7AW. Milli-Q water was added to adjust the volume to 20 mL to produce a 40 mM stock solution of the isotope-labeled 7AW. This stock solution was used in the *in vivo* expression experiments without further processing.

Selection of functional G1PylRS enzymes recognizing 7AW

The selection of G1PylRS enzymes recognizing 7AW used the previously constructed library plasmid pBK-G1RS[2] transformed into *E. coli* DH10B cells that also harbored the selection plasmid pBAD-H6RFP. Following recovery from transformation, the culture was directly inoculated into two flasks, each with 25 mL LB medium containing 100 mg/L carbenicillin, 50 mg/L kanamycin, and 0.4% L-arabinose. One of the cultures was supplied with 1 mM 7AW (Ambeed, USA), which served as the sample for the first round of positive selection (**1P+**). The other culture omitted 7AW and served as control (**1P–**). Overnight expression at 37 °C resulted in a readily detectable level of RFP expression. 0.4 mL of cells were resuspended in 8 mL PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) yielding a concentration suitable for cell sorting by fluorescence activated cell sorting (FACS) on an Aria Fusion high speed cell sorter (BD Biosciences, USA; **Figure S1**).

Cells with high RFP levels were selected from the **1P+** sample (0.4% of the total population), collecting 9.5×10^4 cells, which were subsequently subjected to a round of negative selection without the addition of 7AW and regrown as sample **2N–**. From this sample, 1×10^6 cells with low RFP expression levels (41%) were collected and aliquoted to inoculate media with positive (**3P+**) and negative (**3P–**) conditions. The RFP-positive cells (15.9%, 4×10^5 cells) from the **3P+** sample were collected and recovered under negative condition to obtain the sample **4N–**. Following sorting, cells showing the lowest level of RFP fluorescence (49%, 4×10^5 cells) were selected from the **4N–** sample. They were aliquoted to be recovered as the samples **5P+** and **5P–**, respectively. 43.4% of the cell population in the **5P+** sample produced high RFP fluorescence with 7AW provided compared to 0.8% without the ncAA, indicating the successful accumulation of active G1PylRS variants specific for 7AW. 1.5×10^5 cells with the highest RFP fluorescence, representing 3% of the **5P+** sample, were recovered for storage. An aliquot of 1,000 cells was 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 1 mM 7AW) and negative (without ncAA) 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 (**Figure S2**). Six candidates with the best efficiency and specificity were sequenced. Four individually different sequences were found (**Table S1**). 7AW04 was cloned into the pRSF plasmid to yield pRSF-G1(7AW)RS.

Expression, purification, and cleavage of NT* domain with C-terminal tripeptide motif

E. coli B-95.ΔA cells[3] were co-transformed with pRSF-G1(7AW)RS and the pCDF plasmid containing the gene encoding the NT* domain with C-terminal tripeptide motif. The transformed cells were grown at 37 °C in LB medium containing 25 mg/L kanamycin and 25 mg/L spectinomycin. An aliquot (1 mL) of the overnight culture was used to inoculate 100 mL LB medium supplemented with 25 mg/L kanamycin, 25 mg/L spectinomycin and 1 mM 7AW. 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.

After 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 by sonication (ultrasonic homogenizer Omni-Ruptor 4000, Omni International, USA) on ice (50% power and 50% pulse length for 10 min). The cell lysate was 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 PBS using an Amicon ultrafiltration centrifugal tube (Merck Millipore, Germany) with a molecular weight cut-off of 10 kDa.

To release the C-terminal tripeptide containing 7AW, 0.5 mL NT* fusion was incubated with His_6-TEV protease (10:1 by mass) at 25 °C for 4 h. Afterwards, 0.5 mL Ni Sepharose 6 Fast Flow resin (Cytiva, USA) was added to capture the His_6-tagged proteins, leaving the tripeptide in the supernatant, which was collected by centrifugation at 21,000 *g* for 5 min. This step was repeated three more times. To the supernatant was added D_2O to a final concentration of 10% for analysis by ^1H NMR.

Expression and purification of the Zika virus NS2B-NS3 with 7AW

Zika virus NS2B-NS3 protease samples were produced like the NT* fusion, but the LB media used were supplemented with isotope-labeled 7AW.

Cell-free protein synthesis (CFPS) of ZiPro selectively labeled with $^{15}N/^{13}C$-labeled tryptophan

The CFPS reaction was conducted at 30 °C for 16 h in a dialysis system following a published protocol,[4] where the unlabeled tryptophan was excluded from the amino acid mixture and 1 mM $^{15}N/^{13}C$-labeled tryptophan (Merck, USA) was provided in the outer buffer.

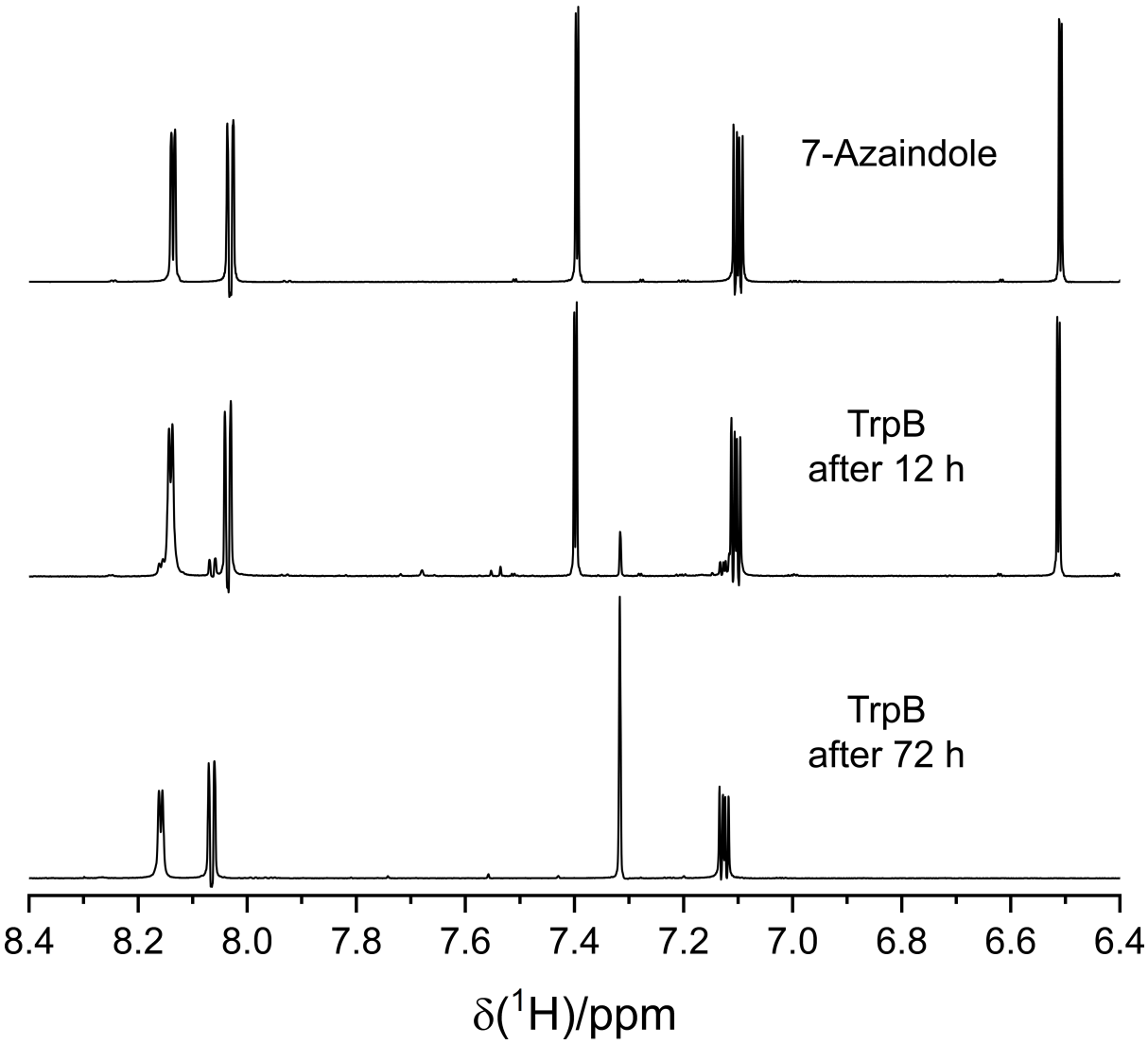

Figure S1. Enzymatic synthesis of 7AW by *Tm*9D8* TrpB from unlabeled 7-azaindole and isotope-labeled serine, monitored by ^{1}H NMR. After 12 h, the aromatic region of the spectrum predominantly shows signals from 7-azaindole, while peaks of 7AW are visible but small. After 72 h, the spectrum shows only signals of 7AW, while the resonances of 7-azaindole have disappeared. The spectra were recorded with a double-spin echo sequence for suppression of the water resonance, which leads to small phase distortions of multiplets split by scalar couplings.[5]

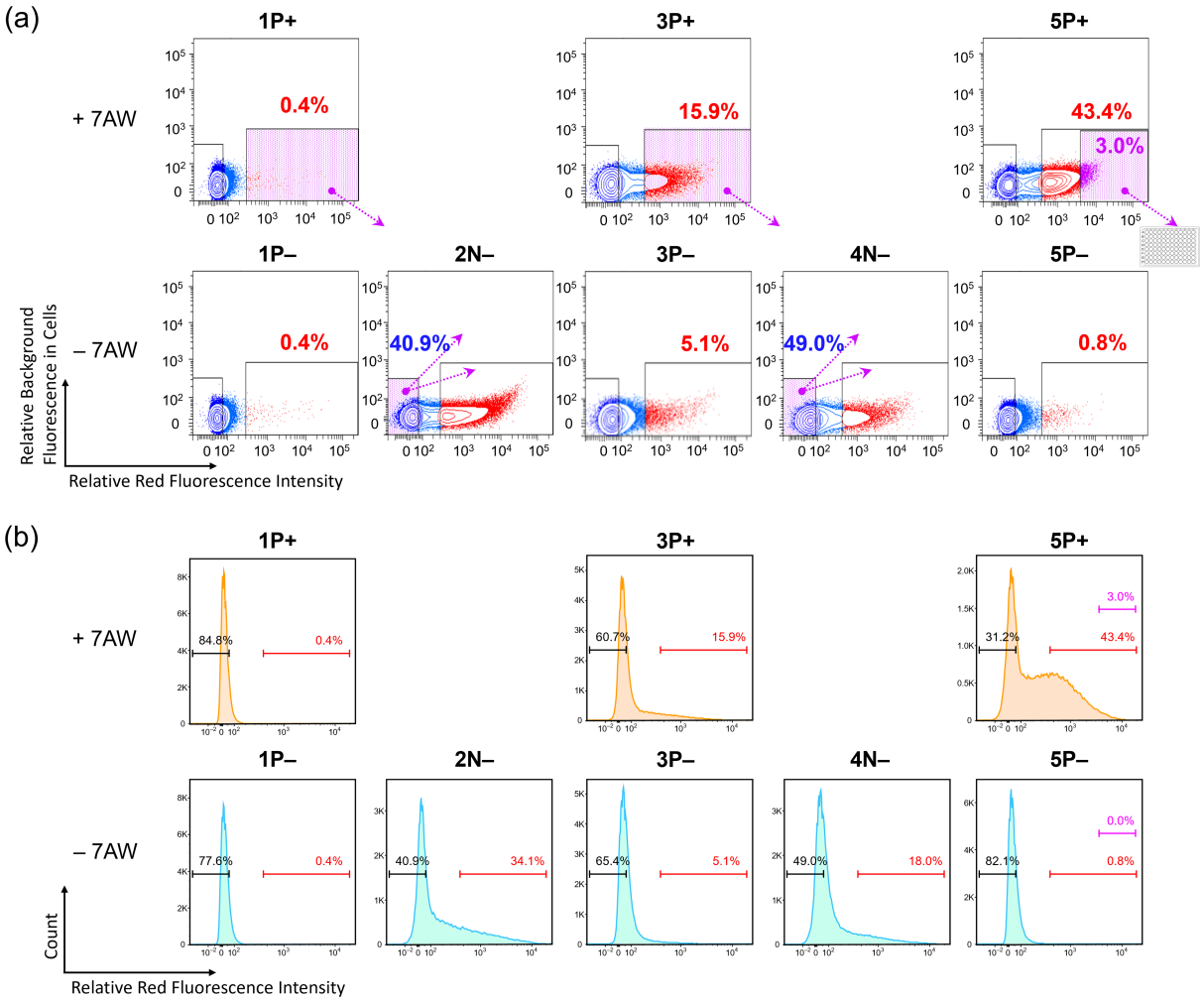

Figure S2. FACS experiments for selection of active G1PylRS enzymes specific for 7AW. (a) Plots of background fluorescence in individual cells excited at 488 nm versus red fluorescence intensity. Violet shades and arrows identify, respectively, the cell populations collected and the following selection scheme, which the selected cells were subjected to following amplification by culturing. (b) Histograms of five selection rounds, plotting the cell count versus the relative intensity of red fluorescence. The difference in RFP fluorescence intensity of cells grown in the presence or absence of 7AW was used as indicator of 7AW-specific RS enzymes in the gene pool.

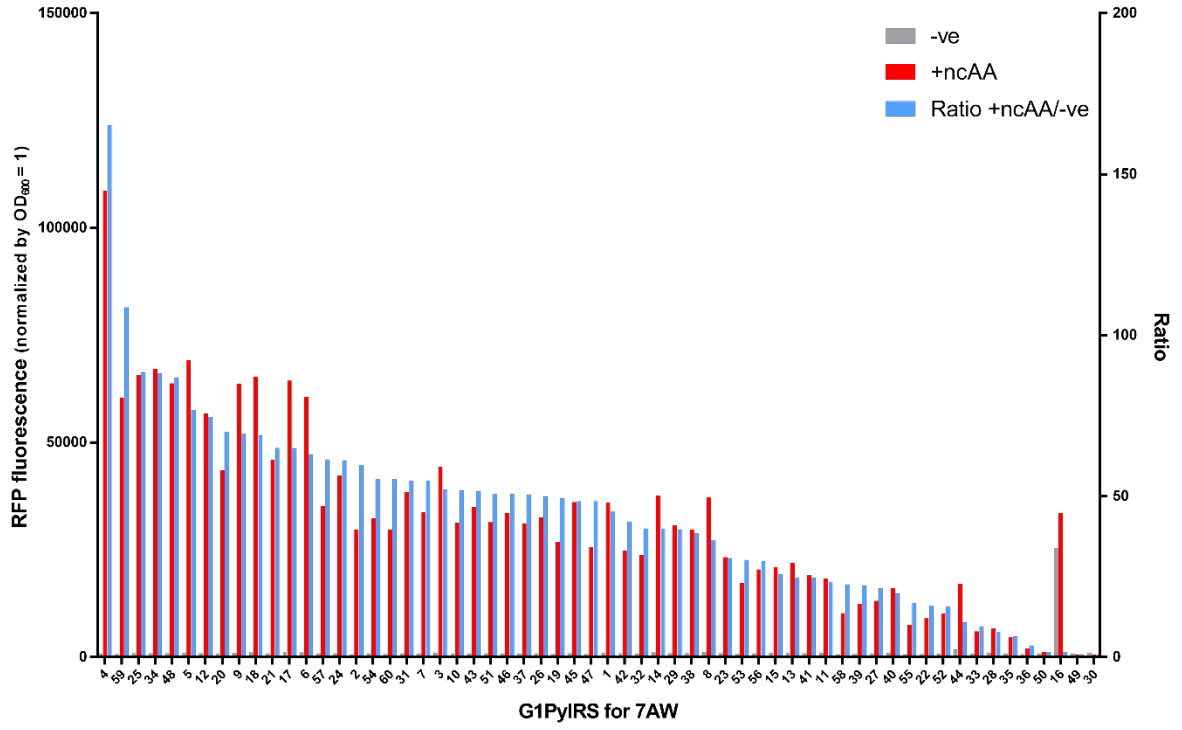

Figure S3. Screening of individual colonies for the activity of G1PylRS for 7AW incorporation. The cells collected from the 3% fraction with the highest red fluorescence in the **5P+** sample were cultured in 96-well plates with and without 1 mM 7AW, and the intensity of red fluorescence was measured as an indicator of the expression level of amber-interrupted RFP. The plot displays the colonies sorted by descending ratio of red fluorescence readouts from +ncAA wells versus –ncAA wells.

Table S1. Mutations found in G1PylRS variants selected to recognize 7AW.[a]

RS Variants	*Randomized Sites*						
*Mm*PylRS-wt	L305	Y306	N346	V348	Y384	V401	W417
*G1*PylRS-wt	L124	Y125	N165	V167	Y204	A221	W237
7AW04	M	M	A	G	Y	G	H
7AW05	L	Y	A	A	Y	G	S
7AW12	L	Y	A	S	Y	G	H
7AW25	L	Y	A	S	Y	G	H
7AW48	H	Y	A	S	Y	G	S
7AW59	L	Y	A	S	Y	G	H

[a] Gray background highlights the amino acids selected for randomization in the wild-type RS. The orange background marks identical variants found in different colonies.

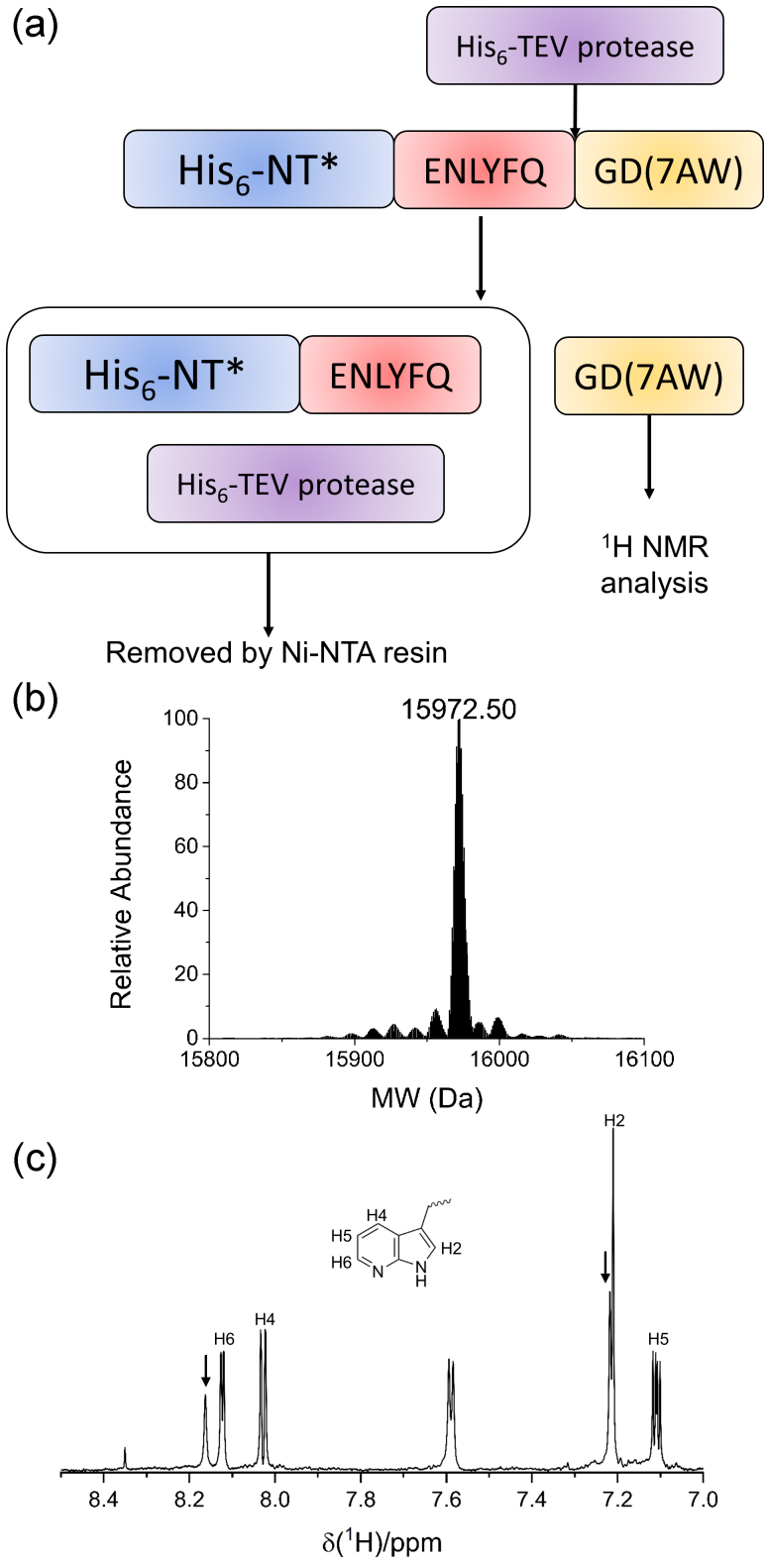

Figure S4. Experiment showing the potential of the selected G1PylRS mutant to discriminate between 7AW and tryptophan. (a) NT* fusion construct used. (b) Intact protein mass spectrometric analysis of 7AW incorporation into the NT* fusion (calculated mass 15972.80 Da). (c) 1D ^{1}H-NMR spectrum of the GD(7AW) tripeptide following release from the NT* fusion by TEV protease. The arrows identify peaks of residual imidazole remaining after Ni-NTA purification. The doublet at 7.59 ppm is of a backbone amide proton.

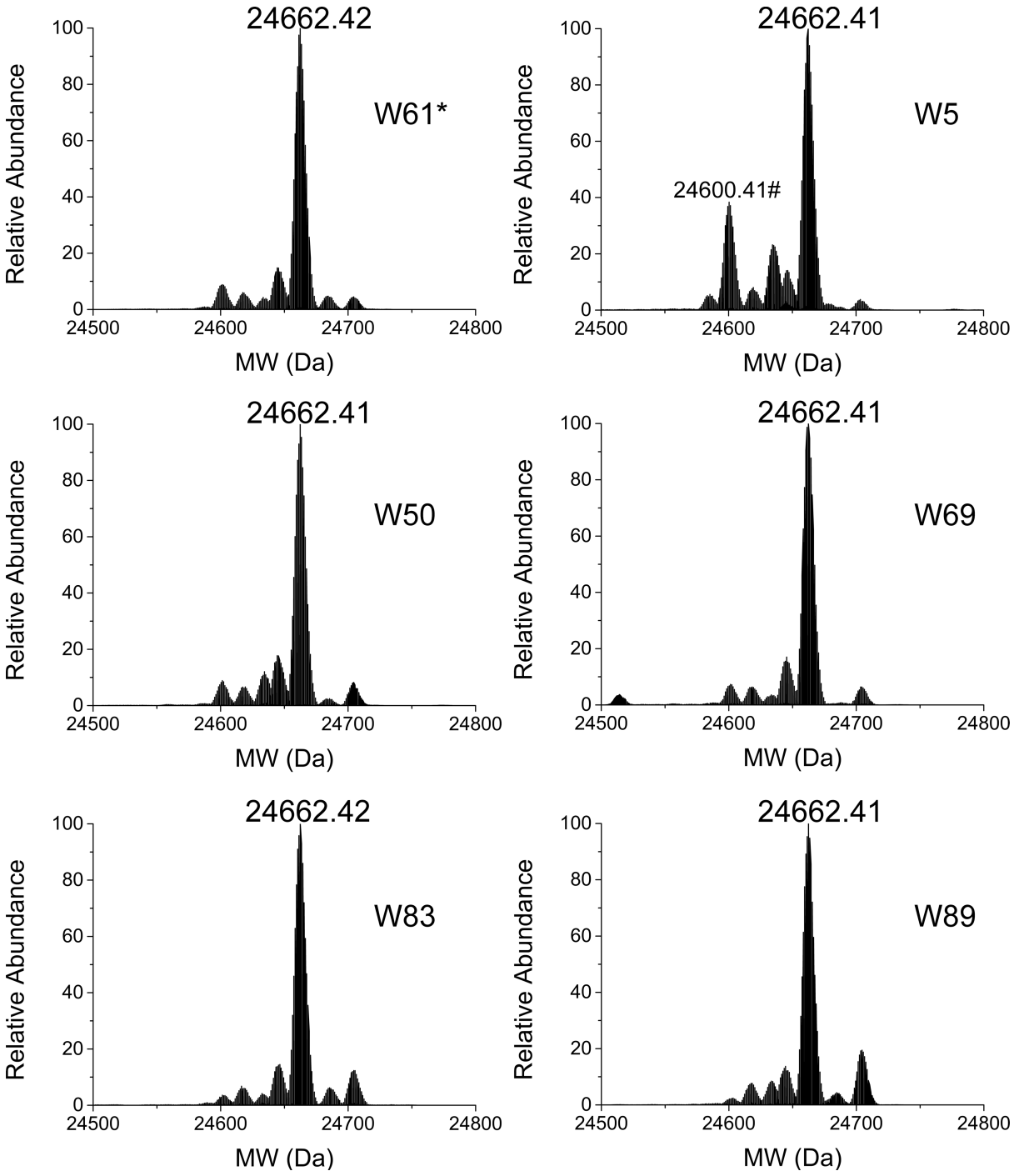

Figure S5. Intact protein mass spectrometric analysis of ZiPro mutants with $^{15}N/^{13}C$-labeled 7AW. The calculated mass (following loss of the N-terminal methionine) is 24662.44 Da. The spectra are annotated with the sequence number of the respective mutation sites. The mass of the peaks marked with a "#" symbol in correspond to amber stop codon suppression by glutamyl-tRNA.

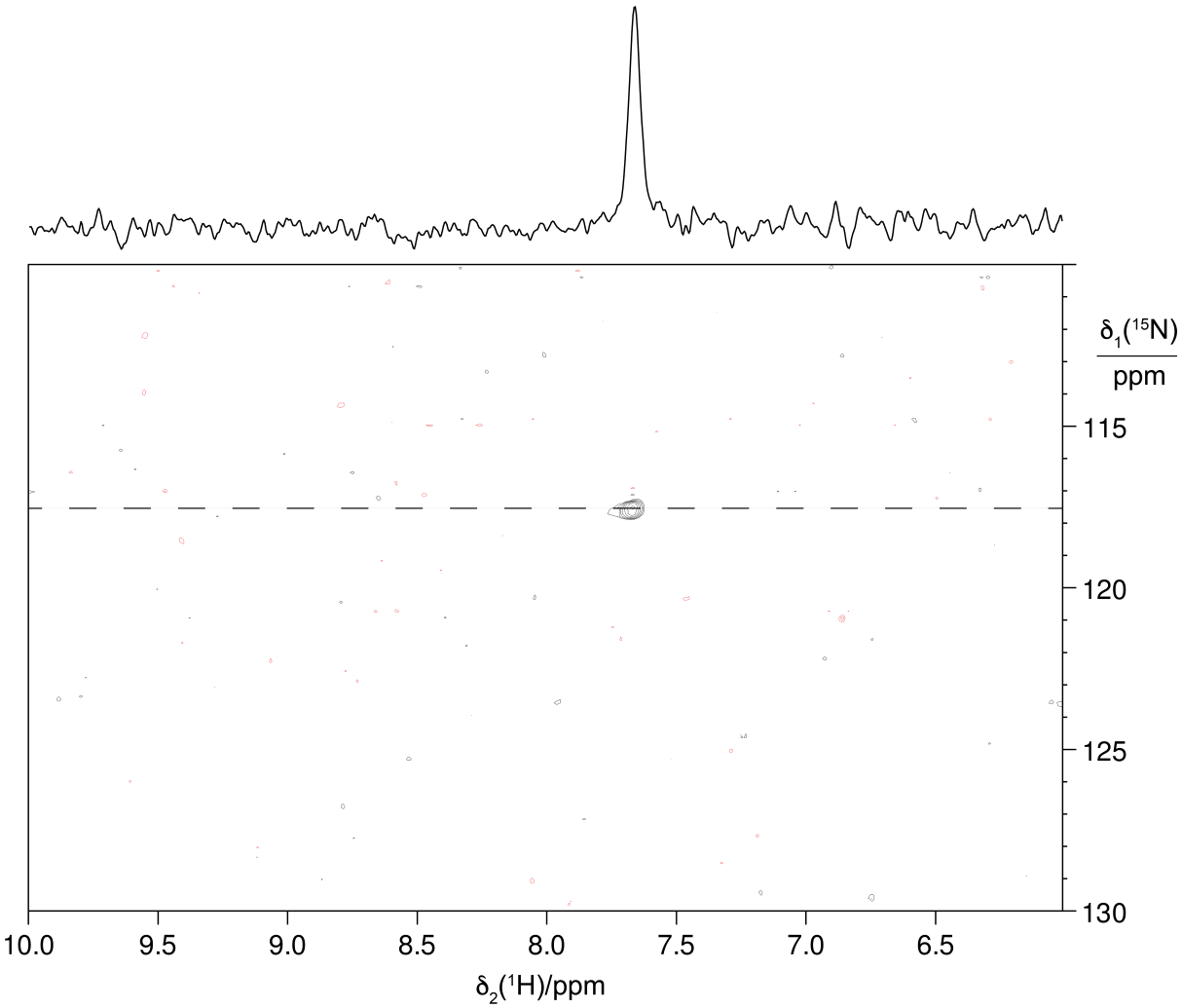

Figure S6. ^{15}N-HSQC spectrum of ZiPro Trp83_7AW (Figure 2e of the main text) illustrating the selectivity of 7AW incorporation. The lowest contour lines of the spectrum were chosen to display the level of white noise. The 1D cross-section through the cross-peak, taken at the dashed line, is shown above the 2D plot. The signal-to-noise ratio of the 7AW cross-peak (about 10:1) and complete absence of any other cross-peaks indicates that misincorporation of 7AW in response to the native tryptophan codons is below 10%.

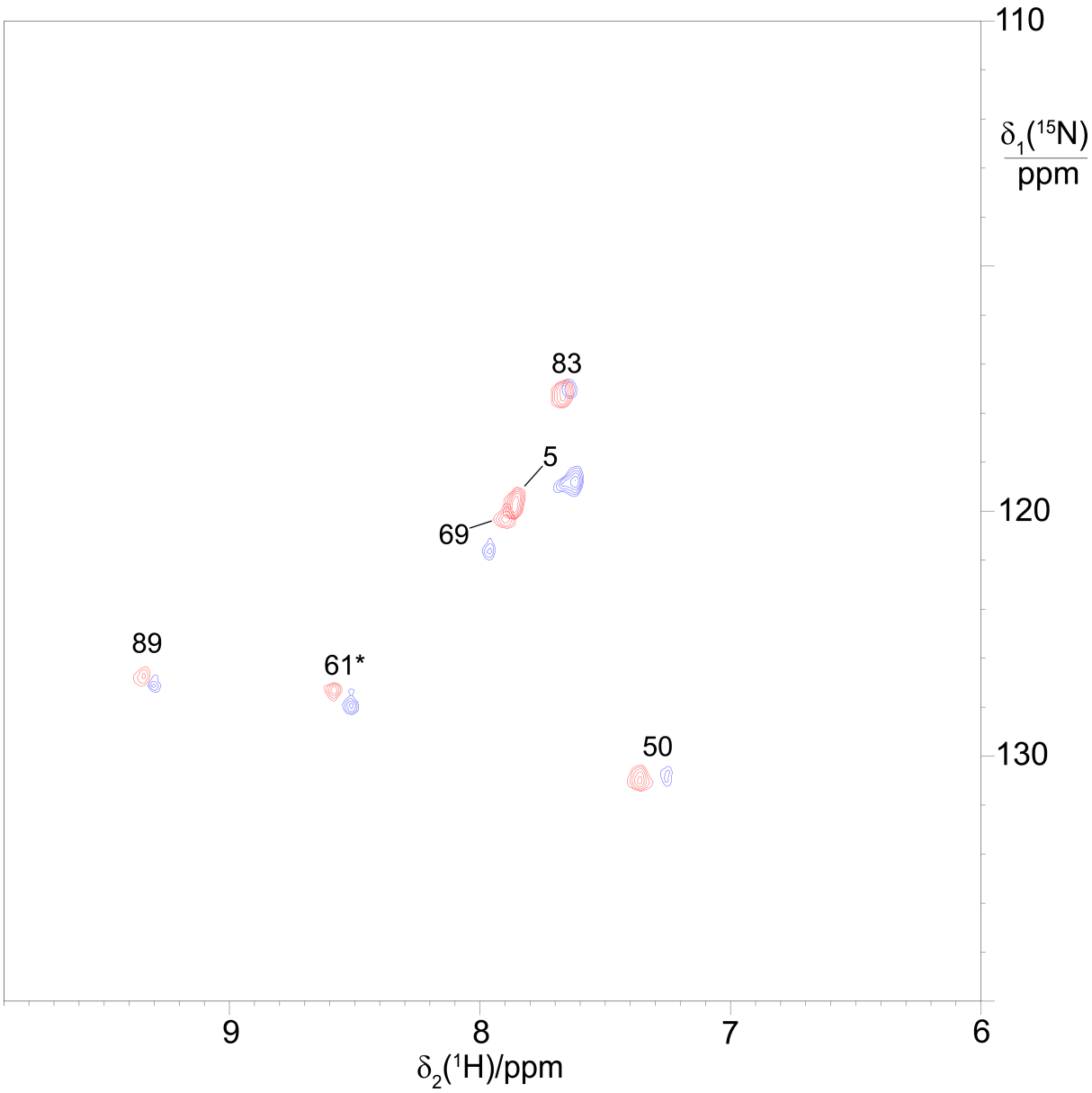

Figure S7. Superimposition of the ^{15}N-HSQC spectra shown in Figure 2 of the main text. The cross-peaks are of backbone amides of the wild-type construct of ZiPro labeled with ^{15}N/^{13}C-labeled tryptophan (blue) and the 7AW mutants (red).

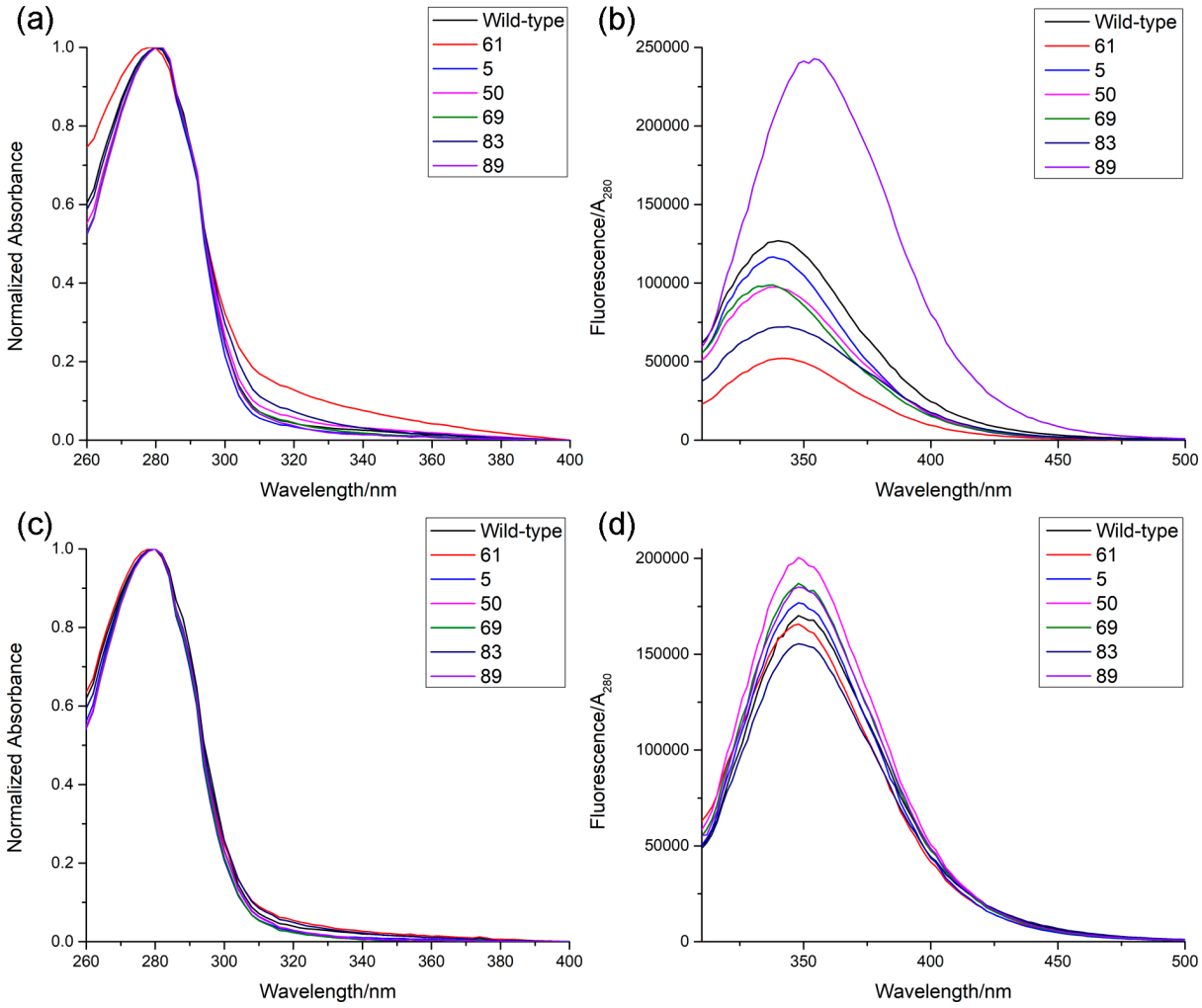

Figure S8. UV absorption and emission fluorescence spectra of the Zika virus NS2B-NS3 7AW mutants of the present work. All measurements were carried out at 25 °C in NMR buffer (20 mM MES, pH 6.5, 150 mM NaCl) at 20 μM protein concentrations. Sample excitation was at 280 nm, using a TECAN Infinite 200 Pro M Plex plate reader. (a) Absorption spectra of native samples. (b) Fluorescence emission spectra of native samples. (c) Absorption spectra in 6 M urea. (d) Fluorescence spectra in 6 M urea.

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

Protein	DNA sequence	Amino acid sequence[a]
NT* C-terminal tripeptide motif	ATGCATCATCATCATCACCACAGCCATACCACACCGTGGACCA ATCCTGGTCTGGCAGAAAACTTTATGAATAGCTTTATGCAGGG TCTGAGCAGCATGCCTGGTTTTACCGCAAGCCAGCTGGACAAA ATGAGCACCATTGCACAGAGCATGGTTCAGAGCATTCAGAGCC TGGCAGCACAGGGTCGTACCAGTCCGAATGATCTGCAGGCACT GAATATGGCATTTGCAAGCAGCATGGCAGAAATTGCAGCAAGC GAAGAAGGTGGCGGTAGCCTGAGCACCAAAACCAGCAGCATTG CAAGCGCAATGAGCAATGCATTTCTGCAGACAACCGGTGTTGT TAATCAGCCGTTTATTAACGAAATTACCCAGCTGGTTAGCATG TTTGCACAGGCAGGTATGAATGATGTTAGCGCAGAAAACCTGT ACTTTCAAGGCGATTAGTAA	MHHHHHHSHTTPWTNPGLAENFMNS FMQGLSSMPGFTASQLDKMSTIAQS MVQSIQSLAAQGRTSPNDLQALNMA FASSMAEIAASEEGGGSLSTKTSSI ASAMSNAFLQTTGVVNQPFINEITQ LVSMFAQAGMNDVSAENLYFQGDX
*Tm*9D8* TrpB	ATGCACCACCACCATCACCATAAAGGCTATTTTGGTCCGTATG GTGGTCAGTATGTTCCGGAAATTCTGATGGGCGCACTGGAAGA ACTGGAAGCAGCATATGAAGGCATCATGAAAGATGAGAGCTTT TGGAAAGAATTTAACGATCTGCTGCGTGATTATGCAGGTCGTC CGACACCGCTGTATTTTGCACGTCGTCTGAGTGAAAAATATGG TGCACGTGTGTATCTGAAACGCGAGGACCTGCTGCATACCGGT GCACATAAAATCAATAATGCAATTGGTCAGGTTCTGCTGGCAA AGCTGATGGGTAAAACCCGTATTATTGCAGAAACCGGTGCAGG TCAGCATGGTGTTGCAACCGCAACCGCAGCAGCACTGTTTGGT ATGGAATGTGTGATTTATATGGGCGAAGAAGATACCATTCGTC AGAAACTGAATGTGGAACGTATGAAACTGCTGGGTGCAAAAGT TGTTCCGGTTAAAAGCGGTAGTCGTACCCTGAAAGATGCAATT GATGAAGCACTGCGCGATTGGATTACCAATCTGCAAACCACCT ATTACGTTTTTGGTAGCGTTGTTGGTCCGCATCCGTATCCGAT TATTGTTCGTAATTTTCAGAAAGTGATCGGCGAGGAAACCAAA AAGCAGATTCCGGAAAAAGAAGGTCGTCTGCCGGATTATATTG TTGCATGTGTTAGCGGTGGTAGCAATGCAGCCGGTATCTTTTA TCCGTTTATTGATAGCGGTGTGAAACTGATTGGTGTTGAAGCC GGTGGTGAAGGTCTGGAAACCGGTAAACATGCAGCAAGCCTGC TGAAAGGTAAAATTGGTTATCTGCATGGCAGCAAAACCTTTGT TCTGCAAGATGATTGGGGTCAAGTTCAGGTTAGCCATAGCGTT AGCGCAGGTCTGGATTATAGTGGCGTTGGTCCTGAACATGCAT ATTGGCGTGAAACCGGCAAAGTTCTGTATGATGCAGTTACCGA TGAAGAGGCACTGGATGCATTTATTGAACTGAGCCGTCTGGAA GGTATTATTCCGGCACTGGAAAGCAGCCATGCACTGGCATATC TGAAAAAGATTAACATCAAAGGCAAGGTCGTTGTGGTTAATCT GAGCGGTCGTGGTGATAAAGATCTGGAAAGCGTTCTGAATCAT CCGTATGTTCGTGAACGTATTCGCTAA	MHHHHHHKGYFGPYGGQYVPEILMG ALEELEAAYEGIMKDESFWKEFNDL LRDYAGRPTPLYFARRLSEKYGARV YLKREDLLHTGAHKINNAIGQVLLA KLMGKTRIIAETGAGQHGVATATAA ALFGMECVIYMGEEDTIRQKLNVER MKLLGAKVVPVKSGSRTLKDAIDEA LRDWITNLQTTYYVFGSVVGPHPYP IIVRNFQKVIGEETKKQIPEKEGRL PDYIVACVSGGSNAAGIFYPFIDSG VKLIGVEAGGEGLETGKHAASLLKG KIGYLHGSKTFVLQDDWGQVQVSHS VSAGLDYSGVGPEHAYWRETGKVLY DAVTDEEALDAFIELSRLEGIIPAL ESSHALAYLKKINIKGKVVVVNLSG RGDKDLESVLNHPYVRERIR
ZiPro Trp61*TAG[b]	ATGTCGGTAGATATGTACATCGAACGCGCAGGAGACATTACCT AGGAGAAGGACGCTGAAGTCACTGGCAATTCACCCCGTTTAGA CGTTGCTCTGGATGAATCGGGTGACTTTAGTCTTGTAGAAGAC GATGGTCCTCCAATGGCGGGGGGAGGTGGCTCAGGAGGGGGTG GGTCTGGTGCATTATGGGACGTACCAGCACCTAAGGAAGTAAA AAACGGCGAAACTACAGACGGTGTGTATCGCGTGATGACACGT GGTTTGTTGGGTAGCACACAAGTGGGCGTCGGAGTAATGCAAG AGGGCGTATTTCACACGATGTGGCATGTAACCAAAGGTTCTGC GTTACGCTCCGGCGAAGGACGCCTTGACCCGTATTGGGGCGAC GTCAAGCAAGATTTGGTTAGCTATAGCGGACCTTGGAAACTGG ACGCGGCATGGGACGGCCACTCGGAGGTACAGTTATTGGCCGT TCCGCCAGGAGAGCGTGCTCGTAACATTCAAACGCTGCCTGGT ATTTTCAAAACTAAGGATGGTGATATCGGGGCGGTGGCCTTAG ATTATCCGGCAGGAACATCCGGGTCTCCCATTTTAGACAAGAG TGGTCGCGTGATTGGGTTATATGGGAATGGGGTTGTAATCAAG	MSVDMYIERAGDITXEKDAEVTGNS PRLDVALDESGDFSLVEDDGPPMAG GGGSGGGGSGALWDVPAPKEVKNGE TTDGVYRVMTRGLLGSTQVGVGVMQ EGVFHTMWHVTKGSALRSGEGRLDP YWGDVKQDLVSYSGPWKLDAAWDGH SEVQLLAVPPGERARNIQTLPGIFK TKDGDIGAVALDYPAGTSGSPILDK SGRVIGLYGNGVVIKNGSYVSAITQ GRRHHHHHH

	AATGGATCTTACGTCTCGGCGATCACACAGGGTCGCCGCCATC ACCACCATCATCACTAA	
ZiPro Trp5TAG	ATGTCGGTAGATATGTACATCGAACGCGCAGGAGACATTACCT GGGAGAAGGACGCTGAAGTCACTGGCAATTCACCCCGTTTAGA CGTTGCTCTGGATGAATCGGGTGACTTTAGTCTTGTAGAAGAC GATGGTCCTCCAATGGCGGGGGGAGGTGGCTCAGGAGGGGGTG GGTCTGGTGCATTATAGGACGTACCAGCACCTAAGGAAGTAAA AAACGGCGAAACTACAGACGGTGTGTATCGCGTGATGACACGT GGTTTGTTGGGTAGCACACAAGTGGGCGTCGGAGTAATGCAAG AGGGCGTATTTCACACGATGTGGCATGTAACCAAAGGTTCTGC GTTACGCTCCGGCGAAGGACGCCTTGACCCGTATTGGGGCGAC GTCAAGCAAGATTTGGTTAGCTATAGCGGACCTTGGAAACTGG ACGCGGCATGGGACGGCCACTCGGAGGTACAGTTATTGGCCGT TCCGCCAGGAGAGCGTGCTCGTAACATTCAAACGCTGCCTGGT ATTTTCAAAACTAAGGATGGTGATATCGGGGCGGTGGCCTTAG ATTATCCGGCAGGAACATCCGGGTCTCCCATTTTAGACAAGAG TGGTCGCGTGATTGGGTTATATGGGAATGGGGTTGTAATCAAG AATGGATCTTACGTCTCGGCGATCACACAGGGTCGCCGCCATC ACCACCATCATCACTAA	MSVDMYIERAGDITWEKDAEVTGNS PRLDVALDESGDFSLVEDDGPPMAG GGGSGGGGSGALXDVPAPKEVKNGE TTDGVYRVMTRGLLGSTQVGVGVMQ EGVFHTMWHVTKGSALRSGEGRLDP YWGDVKQDLVSYSGPWKLDAAWDGH SEVQLLAVPPGERARNIQTLPGIFK TKDGDIGAVALDYPAGTSGSPILDK SGRVIGLYGNGVVIKNGSYVSAITQ GRRHHHHHH
ZiPro Trp50TAG	ATGTCGGTAGATATGTACATCGAACGCGCAGGAGACATTACCT GGGAGAAGGACGCTGAAGTCACTGGCAATTCACCCCGTTTAGA CGTTGCTCTGGATGAATCGGGTGACTTTAGTCTTGTAGAAGAC GATGGTCCTCCAATGGCGGGGGGAGGTGGCTCAGGAGGGGGTG GGTCTGGTGCATTATGGGACGTACCAGCACCTAAGGAAGTAAA AAACGGCGAAACTACAGACGGTGTGTATCGCGTGATGACACGT GGTTTGTTGGGTAGCACACAAGTGGGCGTCGGAGTAATGCAAG AGGGCGTATTTCACACGATGTAGCATGTAACCAAAGGTTCTGC GTTACGCTCCGGCGAAGGACGCCTTGACCCGTATTGGGGCGAC GTCAAGCAAGATTTGGTTAGCTATAGCGGACCTTGGAAACTGG ACGCGGCATGGGACGGCCACTCGGAGGTACAGTTATTGGCCGT TCCGCCAGGAGAGCGTGCTCGTAACATTCAAACGCTGCCTGGT ATTTTCAAAACTAAGGATGGTGATATCGGGGCGGTGGCCTTAG ATTATCCGGCAGGAACATCCGGGTCTCCCATTTTAGACAAGAG TGGTCGCGTGATTGGGTTATATGGGAATGGGGTTGTAATCAAG AATGGATCTTACGTCTCGGCGATCACACAGGGTCGCCGCCATC ACCACCATCATCACTAA	MSVDMYIERAGDITWEKDAEVTGNS PRLDVALDESGDFSLVEDDGPPMAG GGGSGGGGSGALWDVPAPKEVKNGE TTDGVYRVMTRGLLGSTQVGVGVMQ EGVFHTMXHVTKGSALRSGEGRLDP YWGDVKQDLVSYSGPWKLDAAWDGH SEVQLLAVPPGERARNIQTLPGIFK TKDGDIGAVALDYPAGTSGSPILDK SGRVIGLYGNGVVIKNGSYVSAITQ GRRHHHHHH
ZiPro Trp69TAG	ATGTCGGTAGATATGTACATCGAACGCGCAGGAGACATTACCT GGGAGAAGGACGCTGAAGTCACTGGCAATTCACCCCGTTTAGA CGTTGCTCTGGATGAATCGGGTGACTTTAGTCTTGTAGAAGAC GATGGTCCTCCAATGGCGGGGGGAGGTGGCTCAGGAGGGGGTG GGTCTGGTGCATTATGGGACGTACCAGCACCTAAGGAAGTAAA AAACGGCGAAACTACAGACGGTGTGTATCGCGTGATGACACGT GGTTTGTTGGGTAGCACACAAGTGGGCGTCGGAGTAATGCAAG AGGGCGTATTTCACACGATGTGGCATGTAACCAAAGGTTCTGC GTTACGCTCCGGCGAAGGACGCCTTGACCCGTATTAGGGCGAC GTCAAGCAAGATTTGGTTAGCTATAGCGGACCTTGGAAACTGG ACGCGGCATGGGACGGCCACTCGGAGGTACAGTTATTGGCCGT TCCGCCAGGAGAGCGTGCTCGTAACATTCAAACGCTGCCTGGT ATTTTCAAAACTAAGGATGGTGATATCGGGGCGGTGGCCTTAG ATTATCCGGCAGGAACATCCGGGTCTCCCATTTTAGACAAGAG TGGTCGCGTGATTGGGTTATATGGGAATGGGGTTGTAATCAAG AATGGATCTTACGTCTCGGCGATCACACAGGGTCGCCGCCATC ACCACCATCATCACTAA	MSVDMYIERAGDITWEKDAEVTGNS PRLDVALDESGDFSLVEDDGPPMAG GGGSGGGGSGALWDVPAPKEVKNGE TTDGVYRVMTRGLLGSTQVGVGVMQ EGVFHTMWHVTKGSALRSGEGRLDP YXGDVKQDLVSYSGPWKLDAAWDGH SEVQLLAVPPGERARNIQTLPGIFK TKDGDIGAVALDYPAGTSGSPILDK SGRVIGLYGNGVVIKNGSYVSAITQ GRRHHHHHH

ZiPro Trp83TAG	ATGTCGGTAGATATGTACATCGAACGCGCAGGAGACATTACCT GGGAGAAGGACGCTGAAGTCACTGGCAATTCACCCCGTTTAGA CGTTGCTCTGGATGAATCGGGTGACTTTAGTCTTGTAGAAGAC GATGGTCCTCCAATGGCGGGGGGAGGTGGCTCAGGAGGGGGTG GGTCTGGTGCATTATGGGACGTACCAGCACCTAAGGAAGTAAA AAACGGCGAAACTACAGACGGTGTGTATCGCGTGATGACACGT GGTTTGTTGGGTAGCACACAAGTGGGCGTCGGAGTAATGCAAG AGGGCGTATTTCACACGATGTGGCATGTAACCAAAGGTTCTGC GTTACGCTCCGGCGAAGGACGCCTTGACCCGTATTGGGGCGAC GTCAAGCAAGATTTGGTTAGCTATAGCGGACCTTAGAAACTGG ACGCGGCATGGGACGGCCACTCGGAGGTACAGTTATTGGCCGT TCCGCCAGGAGAGCGTGCTCGTAACATTCAAACGCTGCCTGGT ATTTTCAAAACTAAGGATGGTGATATCGGGGCGGTGGCCTTAG ATTATCCGGCAGGAACATCCGGGTCTCCCATTTTAGACAAGAG TGGTCGCGTGATTGGGTTATATGGGAATGGGGTTGTAATCAAG AATGGATCTTACGTCTCGGCGATCACACAGGGTCGCCGCCATC ACCACCATCATCACTAA	MSVDMYIERAGDITWEKDAEVTGNS PRLDVALDESGDFSLVEDDGPPMAG GGGSGGGGSGALWDVPAPKEVKNGE TTDGVYRVMTRGLLGSTQVGVGVMQ EGVFHTMWHVTKGSALRSGEGRLDP YWGDVKQDLVSYSGPXKLDAAWDGH SEVQLLAVPPGERARNIQTLPGIFK TKDGDIGAVALDYPAGTSGSPILDK SGRVIGLYGNGVVIKNGSYVSAITQ GRRHHHHHH
ZiPro Trp89TAG	ATGTCGGTAGATATGTACATCGAACGCGCAGGAGACATTACCT GGGAGAAGGACGCTGAAGTCACTGGCAATTCACCCCGTTTAGA CGTTGCTCTGGATGAATCGGGTGACTTTAGTCTTGTAGAAGAC GATGGTCCTCCAATGGCGGGGGGAGGTGGCTCAGGAGGGGGTG GGTCTGGTGCATTATGGGACGTACCAGCACCTAAGGAAGTAAA AAACGGCGAAACTACAGACGGTGTGTATCGCGTGATGACACGT GGTTTGTTGGGTAGCACACAAGTGGGCGTCGGAGTAATGCAAG AGGGCGTATTTCACACGATGTGGCATGTAACCAAAGGTTCTGC GTTACGCTCCGGCGAAGGACGCCTTGACCCGTATTGGGGCGAC GTCAAGCAAGATTTGGTTAGCTATAGCGGACCTTGGAAACTGG ACGCGGCATAGGACGGCCACTCGGAGGTACAGTTATTGGCCGT TCCGCCAGGAGAGCGTGCTCGTAACATTCAAACGCTGCCTGGT ATTTTCAAAACTAAGGATGGTGATATCGGGGCGGTGGCCTTAG ATTATCCGGCAGGAACATCCGGGTCTCCCATTTTAGACAAGAG TGGTCGCGTGATTGGGTTATATGGGAATGGGGTTGTAATCAAG AATGGATCTTACGTCTCGGCGATCACACAGGGTCGCCGCCATC ACCACCATCATCACTAA	MSVDMYIERAGDITWEKDAEVTGNS PRLDVALDESGDFSLVEDDGPPMAG GGGSGGGGSGALWDVPAPKEVKNGE TTDGVYRVMTRGLLGSTQVGVGVMQ EGVFHTMWHVTKGSALRSGEGRLDP YWGDVKQDLVSYSGPWKLDAAXDGH SEVQLLAVPPGERARNIQTLPGIFK TKDGDIGAVALDYPAGTSGSPILDK SGRVIGLYGNGVVIKNGSYVSAITQ GRRHHHHHH

[a] X indicates the position of the 7AW residue.

[b] The star indicates that the site of mutation is in NS2B.

References

(1) Boville, C. E.; Romney, D. K.; Almhjell, P. J.; Sieben, M.; Arnold, F. H. Improved synthesis of 4-cyanotryptophan and other tryptophan analogues in aqueous solvent using variants of TrpB from *Thermotoga maritima*. *J. Org. Chem*. **2018**, *83*, 7447–7452.

(2) 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. *Angew. Chem. Int. Ed.* **2022**, *61*, e202114154.

(3) 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.

(4) Apponyi, M., Ozawa, K., Dixon, N.E. and Otting, G. (2008) Cell-free protein synthesis for analysis by NMR spectroscopy. In Methods in Molecular Biology 426, Structural proteomics: high-throughput methods. B. Kobe, M. Guss, T. Huber, Eds. Humana Press, Totowa, USA, pp. 257–268.

(5) Hwang, T. L.; Shaka, A. J. Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *J. Magn. Reson. A* **1995**, *112*, 275–279.