Supporting Information

Guo et al. 10.1073/pnas.1419737111

SI Methods

Chemical Synthesis of L-Pyl. Enantiomerically pure L-Pyl was synthesized exactly as described previously (1).

Structure Determination of AcKRS3. All chemicals were purchased from Sigma–Aldrich. Overexpression and purification of the CTD (residues 188–454) of AckRS1 were performed in the same manner as for the CTD of PylRS described previously (2). Crystals of the CTD of AcKRS1 were obtained through multiple rounds of seeding by first using crystals of the CTD of PylRS (2). The CTD of AckRS1 crystallizes as rods and grows at 20 °C within a few days by vapor diffusion. The condition of 50 mM Tris (pH 8.0) and 23% (wt/vol) PEG 2K monoethyl-ether (mme) for crystallization was similar to the condition for the CTD of PylRS (2). Iterative seeding with crystals of the CTD of AcKRS1 increased the size of crystals in later rounds from 200 μm to 2 mm in the largest dimension. Crystals were grown in the presence of AcK (up to 75 mM), ATP, or AMP–β,γ-imidodiphosphate (PNP) (10 mM), and $MgCl_2$ (5 mM) or EDTA (5 mM). Diffraction data were collected from crystals grown, stabilized, and frozen in the presence of AcK (75 mM), ATP, or AMP-PNP (10 mM) and $MgCl_2$ (5 mM) in a stepwise manner.

The cryogens tested and used for successful data collection were 25% PEG 2Kmme and 30% (vol/vol) ethylene glycol or 35% PEG 2Kmme. The crystal mother solution was interchanged six times to exchange the crystallization solution into the cryogen solution (10%, 20%, 40%, 80%, and 100% cryogen). The time between each step was varied from 5 to 60 min. Crystals were frozen into liquid propane or liquid nitrogen, and no noticeable difference was observed in the diffraction patterns. Diffraction data were collected at 100 K using synchrotron X-ray radiation in beam line 24ID-C at the Advanced Photon Source at Argonne National Laboratory (Argonne, IL) and in beam line x29 at the National Synchrotron Light Source at Brookhaven National Laboratory (Upton, NY). The original data were processed and scaled with the HKL-2000 program suite (HKL Research) (3). General handling of scaled data was done using Collaborative Computational Project programs (4).

Structure Determination of IFRS Complexes with ncAA Substrates. Overproduction and purification of the CTD of IFRS were performed in the same manner as for the CTD of PylRS as described previously (2). The purified IFRS was mixed with 5 mM ATP, 5 mM $MgCl_2$, and 5 mM 3-I-Phe or 5 mM 3-Br-ThA and then incubated at room temperature for 30 min. Crystals of IFRS with 3-I-Phe or 3-Br-ThA were obtained by sitting drop vapor diffusion against 100 mM Hepes (pH 7.0), 5 mM $MgCl_2$, and 14% PEG3350 at 20 °C.

The diffraction datasets of IFRS complexes with 3-I-Phe or 3-Br-ThA were collected at 100 K in beam line 24ID-C at the Advanced Photon Source after crystals were soaked into reservoir solution buffer containing 35% ethylene glycol. Diffraction data of IFRS complexes with 3-I-Phe or 3-Br-ThA were indexed, integrated, scaled, and merged using XDS software (5).

After rigid body refinement with the structure of PylRS CTD (PDB ID code 2ZIM) as a start model using phenix.refine (6), the structures of IFRS with 3-I-Phe or 3-Br-ThA were rebuilt and modified manually by using Coot (7). Structure refinements were performed using phenix.refine. Finally, R_{work}/R_{free} factors of IFRS complexes with 3-I-Phe or 3-Br-ThA were converged to 17.3%/20.1% or 17.0%/21.1%, respectively. All figures of IFRS were prepared with PyMOL (PyMOL Molecular Graphics System, version 1.5.0.4).

IFRS-Directed Evolution. IFRS (PylRS-Asn346Ser/Cys348Ile) was selected from three PylRS plasmid [pKTS-MmPylS (8)] libraries. The libraries were generated as described previously (8) by overlap extension PCR using mutagenic primers with NNK (N = A, T, G or C; K = G or T) at the following positions: library 1, Asn346/Cys348/Leu305; library 2, Asn346/Cys348/Tyr306; and library 3, Asn346/Cys348/Leu309. The IFRS mutant (Table 2) was selected from library 1. Successive rounds of positive and negative selection were performed as previously described (8). The positive selection requires translation of the TAG codon at position 112 in chloramphenicol acetyltransferase [expression plasmid pCAT-pylT (8)] in the presence of 1 mM 3-I-Phe (Fig. 1*A*). Chloramphenicol-resistant colonies were pooled, and the selected library was then subjected to negative selection by translation of TAG codons at positions 13 and 44 in toxic protein ccdB [expression plasmid pAraCB2-pylT (8)] in absence of the 3-I-Phe. One subsequent round of positive selection produced to the IFRS mutant.

Preparation of *M. mazeri* AcKRS1, IFRS, MbPylRS and MmPylRS, and the Enzymatic Assay. Given the limited solubility of *M. mazeri* AcKRS1 (MmAcKRS1) and the dispensability of its N terminus for ATP-PP_i exchange activity (9), an N-terminal truncated MmAcKRS1 containing an N-terminal His_6-tag was made with an amino acid 185–454 sequence (8). The truncated MmPylRS and IFRS were also made in the same way for an ATP-PP_i exchange reaction. Aminoacylation activities were measured with full-length MbPylRS, MmPylRS, and IFRS by the method developed by Wolfson et al. (10). All of the enzymes were expressed in *E. coli* BL21(DE3) with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) induction and purified on an nickel-nitrilotriacetic acid (Ni-NTA) column.

Enzyme activity was tested in ATP-PP_i exchange experiments (11) using Pyl, 3-iodo-Phe, 3-bromo-Phe, 3-chloro-Phe, 3-trifluoromethyl-L-Phe, 3-methyl-Phe, 3-methyl-O-Phe, and N^{ε}-trifluoroacetyl-L-Lys. Reactions (final volume of 50 μL) were carried out at 37 °C in 100 mM Na-Hepes (pH 7.2); 10 mM $MgCl_2$; 50 mM KCl; 5 mM DTT; 2 mM potassium fluoride; 5 mM ATP; 5 mM PP_i; 0.2 μCi [γ-^{32}P]-ATP; 10 μM MmAcKRS1, 10 μM IFRS, or 2.5 μM MmPylRS; 0.025–0.8 mM Pyl; and 0.5–32 mM (all other amino acids). One microliter of the reaction mixture was spotted onto TLC PEI cellulose F plates (Merck), and the TLC plates were then developed in 1 M KH_2PO_4 and 1 M urea to separate ATP and PP_i for detection and quantification as described previously (12).

Aminoacylation reactions (25 μL) were carried out as before (13) at 37 °C in 100 mM Na-Hepes (pH 7.2), 25 mM $MgCl_2$, 60 mM NaCl, 5 mM ATP, 1 mM DTT, 15 μM $tRNA^{Pyl}$ 3′-labeled with [α-^{32}P]-ATP, and amino acids ranging in concentration from 0.25- to eightfold K_m concentrations. Enzyme concentrations were 0.2 μM for MbPylRS and MmPylRS and 1 μM for IFRS. Reactions were stopped by removing 2 μL of the reaction and adding it to 3 μL of 2.5 units/μL nuclease 5′-nucleotidehydrolase, 3′-phosphohydrolase (nuclease P1) and incubated at 25 °C for 30 min. One microliter of nuclease P1 digests was spotted onto TLC PEI cellulose F plates, and [α-^{32}P]AMP (from uncharged tRNA) and aminoacyl-[α-^{32}P]AMP (from charged tRNA) were separated in a running buffer of 0.1 M sodium acetate and 5% acetic acid and visualized by using a PhosphorImager system (Molecular Dynamics Storm 860 Molecular Imager, GE HealthCare).

sfGFP Screen for Incorporation of ncAAs. For the purpose of screening chemical libraries of ncAAs, a single colony of *E. coli* BL21 (DE3) cells was cotransformed with pET-sfGFP-pylT (14) and pKTS-AcKRS1 (8), pKTS-AcKRS3, or pKTS-IFRS. The stains were subsequently cultured in 100 mL of LB supplemented with 50 μg/mL kanamycin and 100 μg/mL ampicillin at 37 °C. Cells were pelleted and washed three times in minimal medium. The cells were grown in minimal medium at 37 °C until $A_{600} = 1.0$, when IPTG was added to a final concentration of 1 mM. Aliquots (50 μL of the induced cell suspension) were transferred into the wells of a 384-well and 96-well plates containing a different ncAA at 1 mM in each well. The cultures were incubated at 37 °C for 12 h under constant shaking in a Biotek plate reader. Fluorescence intensity was recorded every 10 min (excitation wavelength = 485/20 nm, emission wavelength = 528/20 nm). In Fig. 1 and Figs. S5 and S6, well A12 lacks added ncAAs and was used to remove the background signal; well A1 contains cells expressing a WT sfGFP construct with an AGC serine codon at position 2.

Chemical Libraries of ncAAs. Two different ncAA libraries were used with the above sfGFP screening system. The first library contains 313 amino acids, including the 20 standard amino acids and 293 ncAAs. The second library contains 94 ncAAs that are Phe and Lys derivatives. The sfGFP screening results from this library appear in Fig. 2 and Figs. S5 and S6. Full details of this chemical library can be found in Dataset S1, and microplate locations of each ncAA can be found elsewhere (supplementary material in ref. 14). With the exception of Pyl (see above), all ncAAs were commercially obtained (Sigma–Aldrich, ChemImpex, or Bachem).

ESI-MS and LC-MS/MS Analysis of Intact Proteins and Digested Peptide. ESI-MS of sfGFP variants containing ncAA **4** (sfGFP-**4**) and ncAA **10** (sfGFP-**10**) was performed on an AB Sciex QSTAR Elite mass spectrometer. A 5-μL aliquot of purified sfGFP (2–5 mg/mL) was diluted with 45 μL of 2% acetonitrile and 0.1% formic acid. The diluted sample was then desalted using C4 ZipTip (EMD Millipore), eluting into 100 μL of 60% acetonitrile/0.1% formic acid. A 25-μL portion of the desalted sample was then directly injected into the mass spectrometer.

sfGFP-**4** and sfGFP-**10** were selected for LC-MS/MS. The full-length protein was chymotrypsin-digested by a standard in-gel digestion protocol, and LC-MS/MS was performed on an LTQ Orbitrap XL equipped with a Waters nanoAcquity UPLC system (Thermo Scientific). A Waters Symmetry C18 180-μm × 20-mm trap column and a 1.7-μm, 100-μm × 250-mm nanoAcquity UPLC column (35 °C) were used for peptide separation. Trapping was done at 15 μL/min in 99% buffer A (100% water and 0.1% formic acid) for 1 min. Peptide separation was performed at 300 nL/min in buffer A (100% water containing 0.1% formic acid) and buffer B (100% CH_3CN containing 0.1% formic acid). A linear gradient (51 min) was run with 5% buffer B at initial conditions, 50% buffer B at 50 min, and 85% buffer B at 51 min. MS was acquired in the Orbitrap using one microscan and a maximum inject time of 900, followed by data-dependent MS/MS acquisitions in the ion trap (via collision-induced dissociation). The Mascot (Matrix Science) search algorithm was used with the expected ncAA input as a variable modification of canonical.

1. Wong ML, Guzei IA, Kiessling LL (2012) An asymmetric synthesis of L-pyrrolysine. *Org Lett* 14(6):1378–1381.
2. Kavran JM, et al. (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc Natl Acad Sci USA* 104(27):11268–11273.
3. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology: Macromolecular Crystallography, Part A* 276:307–326.
4. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235–242.
5. Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.
6. Afonine PV, et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68(Pt 4):352–367.
7. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.
8. Umehara T, et al. (2012) N-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo. *FEBS Lett* 586(6):729–733.
9. Herring S, et al. (2007) The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity. *FEBS Lett* 581(17): 3197–3203.
10. Wolfson AD, Pleiss JA, Uhlenbeck OC (1998) A new assay for tRNA aminoacylation kinetics. *RNA* 4(8):1019–1023.
11. Polycarpo CR, et al. (2006) Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett* 580(28-29):6695–6700.
12. Guo LT, et al. (2007) Human tryptophanyl-tRNA synthetase is switched to a tRNA-dependent mode for tryptophan activation by mutations at V85 and I311. *Nucleic Acids Res* 35(17):5934–5943.
13. Ambrogelly A, et al. (2007) Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc Natl Acad Sci USA* 104(9):3141–3146.
14. Ko JH, et al. (2013) Pyrrolysyl-tRNA synthetase variants reveal ancestral aminoacylation function. *FEBS Lett* 587(19):3243–3248.

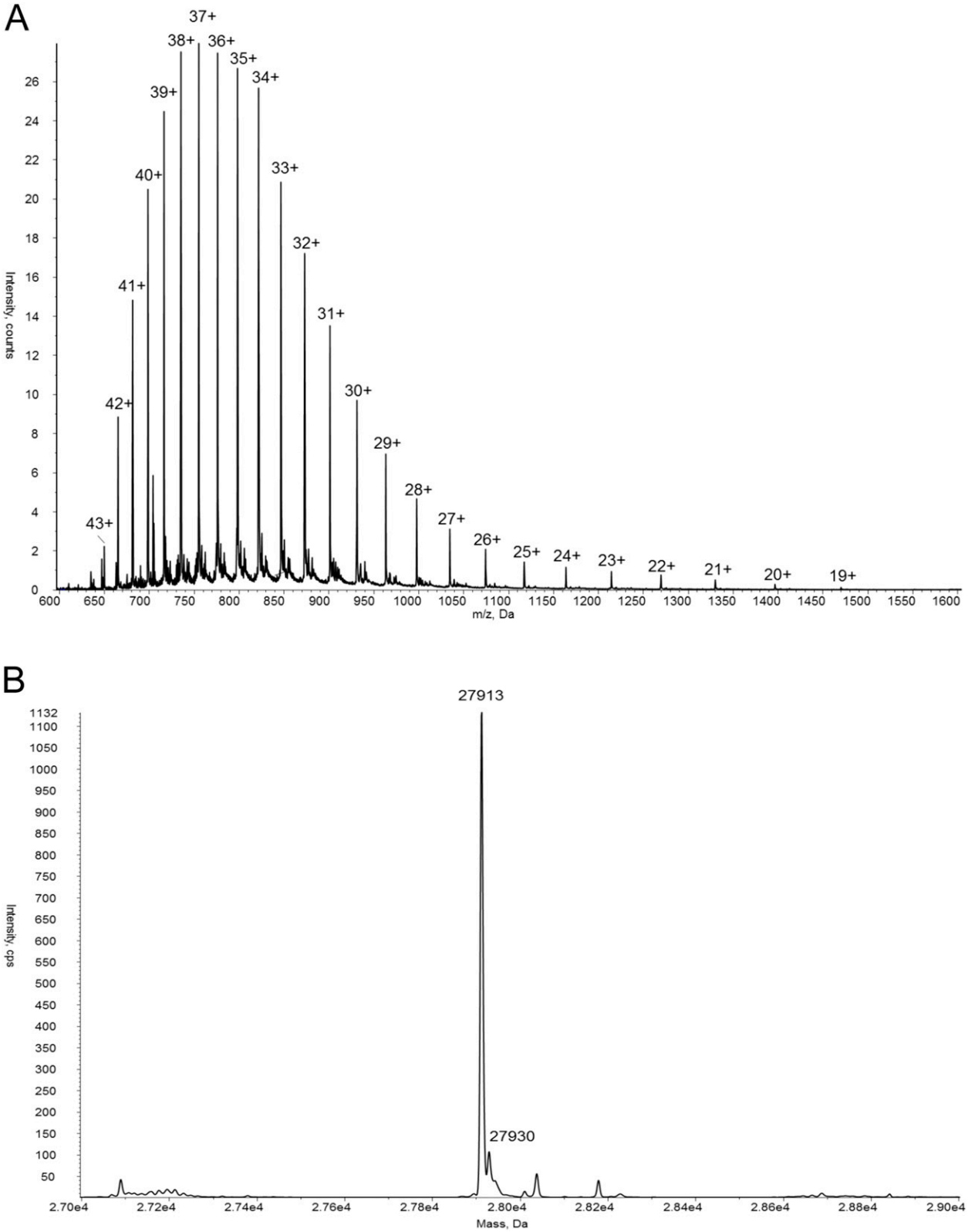

Fig. S1. Molecular mass determination of the protein sfGFP-**4**. The ESI-MS spectrum of sfGFP-**4** (*A*) and the deconvoluted ESI-MS spectrum of sfGFP-**4** (*B*) are shown. The calculated molecular mass is 27,914 Da; we found 27,913 Da and 27,930 Da (a full-length protein containing one oxidized methionine residue).

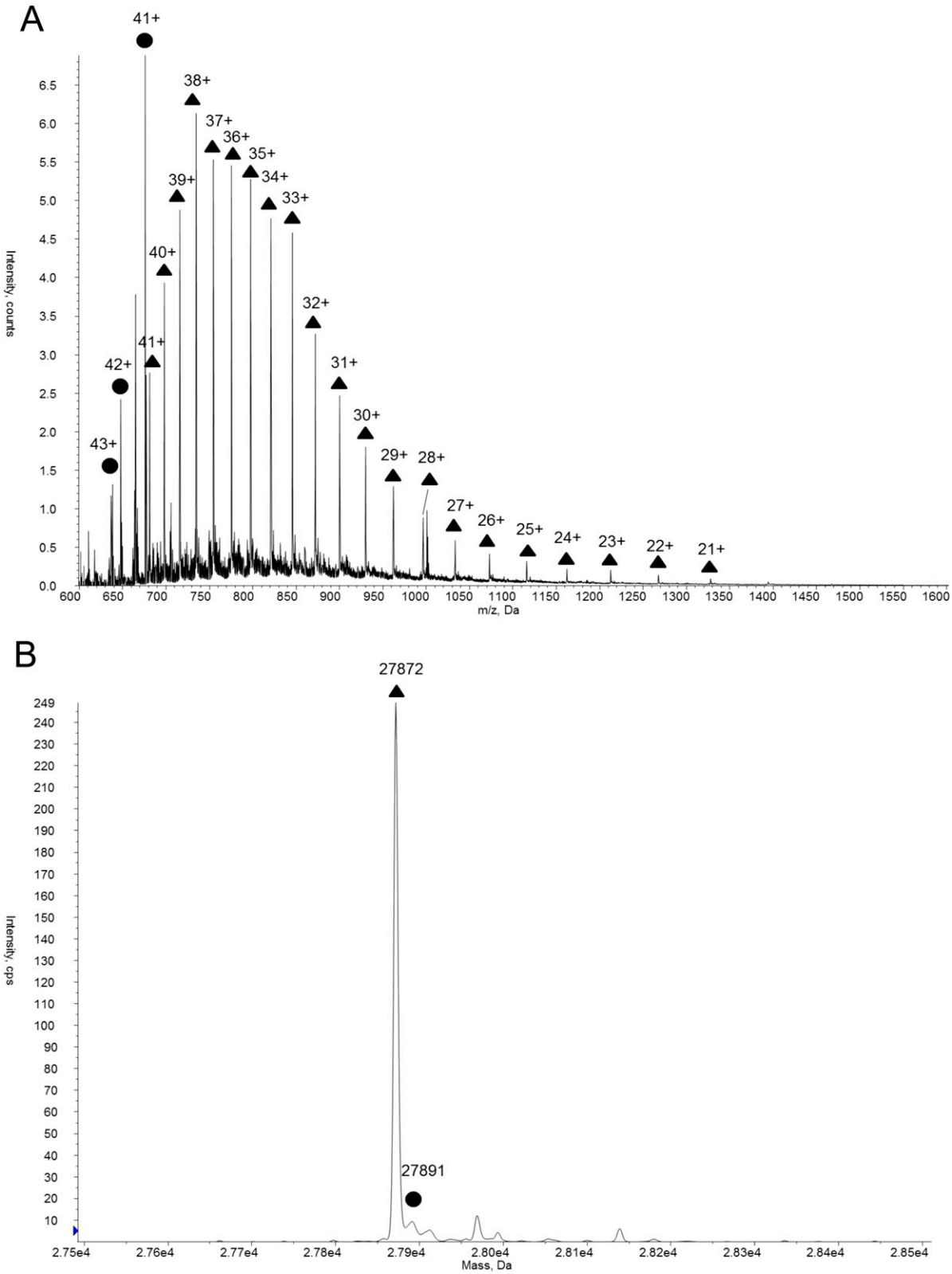

Fig. S2. Molecular mass determination of the protein sfGFP-**10**. The ESI-MS spectrum of sfGFP-**10** (*A*) and the deconvoluted ESI-MS spectrum of sfGFP-**10** (*B*) are shown. The calculated molecular mass is 27,872 Da; we found 27,872 Da and 27,891 Da (a full-length protein containing one oxidized methionine residue).

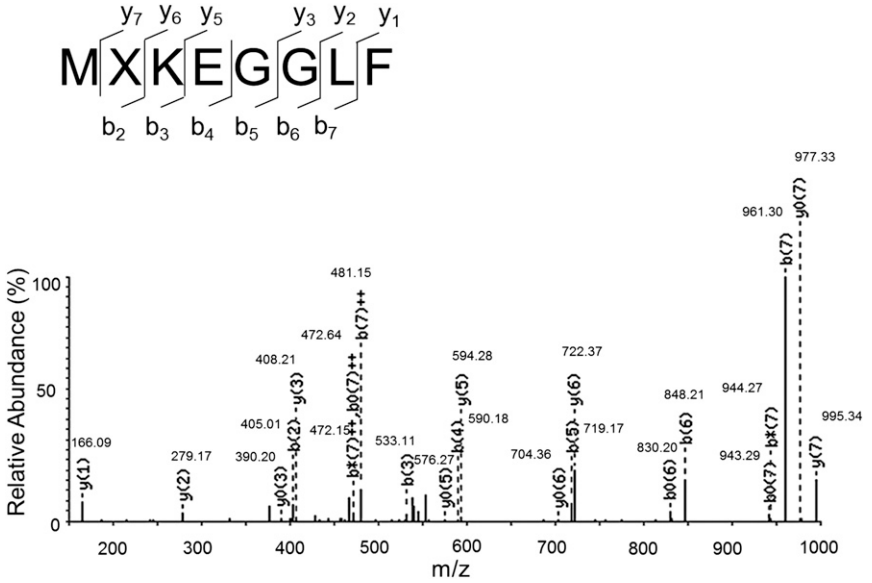

Fig. S3. Tandem mass spectra of the MXKGEELF (X denotes ncAA **4**) fragment from the full-length GFP-**4** protein. The full-length sfGFP-**4** protein was expressed using the AcKRS3-tRNA^{pyl} pair in the presence of 1 mM **4**. b^0, b-H_2O; y^0, y-H_2O.

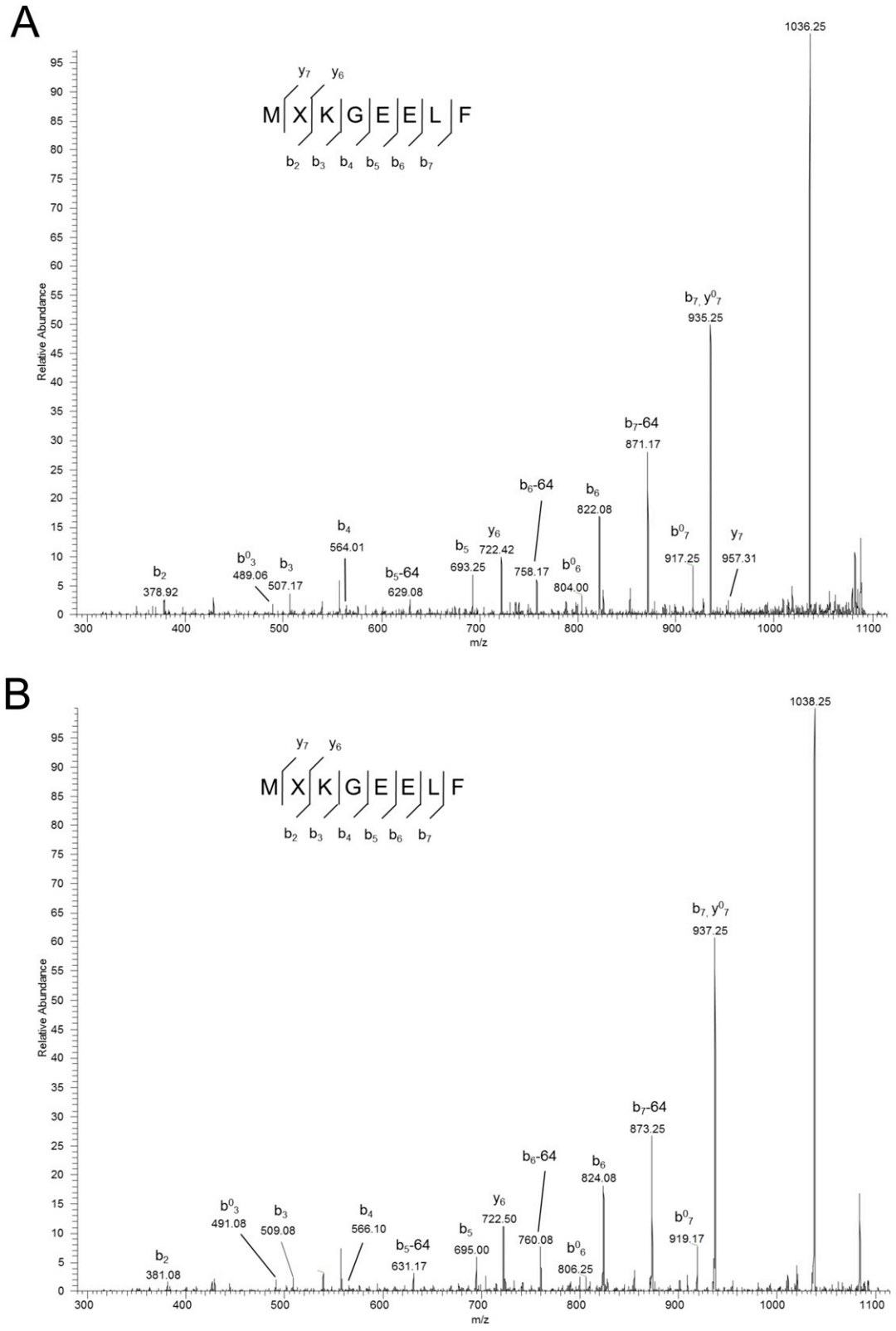

Fig. S4. Tandem mass spectra of the MXKGEELF (X denotes NAA **10**) fragment with methionine oxidation from the full-length GFP-**10** protein. (*A*) Fragments contained **10** with ^{79}Br isotope species. (*B*) Fragments contained **10** with ^{81}Br isotope species. The full-length GFP-**10** protein was expressed using the AcKRS3-tRNA$^{Pyl}_{CUA}$ pair in the presence of 2 mM **10**.

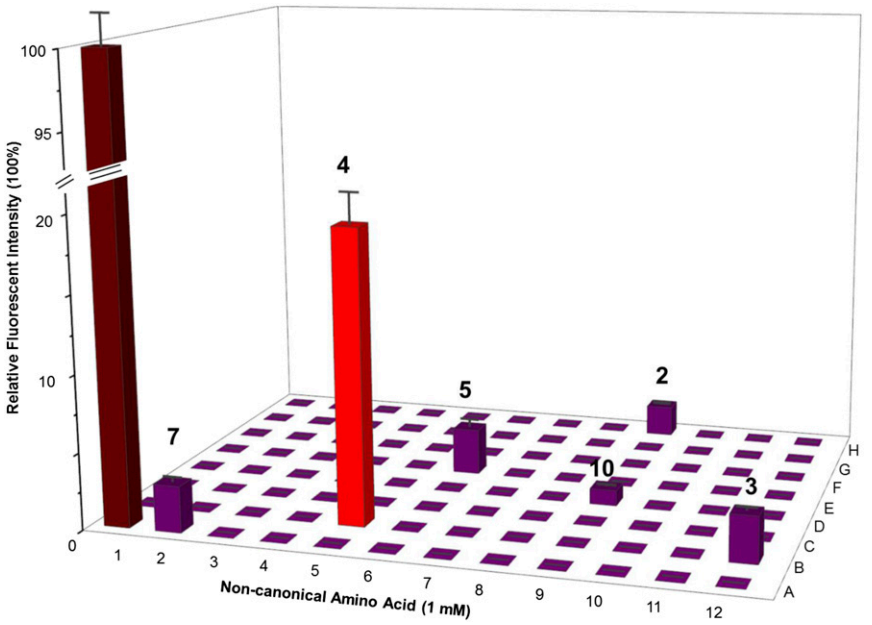

Fig. S5. Substrate specificity range of AcKRS1/tRNA^{Pyl}. Suppression of the *sfGFP-UAG2* gene by the library of ncAA-tRNA was measured by fluorescence intensity. Well A12 was a control experiment lacking ncAA to determine the background signal, which was subtracted from all experiments. Well A1 was a positive control experiment to detect production signal of WT sfGFP (100%). The *E. coli* strain BL21 (DE3) was used in all of these experiments. The 94 ncAA chemical names, structures, and microplate locations are provided in Dataset S1. The labeled ncAA chemical structures are given in scheme 1. Fluorescence data and error values (SD) are represented by bars from three independent experiments in each well and are also given in Dataset S1. Colors (0–4% in purple, 18–24% in red, and 95–100% in brown) have been used to indicate the level of UAG read-through.

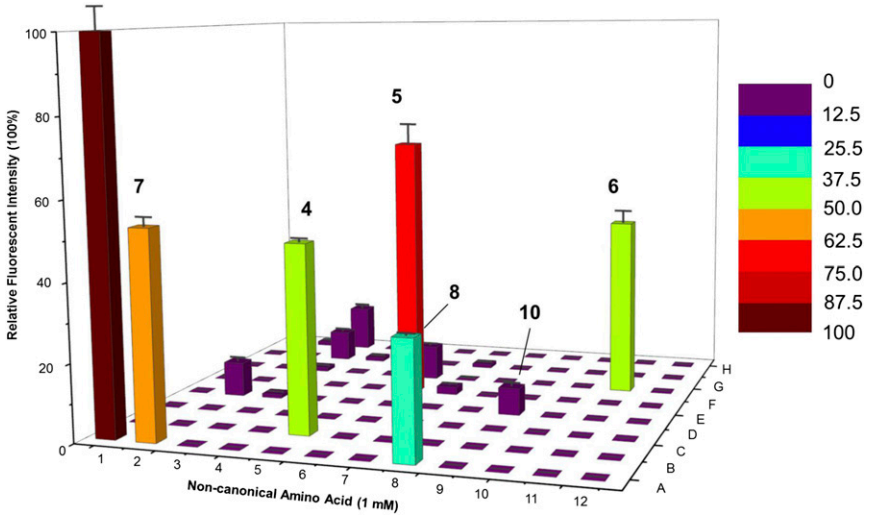

Fig. S6. Substrate specificity range of IFRS/tRNA^{Pyl}. Suppression of the *sfGFP-UAG2* gene by the library of ncAA-tRNA was measured by fluorescence intensity. Well A12 was a control experiment lacking ncAA to determine the background signal, which was subtracted from all experiments. Well A1 was a positive control experiment to detect production signal of WT sfGFP (100%). The *E. coli* strain BL21 (DE3) was used in all of these experiments. The 94 ncAA chemical names, structures, and microplate locations are provided in Dataset S1. The labeled ncAA chemical structures are given in scheme 1. Fluorescence data and error values (SD) are represented by bars from three independent experiments in each well and are also given in Dataset S1. Bar colors indicate the level of UAG read-through according to the scale shown.

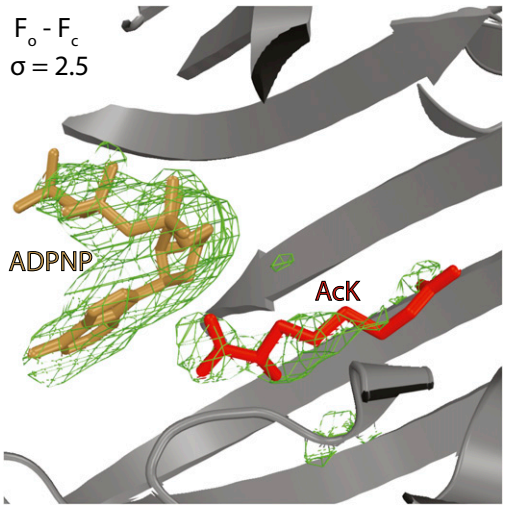

Fig. S7. Electron density in the active site of AcKRS3 (PDB ID code 4Q6G). The Fourier difference (F_o-F_c) electron density map illustrating density for the ADPNP and AcK substrates is shown.

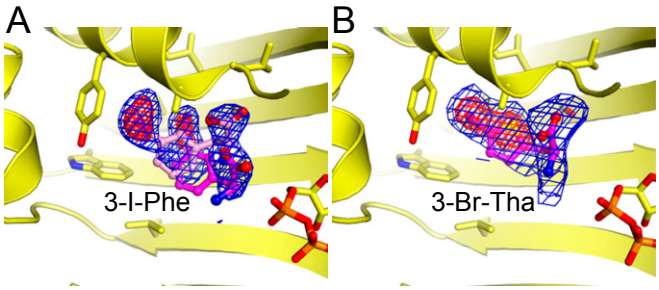

Fig. S8. Complex structures of IFRS with 3-I-Phe and ATP (PDB ID code 4TQD) (*A*) and 3-Br-ThA and ATP (PDB ID code 4TQF) (*B*). Close-up views of the active site are displayed, showing the residual Fourier difference (mF_o-dF_c) electron density map at 2σ (blue) and 4σ (red). The data indicate two possible positions for the iodo atom in the active site (*A*), suggesting two binding conformations of 3-I-Phe (magenta and pink).

Table S1. Sequence comparison of *M. mazei* PylRS variants

PylRS	L301	L305	Y306	L309	N346	C348	Source
AcKRS1	M	L	L	A	N	F	(1)
AcKRS2*	M	L	L	L	N	S	(1)
AcKRS3†	V	I	F	A	N	F	This work
IFRS	L	L	Y	L	S	I	This work

*AcKRS2 has an additional mutation: A315V.
†AcKRS3 has an additional mutation: M382S.

1. Umehara T, et al. (2012) N-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo. *FEBS Lett* 586(6):729–733.

Table S2. Data collection and refinement statistics

	AcKRS3	IFRS/3-I-Phe	IFRS/BrThA
Data collection			
Wavelength, Å	1.0809	0.9789	0.9202
Space group	$P6_4$	$P6_4$	$P6_4$
Cell dimensions			
a, *b*, *c*; Å	105.03, 105.03, 70.31	104.70, 104.70, 70.66	105.11, 105.11, 70.29
Resolution, Å	50–2.25 (2.33–2.25)*	42.1–2.1 (2.2–2.14)	38.2–2.7 (2.8–2.7)
R_{merge}	4.8 (53.8)	6.5 (70.7)	7.3 (80.2)
$I/\sigma I$	41.2 (2.6)	34.2 (4.32)	39.6 (4.63)
Completeness, %	98.7 (90.6)	99.8 (98.1)	99.9 (98.8)
Redundancy	7.5 (5.3)	20.6 (20.6)	20.1 (20.4)
Refinement			
Resolution, Å	50–2.25	42–2.14	38–2.7
No. of reflections	21,010	24,274	12,072
R_{work}/R_{free}	18.8/21.6	15.5/19.3	17.0/21.1
No. of atoms	2,394	2,335	2,249
Protein	2,220	2,129	2,119
Ligand/ion	68	88	73
Water	106	118	57
β-factors			
Protein	61.69	42.40	64.40
Ligand/ion	66.15	46.47	71.50
Water	58.00	45.80	57.00
rmsd			
Bond lengths, Å	0.006	0.014	0.005
Bond angles, °	1.09	1.45	0.91

$I/\sigma I$, average intensity of a group of reflections divided by the average standard deviation (σ) of the same group of reflections.
*Values in parentheses are for the highest resolution shell. Each dataset was collected from a single crystal.

Other Supporting Information Files

Dataset S1 (PDF)