Supporting information

Through-Space Scalar $^{19}F–^{19}F$ Couplings between Fluorinated Non-Canonical Amino Acids for the Detection of Specific Contacts in Proteins

Henry W. Orton,[1] Haocheng Qianzhu,[2] Elwy H. Abdelkader,[1] Edan I. Habel,[2] Yi Jiun Tan,[1] Rebecca L. Frkic,[1] Colin J. Jackson,[1] Thomas Huber,[2] and Gottfried Otting[1]*

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

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

Table of contents:

Supplementary tables

References

Experimental procedures

1. Protein expression and purification

1.1 Cloning of synthetase plasmid and expression plasmids

The gene encoding a previously reported synthetase for CF_3-Phe incorporation (tfm-Phe2 with the mutations Y32Q, L65Q, F108Q, H109E, D158S, L162A)[1] was cloned into a pEVOL plasmid[2] using the same method as reported previously for the construction of pEVOL-SFRS.[3] The resulting pEVOL-CFRS plasmid has a chloramphenicol resistance marker and a *p15A* origin of replication. Expression of the aminoacyl-tRNA synthetase for CF_3-Phe incorporation is under control of the *araBAD* promoter and an optimized suppressor tRNA $Mj^{Tyr}tRNA_{CUA}$ is controlled by the *proK* promoter.[4]

The expression of all PpiB mutants was under control of the T7 promoter. A pET-ppib27/98amb plasmid was constructed based on pET-3a (Novagen, USA), harboring the gene of wildtype PpiB with residues Phe27 and Phe98 mutated to the amber codon TAG. The construct contained a C-terminal His_6-tag. pET-ppib4amb27/98 was then constructed based on pET-ppib27/98amb by mutating also the site of Phe4 to the amber codon.

The gene encoding the PpiB F4A mutant with residues 27 and 98 mutated to the amber codon was transferred into a pCDF-1b plasmid (Novagen, USA), which has a spectinomycin resistance marker, resulting in the expression plasmid pCDF-ppibF4A27/98amb.

All primers and gene fragments were purchased from Integrated DNA Technologies (USA).

1.2 Functional test of *Mj*TyrRS variants for incorporation of CF_3-Tyr and CF_3-Phe

A pCDF-ambRFP plasmid was constructed, which encodes a MASMTG-His_6-amber(TAG)-mCherry red fluorescent protein (RFP) gene in the pCDF vector, where MASMTG stands for the peptide encoded by the 5' nucleotide sequence of the T7 gene 10. The pEVOL-CFRS and pCDF-ambRFP plasmids were co-transformed into *E. coli* B-95.ΔA cells.[5] Similarly, pEVOL-SFRS (which encodes SF61) and pCDF-ambRFP were co-transformed into *E. coli* B-95.ΔA cells. The recovered cells were grown at 37 °C in LB medium containing 17 mg/L chloramphenicol and 50 mg/L spectinomycin for 12 hours. 150 μL of culture was aliquoted into 96-well plates, with 0.4% arabinose supplied. The cells in the plates were grown at 37 °C for 6 hours under different conditions: without the addition of unnatural amino acid (uAA), with 2 mM CF_3-Phe (Chem-Impex International Inc., USA) or with 2 mM CF_3-Tyr (AMATEK Scientific, China). Subsequently, the temperature was reduced to 25 °C and protein expression was induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The RFP fluorescence level and cell density by OD_{600} were measured after expression overnight, using an Infinite® 200 PRO plate reader (Tecan Trading AG, Switzerland). Figure S1 reports the average values of the fluorescence level normalized by OD_{600}, illustrating the high efficiency and specificity of CF_3-Phe/CF_3-Tyr incorporation by both tfm-Phe2 and SF61.

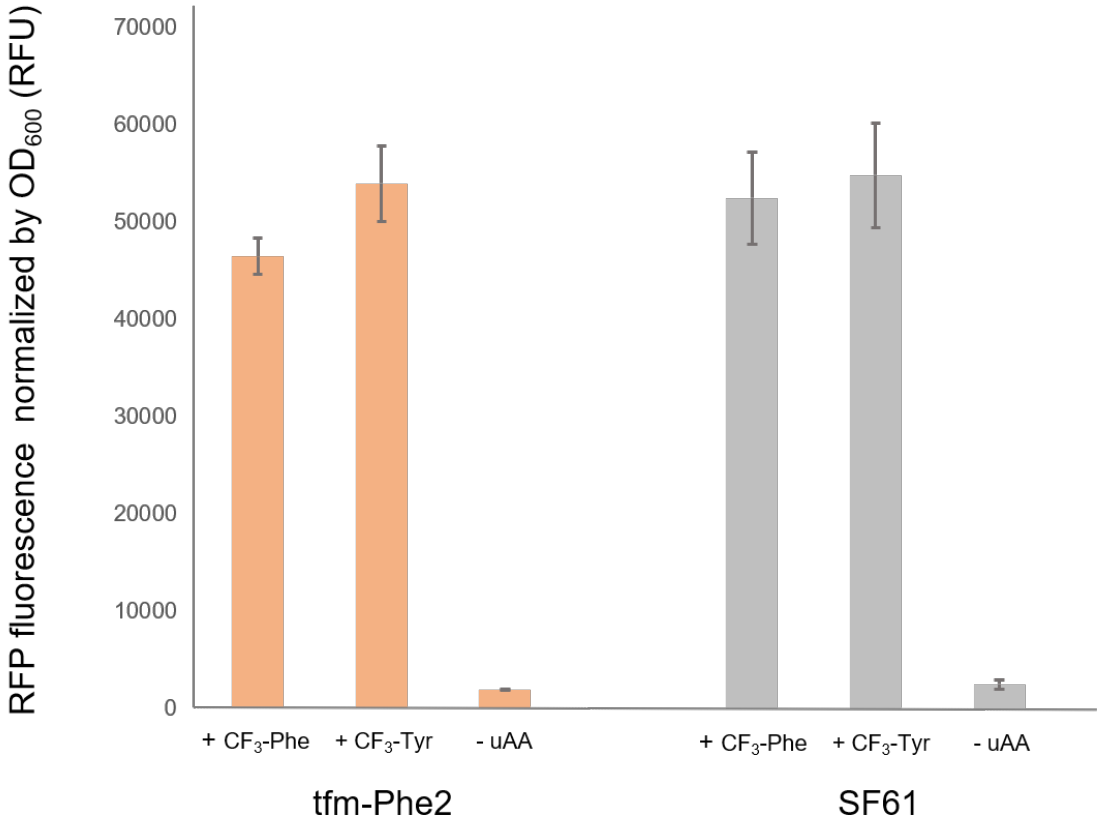

Figure S1. RFP fluorescence levels normalized by OD_{600} in cell cultures producing RFP with CF_3-Phe or CF_3-Tyr incorporated by amber codon read-through by tfm-Phe2 or SF61. Error bars were derived from six biological replicates. SF61 is an aminoacyl-tRNA synthetase selected for incorporation of SF_5-Phe into proteins.[3] Its ready recognition of CF_3-Phe and CF_3-Tyr demonstrates multi-specificity.

1.3 Expression and purification of PpiB CF_3-Phe(27,98) and CF_3-Tyr(27,98)

PpiB samples with site-specific incorporation of CF_3-Tyr or CF_3-Phe at positions 27 and 98 were produced by co-transforming *E. coli* B-95.ΔA cells with the plasmids pEVOL-CFRS and pET-PpiB2798amb. The recovered cells were grown at 37 °C in LB medium containing 17 mg/L chloramphenicol and 100 mg/L ampicillin. A 10 mL overnight culture was used to inoculate 1 L LB medium supplemented with antibiotics. The cells were grown at 37 °C to an OD_{600} of 0.6. At this point, the cells were pelleted by centrifugation at 4000 *g* for 10 minutes at 25 °C and resuspended in fresh 250 mL LB medium supplied with 0.2% arabinose (13.3 mM), 17 mg/L chloramphenicol, 100 mg/L ampicillin and 2 mM CF_3-Tyr or CF_3-Phe. The cells were grown at 37 °C for another 1 hour to initiate expression of the synthetase. The temperature was subsequently reduced to 18 °C and protein expression induced by the addition of 1 mM IPTG.

The harvested cells were resuspended in buffer A (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% glycerol, 10 mM imidazole) followed by lysis using an Avestin Emulsiflex C5 (Avestin, Canada; two passes using 10,000–15,000 psi). The clarified cell lysate was loaded onto a 5 mL HisTrap™ FF column (Cytiva, USA) connected to an ÄKTA pure 25 chromatography system. The column was washed with 20 column volumes buffer B (same as buffer A but with 20 mM imidazole) and the protein was eluted with 5 column volumes buffer C (same as buffer A but with 500 mM imidazole). The eluted protein was desalted using a HiPrep Desalting 26/10 column (Cytiva, USA) equilibrated with buffer D (50 mM Tris-HCl pH 7.5, 100 mM NaCl). The desalted protein was concentrated for NMR measurement and analysed by mass spectrometry and SDS-PAGE (Figure S2a,b).

PpiB CF_3-Tyr(4,27,98) was produced and purified in the same way, except that the plasmid pET-PpiB2798amb was replaced by pET-PpiB4amb2798 in the initial transformation step. Figure S2c depicts the mass spectrometric and SDS-PAGE results obtained for purified PpiB CF_3-Tyr(4,27,98).

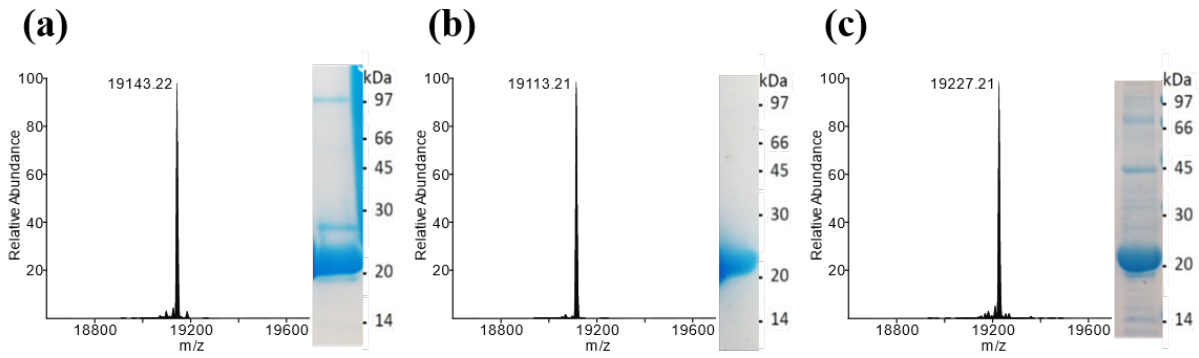

Figure S2. Mass spectra and SDS-PAGE of PpiB mutants with CF_3-Tyr or CF_3-Phe residues. (a) PpiB CF_3-Tyr(27,98); calculated mass 19,144.36 Da. (b) PpiB CF_3-Phe(27,98); calculated mass 19112.36 Da. (c) PpiB CF_3-Tyr(4,27,98); calculated mass 19,228.36 Da.

1.4 Expression and purification of PpiB F4A/CF_3-Tyr(27,98) and PpiB F4A/CF_3-Phe(27,98)

The Phe4Ala mutants of PpiB with CF_3-Tyr or CF_3-Phe in positions 27 and 98 were produced by co-transforming *E. coli* B-95.ΔA cells with the plasmids pEVOL-CFRS and pCDF-PpiBF4A2798amb. The recovered cells were grown at 37 °C in LB medium containing 17 mg/L chloramphenicol and 50 mg/L spectinomycin. 10 mL overnight culture was used to inoculate 1 L LB medium supplemented with antibiotics. The cells were grown at 37 °C to an OD_{600} of 0.75, pelleted by centrifugation at 4000 *g* for 10 minutes at 25 °C and resuspended in fresh 300 mL LB medium supplied with 0.4% arabinose, 17 mg/L chloramphenicol, 50 mg/L spectinomycin and 2 mM CF_3-Tyr or CF_3-Phe. The cells were grown for another 1 hour at 37 °C, the temperature was reduced to 18 °C and protein expression induced by the addition of 1 mM IPTG. The proteins were purified and concentrated in buffer D for NMR as described above for the other PpiB mutants. Figure S3 shows the results of the analysis by mass spectrometry and SDS-PAGE.

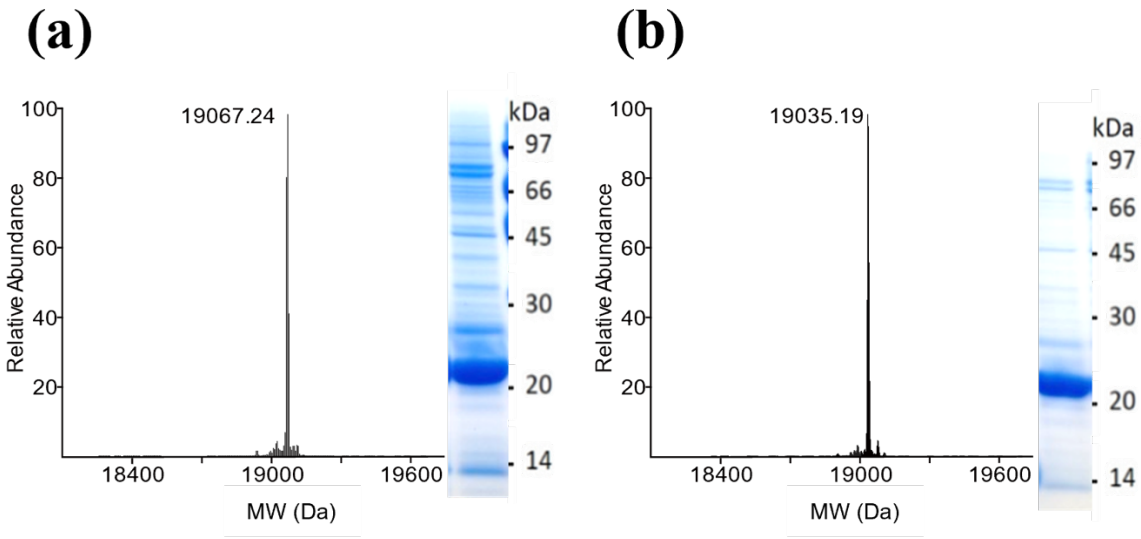

Figure S3. Mass spectra and SDS-PAGE of PpiB Phe4Ala mutants with CF_3-Tyr or CF_3-Phe residues. (a) PpiB F4A/CF_3-Tyr (27,98); calculated mass 19,068.26 Da. (b) PpiB F4A/CF_3-Phe (27,98); calculated mass 19,036.26 Da.

2. Crystallization of PpiB mutants

Prior to crystallization of PpiB CF_3-Tyr(27,98) and PpiB F4A/CF_3-Phe(27,98), the proteins were purified further by size exclusion chromatography, using a HiLoad 26/600 Superdex 75 pg column (Cytiva, USA) equilibrated with buffer E (50 mM Tris-HCl pH 7.5, 500 mM NaCl). Following gel filtration, the fractions containing the target protein were combined, concentrated using an Amicon ultrafiltration centrifugal tube (Merck Millipore, USA) with a molecular weight cut-off of 10 kDa, and buffer-exchanged to buffer F (50 mM imidazole, 1 mM EDTA, 1 mM DTT). The final concentration of the protein samples was 8 mg/mL.

The PpiB mutants were crystallized by the hanging-drop vapor diffusion method at 18 °C. PpiB CF_3-Tyr(27,98) and PpiB F4A/CF_3-Phe(27,98) crystals formed after 4 days of equilibrating against a reservoir solution of 34% PEG 3350, 0.1 M Tris-HCl (pH 8.0), and 0.2 M sodium acetate, and 40% PEG 3350, 0.1 M Tris-HCl (pH 8.0), respectively. These conditions are near those reported previously for the crystallization of wildtype PpiB.[6] Crystals were flash frozen without additional cryoprotection prior to data collection.

Diffraction data were collected at 100 K using the MX2 beamline at the Australian Synchrotron.[7] Reflections collected were indexed and integrated using XDS[8] and scaled in Aimless (CCP4).[9] The phase problem was overcome by molecular replacement in Phaser MR (CCP4),[10] using PDB ID 2NUL[11] as the search model. The structure was refined by iterative rounds of rebuilding in Coot,[12] and refinement in phenix.refine,[13,14] or Refmac.[15,16] Data collection and refinement statistics are given in Table S3. The final structures were deposited in the Protein Data Bank (PDB: 7N3J and 7RFD).

Table S1. Amino acid sequences of the reporter proteins used in this study.[a]

Gene name	Plasmid name	Amino acid sequence
tfm-Phe2	pEVOL-CFRS	MDEFEMIKRNTSEIISEEELREVLKKDEKSAQIGFEPSGKIHLG HYLQIKKMIDLQNAGFDIIIQLADLHAYLNQKGELDEIRKIGD YNKKVFEAMGLKAKYVYGSEQELDKDYTLNVYRLALKTTL KRARRSMELIAREDENPKVAEVIYPIMQVNSIHYAGVDVAV GGMEQRKIHMLARELLPKKVVCIHNPVLTGLDGEGKMSSSK GNFIAVDDSPEEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYP LTIKRPEKFGGDLTVNSYEELESLFKNKELHPMDLKNAVAEE LIKILEPIRKRL
SF61	pEVOL-SFRS	MDEFEMIKRNTSEIISEEELREVLKKDEKSAVIGFEPSGKIHLG HYLQIKKMIDLQNAGFDIIISLADLHAYLNQKGELDEIRKIGD YNKKVFEAMGLKAKYVYGSEWALDKDYTLNVYRLALKTT LKRARRSMELIAREDENPKVAEVIYPIMQVNGVHYKGVDVA VGGMEQRKIHMLARELLPKKVVCIHNPVLTGLDGEGKMSSS KGNFIAVDDSPEEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEY PLTIKRPEKFGGDLTVNSYEELESLFKNKELHPMRLKNAVAE ELIKILEPIRKRL
PpiB(4, 27, 98)	pET-PpiB27/98amb pET-PpiB4amb27/98 pCDF-PpiBF4A27/98amb	MVTFHTNHGDIVIKTFDDKAPETVKNFLDYCREGFYNNTIFH RVINGFMIQGGGFEPGMKQKATKEPIKNEANNGLKNTRGTL AMARTQAPHSATAQFFINVVDNDFLNFSGESLQGWGYCVFA EVVDGMDVVDKIKGVATGRSGMHQDVPKEDVIIESVTVSEH HHHHH
AncCDT-1(68,162)	pCDF-CDT68/162amb	MIAASTLDEIMKRGTLRVGTDADYKPFSFKDKNGQYTGFDID LAKALAKELGVKVEFVPTTWDGIIPALQTGKFDIVMSGMTIT PERKKKVDFSDPYMTAGQTILVKKDNADKIKSFEDLNKPDV KVAVQLGTTSEQAAKEFLPKAKIRTFENNAEAFQEVVSGRA DAMVTDSPVAAYYAKKNPGLAVVVVDEPFTHEPLGFAIRKG DPELLNWVNNWLKQMKKDGTYDKLYEKWFKLHHHHHH
mRFP (192,225)	pCDF-mRFP192/225amb	MASMTGHHHHHHMASSEDVIKEFMRFKVRMEGSVNGHEFE IEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFQYGSK AYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDS SLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTERMYP EDGALKGEIKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGA YKTDIKLDITSHNEDYTIVEQYERAEGRHSTGA

[a] Amino acid residues substituted by non-canonical amino acids are highlighted in red.

Table S2. Yields of the PpiB mutants expressed in the present study reported as mg per liter of cell culture with CF_3-Tyr or CF_3-Phe incorporated.

Protein	Yield (mg per liter of cell culture)
PpiB CF_3-Tyr(27,98)	38.0
PpiB CF_3-Phe(27,98)	32.3
PpiB CF_3-Tyr(4,27,98)	19.2
PpiB F4A/CF_3-Tyr(27,98)	20.6
PpiB F4A/CF_3-Phe(27,98)	25.3

Table S3. Crystallographic data statistics for PpiB CF_3-Tyr(27,98) and PpiB F4A/CF_3-Phe(27,98) crystal structures.[a]

	PpiB CF_3-Tyr(27,98)	PpiB-F4A/CF_3-Phe(27,98)
PDB ID	7N3J	7RFD
Data collection		
Source	Australian Synchrotron MX2	Australian Synchrotron MX2
Wavelength	0.9537	0.9537
Space group	P 1 2_1 1	P 6_1
Cell dimensions		
a, b, c (Å)	34.14, 39.22, 103.71	34.89, 34.89, 210.05
α, β, γ (°)	90, 95.35, 90	90, 90, 120
Resolution (Å)	34.42–2.0 (2.05–2.0)	35.01–1.35 (1.40–1.35)
R_{merge}	0.308 (2.99)	0.209 (2.29)
R_{pim}	0.131 (1.33)	0.049 (0.71)
I/σI	3.9 (0.8)	15.5 (1.3)
$CC_{1/2}$	0.98 (0.60)	0.99 (0.38)
Completeness (%)	99.1 (98.9)	99.9 (99.9)
Multiplicity	6.5 (6.0)	17.4 (10.6)
Refinement		
Resolution (Å)	34.42–2.0 (2.05–2.0)	35.01–1.35 (1.40–1.35)
No. reflections	18541 (1815)	31557 (3180)
R_{work}/R_{free}	0.2481 (0.3962) / 0.2883 (0.4149)	0.1317 (0.0962) / 0.1595 (0.1202)
No. atoms	2680	1537
Protein	2596	1318
Ligand/ion	64	30
Water	20	189
B-factors (overall)	59.79	14.72
Protein	60.03	13.36
Ligand/ion	52.86	14.11
Water	50.90	24.29
R.m.s. deviations		
Bond lengths (Å)	0.004	0.018
Bond angles (°)	0.65	2.24

[a] Statistics for the highest resolution shell are shown in parentheses.

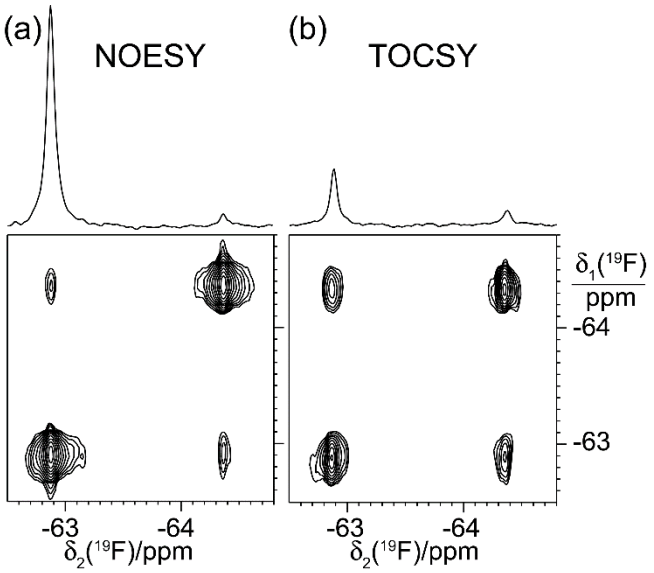

Figure S4. $^{19}F–^{19}F$ NOESY and TOCSY spectra of 0.6 mM PpiB F4A/CF_3-Phe(27,98). The spectra were recorded using t_{1max} = 5.9 ms, t_{2max} = 135 ms, and a total recording time of about 1 h for each spectrum. (a) NOESY spectrum recorded with a mixing time of 200 ms. (b) TOCSY spectrum recorded with 41 ms DIPSI-2 mixing. The panel above shows cross-sections through the low-field diagonal peak of the 2D spectra underneath.

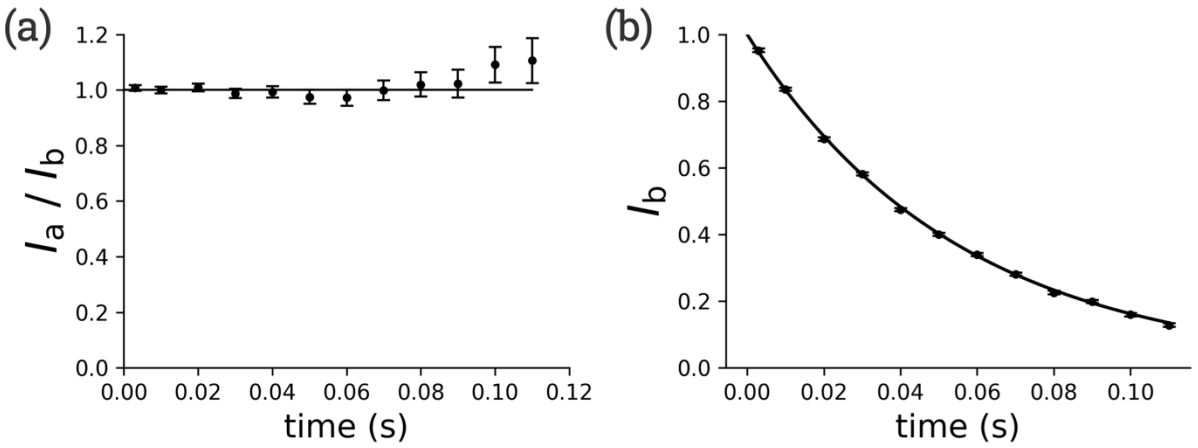

Figure S5. Control experiment of the quantitative *J* evolution experiment of Figures 7 and 8 conducted with a protein containing a single fluorine, GB1 with Trp43 replaced by 7-fluoro-tryptophan. (a) Relative peak intensities observed in experiments a and b of Figure 7 plotted versus the spin echo delay τ. (b) Decay of the signal in experiment b of Figure 7. The fit determined the transverse relaxation rate R_2 as 18.5±0.07 s^{-1}.

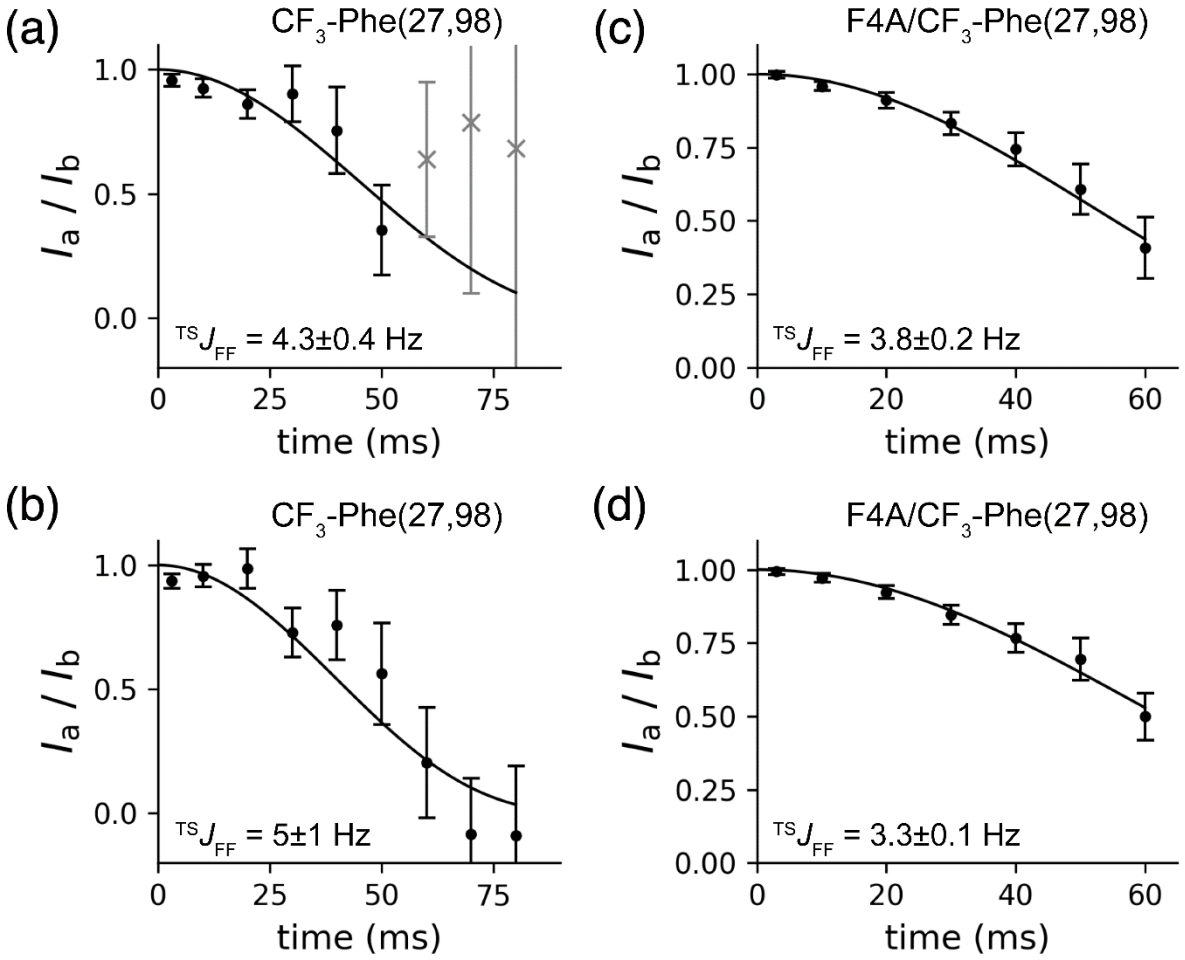

Figure S6. Relative peak intensities observed in quantitative *J* evolution experiments of PpiB CF_3-Phe(27,98) and PpiB F4A/CF_3-Phe(27,98) conducted with and without refocusing of *J* couplings during the spin-echo delay. See caption of Figure 8 for details of the measurement and data fitting. The $^{TS}J_{FF}$ couplings derived by best fits are indicated in the plots. (a) Signal at -62.9 ppm in PpiB CF_3-Phe(27,98). (b) Same as (a), but for the peak at -63.9 ppm. The gray data points were excluded from the fit because of their large uncertainties. (c) Signal at -62.9 ppm in PpiB F4A/CF_3-Phe(27,98). (d) Same as (c), but for the peak at -64.4 ppm.

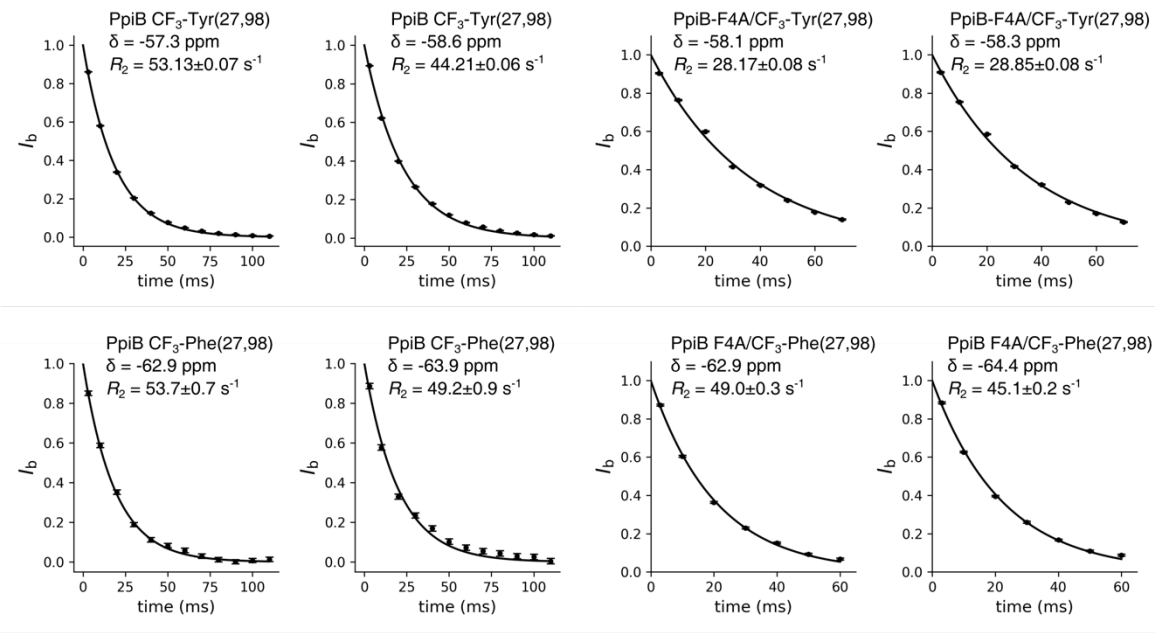

Figure S7. Decay of transverse ^{19}F magnetization in different CF_3 mutants of PpiB determined from fitting the data measured in the quantitative *J* evolution reference experiment. The ^{19}F chemical shifts and transverse relaxation rates are indicated in the plots.

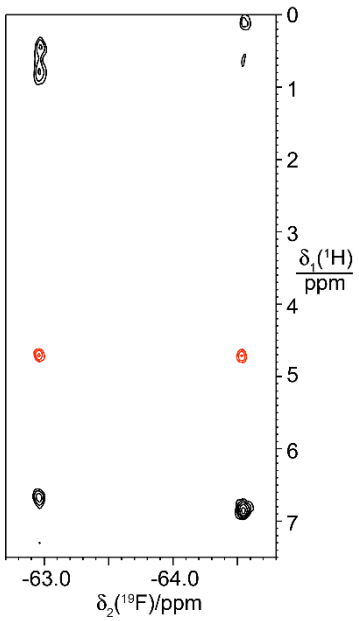

Figure S8. $^{1}H–^{19}F$ NOESY spectrum of 0.6 mM PpiB F4A/CF_3-Phe(27,98). The spectrum was recorded in 13.5 h, using a mixing time of 200 ms, t_{1max} = 7.3 ms, and t_{2max} = 135 ms. Positive and negative peaks are plotted with black and red contour lines, respectively. Negative cross-peaks at the ^{1}H NMR frequency of the water resonance indicate that the $^{1}H–^{19}F$ NOE is governed by rapid changes in the dipolar interaction characteristic of water molecules moving rapidly in the vicinity of the CF_3 groups. For a 2-spin system comprising one ^{1}H spin and one ^{19}F spin, the NOE changes sign when $\tau_c/[1 + (\omega_H - \omega_F)^2\tau_c^2] = 6\ \tau_c/[1 + (\omega_H + \omega_F)^2\tau_c^2]$, where τ_c is the correlation time of the internuclear vector, and ω_H and ω_F are the Larmor frequencies of the ^{1}H and ^{19}F spins, respectively. At 9.4 T, the NOE is zero at τ_c = 0.46 ns.

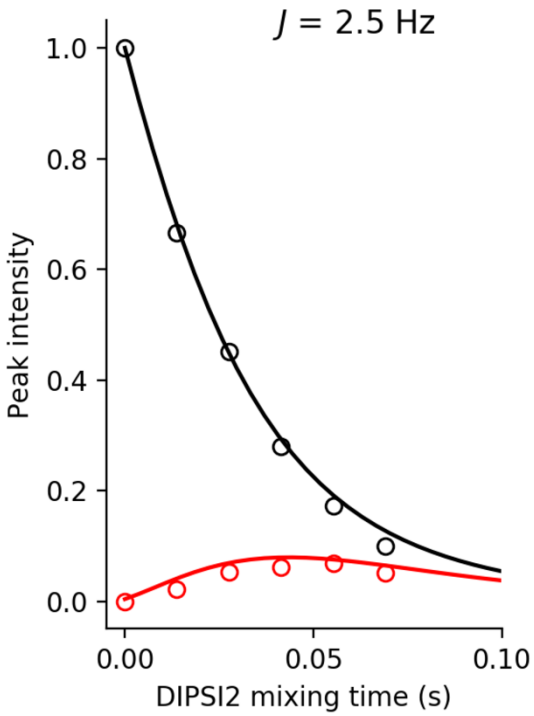

Figure S9. Diagonal and cross-peak intensities in ^{19}F–^{19}F TOCSY spectra as a function of mixing time. Data were simulated using the program Spinach for two CF_3 groups as detailed in the experimental section.[17] The simulations applied DIPSI2 mixing to longitudinal magnetization, followed by an excitation pulse. To alleviate antiphase dispersive components, the peak heights were monitored after apodization with a 50 Hz exponential window function. Experimentally measured data of PpiB CF_3-Tyr(27,98) are shown as open circles. The experimental data were scaled by setting the diagonal peak height at zero mixing time to one. Black and red lines indicate the diagonal and cross-peak heights, respectively.

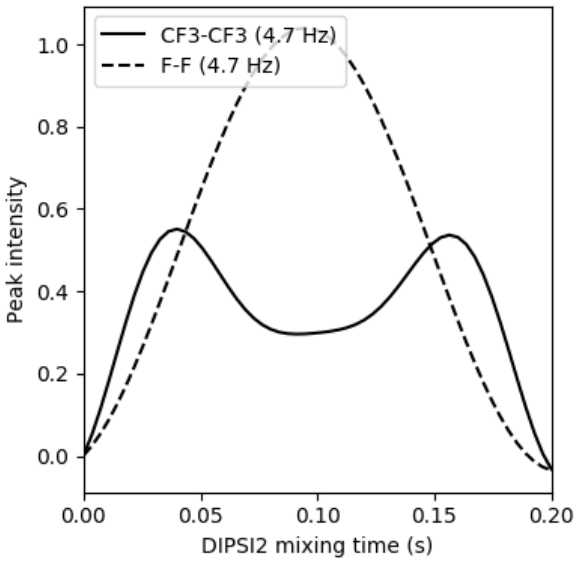

Figure S10. Cross-peak intensities relative to the diagonal peak intensities in $^{19}F–^{19}F$ TOCSY spectra as a function of DIPSI2 mixing time and in the absence of relaxation. Data simulated using the program Spinach.[17] Solid and dashed curves show the results for J_{FF} couplings between two CF_3 groups and two single ^{19}F spins, respectively, with a coupling constant of 4.7 Hz.

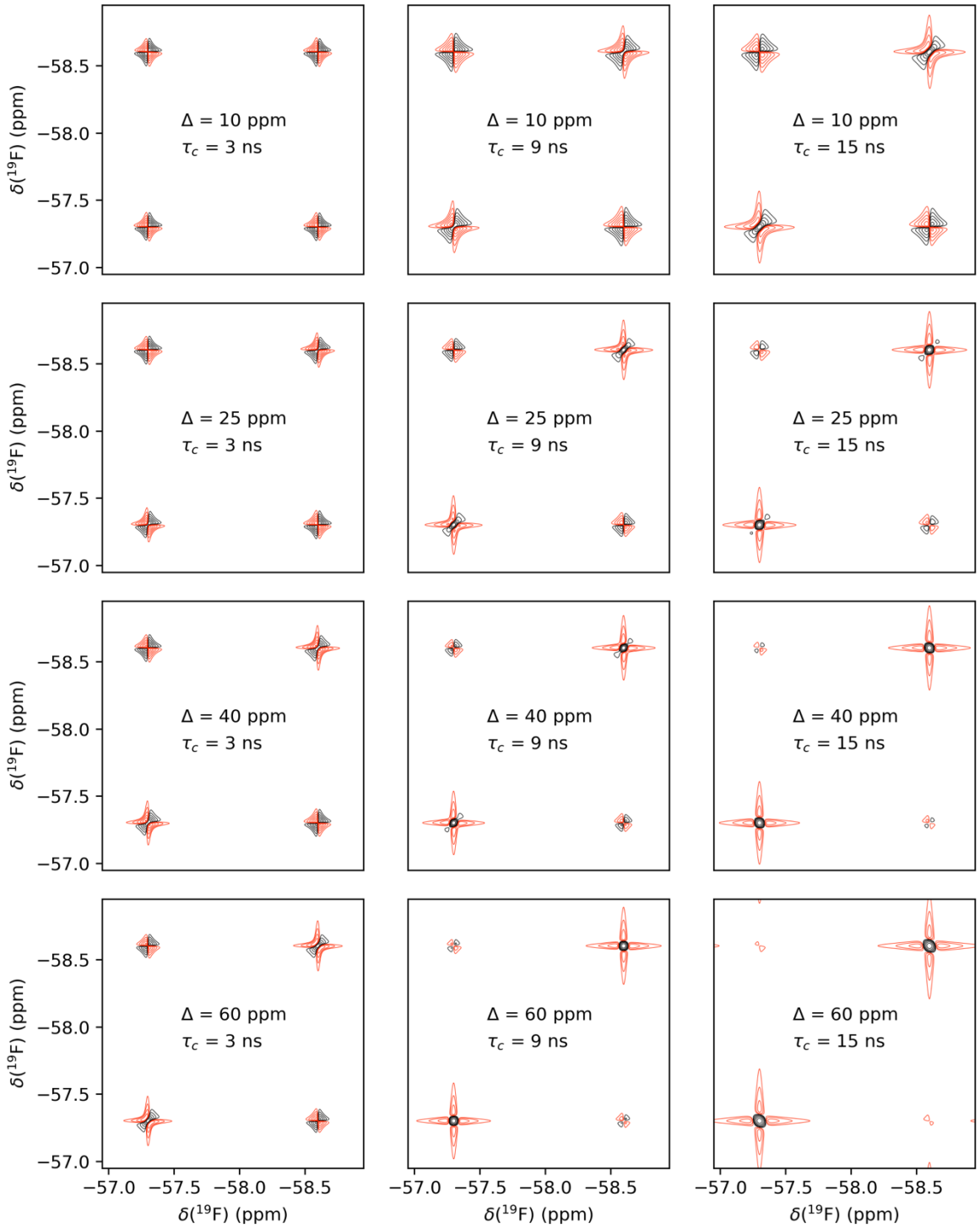

Figure S11. ^{19}F–^{19}F DQF-COSY spectra of PpiB CF_3-Tyr(27,98) simulated using the program Spinach with $^{TS}J_{FF} = 2.5$ Hz.[17] Each spectrum was calculated for a given CSA tensor anisotropy Δ and rotational correlation time τ_c as shown. All spectra were plotted with the lowest contour level set at 1.7% of the height of the diagonal peak and using a factor of 2 between subsequent contour levels.

3. Selection of functional G1PylRS enzymes recognizing TFA-lysine

The previously constructed library plasmid pBK-G1RS[18] encodes a library of pyrrolysyl-tRNA synthetase mutants of the methanogenic archaeon ISO4-G1. This G1PylRS library contains an estimated 2.7 x 10^7 variants by randomization of seven sites lining the amino acid recognition pocket (Leu124, Tyr125, Ala221, and Trp237 fully randomized to the 20 natural amino acids; Asn165 selectively mutated to Gly/Ala/Val/Ser/Asn/Thr/Ile/Asp; Val167 selectively mutated to Gly/Ala/Val/Ser/Cys/Leu/Phe; Tyr204 selectively mutated to Phe/Tyr/Trp). The library was used in conjunction with the selection plasmid pBAD-H6RFP, which encodes the gene of mCherry red fluorescent protein (RFP) preceded by an amber stop codon and an N-terminal His_6 tag (His_6-TAG-RFP). To carry out the selection, the library plasmid pBK-G1RS was transformed into *E. coli* DH10B cells harboring the selection plasmid pBAD-H6RFP. 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 3 mM TFA-lysine, which served as the sample for the first round of positive selection (**1P+**). Overnight expression at 37 °C led to a well-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 fluorescence activated cell sorting (FACS) on an Aria II high speed cell sorter (BD Biosciences, USA; Figure S12).

0.8% of the cells with high RFP levels were selected from the selection round **1P+** (identified a rectangle in Figure S12), collecting 4×10^5 cells in 60 minutes by FACS. The cells collected were grown under conditions of negative selection without the addition of TFA-lysine. 40.5% of the cells from this **2N–** sample showing low RFP expression levels were collected (1×10^6 cells in 10 minutes). These cells were aliquoted to inoculate media with positive (**3P+**) and negative (**3P–**) conditions. The RFP-positive cells (1.2% of the population) from **3P+** were collected (4×10^5 cells in 45 minutes) and recovered under negative growth condition to obtain the **4N–** stage. Following sorting (1×10^6 cells in 15 minutes), the 29.4% fraction of cells showing the lowest level of RFP fluorescence was selected and aliquoted for recovery as **5P+** and **5P–** samples. The cell population with high RFP fluorescence in the **5P** round was 21.9% and 4.6% in the presence and absence of TFA-lysine, respectively, indicating the successful accumulation of active G1PylRS variants specific for TFA-lysine. 6×10^5 cells collected in 20 minutes from the top 5% RFP fluorescent cells of the **5P+** sample were recovered for storage. An aliquot of 2,000 cells were grown 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 and negative growth conditions (with and without 3 mM TFA-lysine, respectively). 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. Sequencing of the six candidates with the highest RFP level in the positive condition revealed a single mutation set with the mutations L124I, Y125L, V167A, Y204W, and A221C, while residues N165 and W237 were unchanged. This G1PylRS mutant was used in the subsequent experiments and referred to as G1TFAKRS. The plasmid pRSF-G1TFAKRS is available from Addgene (Watertown, MA, USA; plasmid #177311).

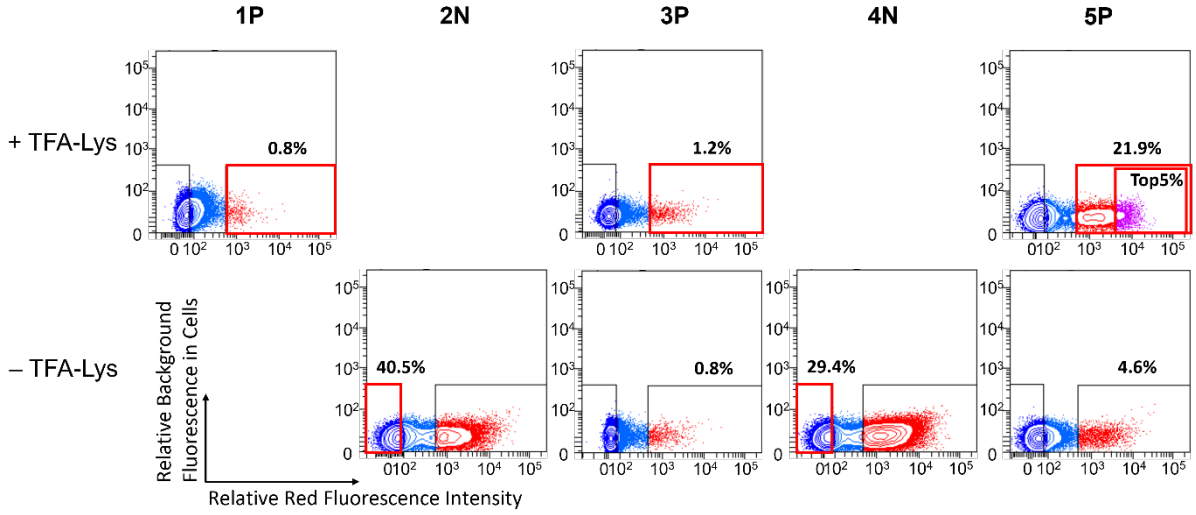

Figure S12. Selection of an aminoacyl-tRNA synthetase specific for N^6-(trifluoroacetyl)-L-lysine (TFA-Lys). FACS experiments for selection of active and specific G1PylRS enzymes that recognize TFA-Lys. The horizontal axis plots the level of red fluorescence and the vertical axis indicates background fluorescence by excitation at 488 nm. **P** and **N** indicate positive and negative selection rounds, where cells were cultured in the presence and absence of TFA-Lys, respectively. Red rectangles indicate fractions of cells collected and re-cultured, where cells with low and high RFP expression were selected following **N** and **P** rounds, respectively, for use in following selection rounds. At the **3P** and **5P** stages, cell cultures were grown with and without TFA-Lys served as control to indicate the presence of TFA-lysine-specific RS enzymes in the gene pool. The 5% fraction of cells in the **5P+** sample showing the brightest RFP fluorescence was collected for further characterization.

4. Expression and purification of AncCDT-1 TFA-Lys(68,162)

AncCDT-1 with Pro68 and Asn162 mutated to TFA-Lys was produced by co-transforming *E. coli* B-95.ΔA cells with the plasmids pRSF-G1TFAKRS and a pCDF plasmid containing the gene of AncCDT-1 with amber stop codons at the sites of TFA-Lys incorporation. The cells were grown at 37 °C in LB medium containing 50 mg/L kanamycin and 50 mg/L spectinomycin. 6 mL of an overnight culture were used to inoculate 250 mL terrific broth (TB) medium supplemented with 50 mg/L kanamycin, 50 mg/L spectinomycin and 20 mM TFA-Lys. The cells were grown at 37 °C to an OD_{600} of 1.0. The temperature was reduced to 25 °C and protein expression was induced by 1 mM IPTG.

Following expression for 16 h, the cells were harvested by centrifugation. The following protein purification steps closely followed the protocol described in Section 1.3 with minor modifications. Cells were resuspended in buffer A and lysed. The cell lysate was centrifuged for 1 h at 30,000 *g* and the supernatant loaded onto a 1 mL His GraviTrap TALON® column (Cytiva, USA). The column was washed with 20 column volumes buffer A and the protein eluted with 5 column volumes buffer C. The buffer was exchanged to storage buffer (50 mM HEPES, pH 7.5, 150 mM NaCl) using an Amicon ultrafiltration centrifugal tube (molecular weight cut-off 10 kDa). The yield was 8.9 mg of purified protein per liter of cell culture.

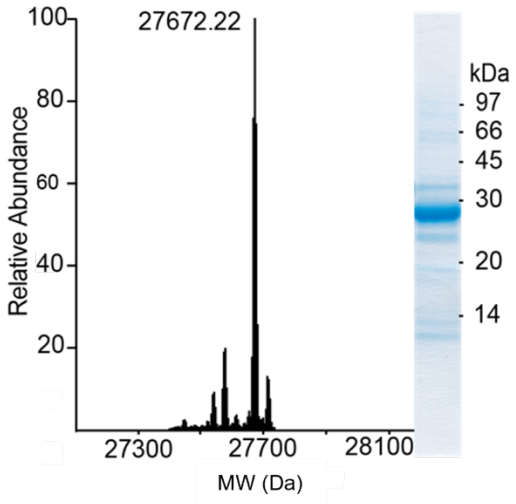

Figure S13. Mass spectrum and SDS-PAGE of AncCDT-1 TFA-Lys(68,162). The calculated mass is 27,673.67 Da. The second largest peak in the mass spectrum corresponds to a glutamine mis-incorporated in place of a TFA-Lys residue (experimental mass: 27,576.17 Da, calculated: 27,577.61 Da).

5. Expression and purification of mRFP TFA-Lys(192,225)

mRFP with Ala192 and Ala225 mutated to TFA-Lys was produced by co-transforming *E. coli* B-95.ΔA cells with the plasmids pRSF-G1TFAKRS and a pCDF plasmid containing the gene of mRFP with amber stop codons at the sites of TFA-Lys incorporation. The cells were grown at 37 °C in LB medium containing 25 mg/L kanamycin and 25 mg/L spectinomycin. 10 mL of an overnight culture were used to inoculate 1 L LB medium supplemented with 25 mg/L kanamycin, 25 mg/L spectinomycin and 20 mM TFA-Lys. The cells were grown at 37 °C to an OD_{600} of 0.6–1.0. The temperature was reduced to 25 °C and protein expression was induced with 1 mM IPTG.

Following expression for 16 h, the cells were harvested by centrifugation. Cells were resuspended in buffer A and lysed as mentioned in section 1.3. The cell lysate was centrifuged for 1 h at 30,000 *g* and the supernatant loaded onto a 1 mL His GraviTrap column (Cytiva, USA). The column was washed with 20 column volumes buffer B and the protein was eluted with 5 column volumes buffer C. The buffer was exchanged to PBS using an Amicon ultrafiltration centrifugal tube (molecular weight cut-off 10 kDa). The yield was 21 mg of purified protein per liter of cell culture.

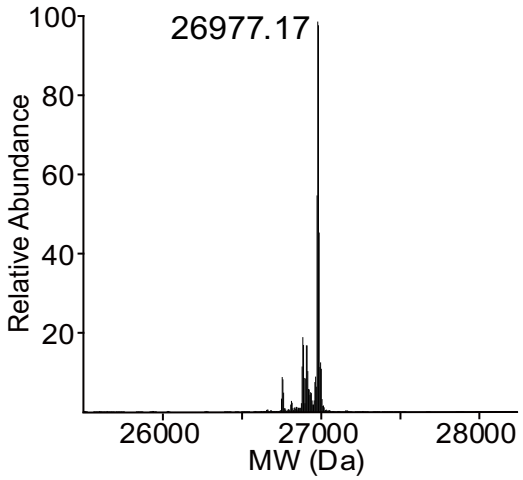

Figure S14. Mass spectrum of mRFP TFA-Lys(192,225). The calculated mass (following loss of the N-terminal methionine) is 26,978.28 Da. The second largest peak in the mass spectrum corresponds to a glutamine mis-incorporated in place of a TFA-Lys residue (experimental mass: 26882.17 Da, calculated: 26,882.23 Da).

6. Bruker pulse program for measuring $^{TS}J_{FF}$ couplings

```
;smallhomoj.ho
;for measuring small homonuclear J-coupling constants
;selecteive decoupling using jump return elements
;place carrier on one resonance and cnst2 on coupling partner

;cnst2 : absolute chemical shift of coupled partner in ppm
;l1 : experiment repeats
;ds : MUST EQUAL 0!!!

#include <Avance.incl>
#include <Grad.incl>

"p2=p1*2"
"d11=30m"

define delay Jret
define delay Jret90
define delay Jret180

"Jret = 1s / (abs(cnst2 - o1/bf1)*bf1)"
"Jret90 = 0.25*Jret"
"Jret180 = 0.5*Jret"

define list<delay> tau = <$VDLIST>

define loopcounter tauloop
"l0=0"
"nbl = td1"
"tauloop = td1 / 2"

1 zd
  d11 st0
  1m pl1:f1
2 6m
3 6m
4 3m
5 3m
  d1
  "d40=0.5*tau - p16 - d16"
  "d41=0.5*tau - p16 - d16 - 4u"
  8u
  (p1 ph4)
  Jret90 UNBLKGRAD
  (p1 ph5)
  p16:gp1
  d16
  d40
if "l0==0"
```

```
{
  ;Inactive coupling
  (p1 ph0)
  Jret180
  (p1 ph2)
}
else
{
  ;Active coupling
  Jret180*0.5*0.69314718056
  (p2 ph0)
  Jret180*0.5*0.69314718056
}
  d41
  p16:gp1
  d16
  4u BLKGRAD
  (p1 ph6)

if "l0==0"
{
  goscnp ph30
}
else
{
  goscnp ph31
}
  ;Phase cycle loop
  3m ipp4 ipp5 ipp6 ipp30 ipp31
  lo to 2 times ns
  3m rppall

  ;Inactive/Active coupling loop
  3m st iu0
  lo to 3 times 2
  3m ru0

  ;Coupling delay loop
  3m tau.inc
  lo to 4 times tauloop
  3m st0 tau.res

  ;Experiment multiplicity loop
  lo to 5 times l1

  d11 wr #0 if #0

exit

ph0=0
```

ph1=1
ph2=2
ph3=3
ph4=0 2
ph5=2 0
ph6=0 0 2 2
ph30=0 2 0 2
ph31=2 0 2 0

References

(1) Jackson, J. C.; Hammill, J. T.; Mehl, R. A. Site-specific incorporation of a ^{19}F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J. Am. Chem. Soc.* **2007**, *129*, 1160–1166.

(2) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J. Mol. Biol.* **2010**, *395*, 361–374.

(3) Qianzhu, H.; Welegedara, A. P.; Williamson, H.; McGrath, A. E.; Mahawaththa, M. C.; Dixon, N. E.; Otting, G.; Huber, T. Genetic encoding of *para*-pentafluorosulfanyl phenylalanine: a highly hydrophobic and strongly electronegative group for stable protein interactions. *J. Am. Chem. Soc.* **2020**, *142*, 17277–17281.

(4) Chatterjee, A.; Sun, S. B.; Furman, J. L.; Xiao, H.; Schultz, P. G. A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. *Biochemistry* **2013**, *52*, 1828–1837.

(5) 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*, 9699–9798.

(6) Edwards, K. J.; Ollis, D. L.; Dixon, N. E. Crystal structure of cytoplasmic *Escherichia coli* peptidyl-prolyl isomerase: evidence for decreased mobility of loops upon complexation. *J. Mol. Biol.* **1997**, *271*, 258–265.

(7) Aragão, D.; Aishima, J.; Cherukuvada, H.; Clarken, R.; Clift, M.; Cowieson, N. P.; Ericsson, D. J.; Gee, C. L.; Macedo, S.; Mudie, N.; Panjikar S.; Price, J. R.; Riboldi-Tunnicliffe, A.; Rostan, R.; Williamson, R.; Caradoc-Davies, T. T.; MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. *J. Synchrotron Rad.*, **2018**, *25*, 885–891.

(8) Kabsch, W. XDS. *Acta Crystallogr. D* **2010**, *66*, 125–132.

(9) Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **2013**, *69*, 1204–1214.

(10) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Cryst.* **2007**, *40*, 658–674.

(11) Edwards, K. J.; Ollis, D. L.; Dixon, N. E. Crystal structure of cytoplasmic *Escherichia coli* peptidyl-prolyl isomerase: evidence for decreased mobility of loops upon complexation. *J. Mol. Biol*. **1997**, *271*, 258–265.

(12) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **2010**, *66*, 486–501.

(13) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart P. H.; Adams, P. D. Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallogr. D* **2012**, *68*, 352–367.

(14) Zwart, P. H.; Afonine, P. V.; Grosse-Kunstleve, R. W.; Hung L-W; Ioerger, T. R.; McCoy, A. J.; McKee, E.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Storoni, L. C.; Terwilliger, T. C.; Adams, P. D. Automated structure solution with the PHENIX suite. In *Structural proteomics. Methods in Molecular Biology;* Humana Press, **2008**, pp 419–435.

(15) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **2011**, *67*, 355–367.

(16) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **1997**, *53*, 240–255.

(17) Hogben, H. J.; Krzystyniak, M.; Charnock, G. T. P.; Hore, P. J.; Kuprov, I. Spinach – A software library for simulation of spin dynamics in large spin systems. *J. Magn. Reson*. **2011**, *208*, 179–194.

(18) Abdelkader, E. H.; Qianzhu, H.; George, J.; Frkic, R.; Jackson, C.; Nitsche, C.; Otting, G.; Huber, T. Genetic encoding of cyanopyridylalanine for in-cell protein macrocyclization by the nitrile-aminothiol click reaction. ChemRxiv, Preprint, 10-27-**2021**. DOI: 10.33774/chemrxiv-2021-r5dn6.