Title: Orthogonal synthetases for polyketide precursors

One sentence summary: Structurally characterized PylRS variants accept non-L-α-amino acids and support their incorporation into hetero-oligomers.

Authors: Riley Fricke[1,2]†, Cameron Swenson[1,2]†, Leah Tang Roe[1,2], Noah Hamlish[2,3], Omer Ad[4], Sarah Smaga[1,2], Christine L. Gee[3], & Alanna Schepartz[1,2,3,5,6]*

Affiliations: [1]Department of Chemistry, University of California, Berkeley, CA 94720.
[2]Center for Genetically Encoded Materials, University of California, Berkeley, CA 94720.
[3]Department of Molecular and Cellular Biology, University of California, Berkeley, CA 94720.
[4]Department of Chemistry, Yale University, New Haven, CT 06520.
[5]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720.
[6]Chan Zuckerberg Biohub, San Francisco, CA 94158
*e-mail: schepartz@berkeley.edu
†These authors contributed equally
*To whom correspondence should be addressed

Abstract:

The absence of orthogonal aminoacyl-tRNA synthetases that accept non-L-α-amino acids is the primary bottleneck hindering the *in vivo* translation of sequence-defined hetero-oligomers. Here we report PylRS enzymes that accept α-hydroxy acids, α-thio acids, *N*-formyl-L-α-amino acids, and α-carboxyl acid monomers (malonic acids) that are formally precursors to polyketide natural products. These monomers are all accommodated and accepted by the translation apparatus *in vitro*. High-resolution structural analysis of the complex between one such PylRS enzyme and a *meta*-substituted 2-benzylmalonate derivative reveals an active site that discriminates *pro*-chiral carboxylates and accommodates the large size and distinct electrostatics of an α-carboxyl acid substituent. This work emphasizes the potential of PylRS for evolving new enzymes capable of encoding diverse non-L-α-amino acids in synergy with natural or evolved ribosomes.

Introduction:

All extant organisms biosynthesize polypeptides in an mRNA template-dependent manner using a translation apparatus composed of ribosomes, aminoacyl-tRNA synthetases, tRNAs, and a host of ancillary factors. Repurposing this translation apparatus for the templated synthesis of mixed-sequence oligomers–especially those containing non-L-α-amino acids–would provide a biological route to sequence-defined non-peptide polymers with novel, tunable, and evolvable properties as well as protein therapeutics with improved stability and expanded recognition potential. Chemical methods support the synthesis of certain sequence-defined (as opposed to sequence-controlled) non-peptide polymers (*1*, *2*) but are limited to small scale and generate considerable waste. Cellular methods based on the translation apparatus generate less waste, especially on a large scale, achieve greater chain lengths, and are scalable for industrial production. Over the past decade, it has become clear that many non-L-α-amino acids, including α-hydroxy acids (*3*, *4*), D-α- (*5*), linear (*6–8*) and cyclic (*9*) β-, γ- (*10*, *11*), and long chain amino acids (*12*), α-aminoxy and α-hydrazino acids (*13*), α-thio acids (*14*), aramids (*15*, *16*)–even 1,3-dicarbonyl monomers that resemble polyketide precursors (*16*)–are accepted by *E. coli* ribosomes in small-scale *in vitro* reactions. The structural and electronic diversity of these monomers reiterates the important role of proximity in promoting bond-forming reactions within the *E. coli* peptidyl transferase center (PTC) (*17*). However, there are scant examples in which non-L-α-amino acids have been incorporated into proteins *in vivo* (*18–23*) [see also (*24*)]. The absence of orthogonal aminoacyl-tRNA synthetase (aaRS) enzymes that accept non-L-α-amino acid substrates is the primary bottleneck limiting the production of sequence-defined non-peptide polymers *in vivo* using wild-type or engineered ribosomes.

In *E. coli*, two families of orthogonal aaRS enzymes have been employed widely for the incorporation of hundreds of different non-canonical α-amino acid monomers (*25–29*). The first includes pyrrolysyl-tRNA synthetase (PylRS) enzymes from methanogenic archaea and bacteria whose natural substrate is pyrrolysine (*30*), an L-α-amino acid found in the active sites of certain enzymes involved in methane metabolism (*31*). The second includes a large family of enzymes derived from *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS) (*25*, *32*). Towards the goal of charging tRNAs with non-L-α-amino acid monomers, PylRS enzymes have advantages of *Mj*TyrRS in how they recognize the α-amine of their substrate. While *Mj*TyrRS

recognizes the substrate α-amine *via* multiple, direct hydrogen bonds to the side chains of Q173, Q176, and Y151 (PDB: 1J1U) (*33*), *Methanosarcina mazei* PylRS (*Mm*PylRS) instead uses water-mediated hydrogen bonds to the N346 side chain and L301 and A302 backbone amides (PDB: 2ZCE; Fig. 1A) (*34*). These differences were exploited by Kobayashi *et al.* to acylate the cognate tRNA *Mm*-$tRNA^{Pyl}$ with a series of conservative N^{ε}-Boc-L-lysine (L-BocK) analogs containing -OH, -H and $-NHCH_3$ in place of the α-amine (Fig. 1B) (*20*).

We hypothesized that the water-mediated α-amine recognition employed by *Mm*PylRS could provide sufficient space and flexibility for substrates with less conservative substituents in place of the α-amine. Here we report that the tolerance of *Mm*PylRS for substrates with multiple alternative substituents in place of the α-amine extends to *Methanomethylophilus alvus* PylRS (*Ma*PylRS) (*35*, *36*) as well as *Ma*PylRS variants (*Ma*FRS1 and *Ma*FRS2 (*37*)) that recognize diverse L-phenylalanine derivatives. More importantly, *Ma*FRS1 and *Ma*FRS2 also accept phenylalanine derivatives with α-thio, *N*-formyl-L-α-amino, as well as an α-carboxyl substituent: 2-benzylmalonic acid. A final variant, *Ma*FRSA (*38*), is selective for ring-substituted 2-benzylmalonate derivatives over L-Phe. Malonates contain a 1,3-dicarbonyl unit that represents the defining backbone element of polyketide natural products, and after decarboxylation have the potential to support Claisen-type condensation within the PTC to form a carbon-carbon bond. Structural analysis of *Ma*FRSA complexed with a *meta*-substituted 2-benzylmalonate derivative and a non-hydrolyzable ATP analogue reveals how the enzyme uses a novel pattern of hydrogen bonds to differentiate the two pro-chiral carboxylates in the substrate and accommodate the larger size and distinct electrostatics of an α-carboxyl substituent. *In vitro* translation studies confirm that tRNAs carrying α-thio, α-carboxyl, and *N*-formyl-L-α-amino acid monomers are effectively delivered to and accommodated by *E. coli* ribosomes. This work describes the first orthogonal aminoacyl-tRNA synthetase that accepts α-thio acids and the α-carboxyl acid precursors capable of effecting carbon-carbon bond formation within the ribosome. These novel activities emphasize the potential of PylRS as a scaffold for evolving new enzymes capable of encoding more diverse non-L-α-amino acids in synergy with natural or evolved ribosomes.

Methods

I. Expression and purification of *Ma*PylRS, *Ma*FRS1, *Ma*FRS2, and *Ma*FRSA for use in biochemical experiments

Cloning. Plasmids used to express wild-type (WT) *Ma*PylRS (pET32a-*Ma*PylRS) and *Ma*FRS1 (pET32a-*Ma*FRS1) were constructed by inserting synthetic dsDNA fragments (Table S1) into the NdeI-NdeI cut sites of a pET32a vector using the Gibson method (*71*). pET32a-*Ma*FRS2 and pET32a-*Ma*FRSA were constructed from pET32a-*Ma*FRS1 using a Q5® Site-Directed Mutagenesis Kit (NEB). Primers pRF31 & pRF32, and pRF32 & pRF33 (Table S1) were used to construct pET32a-*Ma*FRS2 and pET32a-*Ma*FRSA, respectively. The sequences of the plasmids spanning the inserted regions were confirmed via Sanger sequencing at the UC Berkeley DNA Sequencing Facility using primers T7 F and T7 R (Table S1) and the complete sequence of each plasmid was confirmed by the Massachusetts General Hospital CCIB DNA Core.

Chemically competent cells and transformation. Chemically competent cells were prepared by following a modified published protocol (*72*). Briefly, 5 mL of LB was inoculated using a freezer stock of BL21-Gold (DE3)pLysS cells. The following day, 50 mL of LB was inoculated with 0.5 mL of the culture from the previous day and incubated at 37 ºC with shaking at 200 rpm until the culture reached an OD_{600} between 0.3-0.4. The cells were collected by centrifugation at 4303 × g for 20 min at 4 ºC. The cell pellet was resuspended in 5 mL of sterile filtered TSS solution (10% w/v polyethylene glycol 8000, 30 mM $MgCl_2$, 5% v/v DMSO in 25 g/L LB). The chemically competent cells were portioned into 100 µL aliquots in 1.5 mL microcentrifuge tubes, flash frozen in liquid N_2, and stored at -80 ºC until use. The following protocol was used to transform plasmids into chemically competent cells: 20 µL of KCM solution (500 mM KCl, 150 mM $CaCl_2$, 250 mM $MgCl_2$) was added to a 100 µL aliquot of cells on ice along with approximately 200 ng of the requisite plasmid and water to a final volume of 200 µL. The cells were incubated on ice for 30 min and then heat-shocked by placing them for 90 s in a water-bath heated to 42 ºC. Immediately after heat shock the cells were placed on ice for 2 min, after which 800 µL of LB was added. The cells then incubated at 37 ºC with shaking at 200 rpm for 60 min. The cells were plated onto LB-agar plates with the appropriate antibiotic and incubated overnight at 37 ºC.

Expression. Plasmids used to express wild type (WT) *Ma*PylRS, *Ma*FRS1, *Ma*FRS2 and *Ma*FRSA were transformed into BL21-Gold (DE3)pLysS chemically competent cells and plated onto LB agar plates supplemented with 100 µg/mL carbenicillin. Colonies were picked the following day and used to inoculate 10 mL of LB supplemented with 100 µg/mL carbenicillin. The cultures were incubated overnight at 37 ºC with shaking at 200 rpm. The following day the 10 mL cultures were used to inoculate 1 L of LB supplemented with 100 µg/mL carbenicillin in 4 L baffled Erlenmeyer flasks. Cultures were incubated at 37 ºC with shaking at 200 rpm for 3 h until they reached an OD_{600} of 0.6-0.8. At this point, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM and incubation was continued for 6 h at 30 ºC with shaking at 200 rpm. Cells were harvested by centrifugation at 4303 × g for 20 min at 4 ºC and the cell pellets were stored at -80 ºC until the expressed protein was purified as described below.

Purification. The following buffers were used for protein purification: Wash buffer: 50 mM sodium phosphate (pH 7.4), 500 mM NaCl, 20 mM β-mercaptoethanol (BME), 25 mM imidazole; Elution buffer: 50 mM sodium phosphate (pH 7.4), 500 mM NaCl, 20 mM β-mercaptoethanol (BME), 100 mM imidazole; Storage buffer: 100 mM HEPES-K, pH 7.2, 100 mM NaCl, 10 mM $MgCl_2$, 4 mM dithiothreitol (DTT), 20% v/v glycerol. 1 cOmplete Mini EDTA-free protease inhibitor tablet was added to Wash and Elution buffers immediately before use. To isolate protein, cell pellets were resuspended in Wash buffer (5 mL/g cells). The resultant cell paste was lysed at 4 ºC by sonication (Branson Sonifier 250) over 10 cycles consisting of 30 sec sonication followed by 30 sec manual swirling. The lysate was centrifuged at 4303 × g for 10 min at 4 °C to separate the soluble and insoluble fractions. The soluble lysate was incubated at 4 ºC with 1 mL of Ni-NTA agarose resin (washed with water and equilibrated with Wash buffer) for 2 h. The lysate-resin mixture was added to a 65 g RediSep® Disposable Sample Load Cartridge (Teledyne ISCO) and allowed to drain at RT. The protein-bound Ni-NTA agarose resin was then washed with three 10 mL aliquots of Wash buffer. The protein was eluted from Ni-NTA agarose resin by rinsing the resin three times with 10 mL Elution buffer. The elution fractions were pooled and concentrated using a 10 kDa MWCO Amicon® Ultra-15 Centrifugal Filter Unit (4303 × g, 4 ºC). The protein was then buffer-exchanged into Storage buffer until the [imidazole] was < 5 µM using the same centrifugal filter unit. The protein was dispensed into 20

µL single-use aliquots and stored at -80 ºC for up to 8 months. Protein concentration was measured using the Bradford assay (*73*). Yields were between 8 and 12 mg/L. Proteins were analyzed by SDS-PAGE (fig. S2A) using Any kD™ Mini-PROTEAN® TGX™ Precast Protein Gels (BioRad). The gels were run at 200 V for 30 min.

LC-MS analysis. Proteins were analyzed by LC-MS to confirm their identities (fig. S2B). Samples analyzed by mass spectrometry were resolved using a Poroshell StableBond 300 C8 (2.1 x 75 mm, 5 µm, Agilent Technologies part #660750-906) using a 1290 Infinity II UHPLC (G7120AR, Agilent). The mobile phases used for separation were (A) 0.1% formic acid in water and (B) 100% acetonitrile, and the flow rate was 0.4 mL/min. After an initial hold at 5% (B) for 0.5 min, proteins were eluted using a linear gradient from 5 to 75% (B) for 9.5 min, a linear gradient from 75 to 100% (B) for 1 min, a hold at 100% (B) for 1 min, a linear gradient 100 to 5% (B) for 3.5 min, and finally a hold at 5% (B) for 4.5 min. Protein masses were analyzed using LC-HRMS with an Agilent 6530 Q-TOF AJS-ESI (G6530BAR). The following parameters were used: gas temperature 300 ºC, drying gas flow 12 L/min, nebulizer pressure 35 psi, sheath gas temperature 350 ºC, sheath gas flow 11 L/min, fragmentor voltage 175 V, skimmer voltage 65 V, Oct 1 RF Vpp 750 V, Vcap 3500 V, nozzle voltage 1000 V, 3 spectra/s.

Size Exclusion Chromatography. Analytical size exclusion chromatography (fig. S2C) was performed on an ÄKTA Pure 25. A flow rate of 0.5 mL/min was used for all steps. A Superdex® 75 Increase 10/300 GL column (stored and operated at 4 °C) was washed with 1.5 column volumes (CV) of degassed and sterile filtered MilliQ water. The column equilibrated in 1.5 column volumes of SEC Buffer: 100 mM HEPES (pH 7.2), 100 mM NaCl, 10 mM $MgCl_2$, 4 mM DTT. Approximately 800 µg of protein in 250 µL SEC Buffer was loaded into a 500 µL capillary loop. The sample loop was washed with 2.0 mL of SEC Buffer as the sample was injected onto the column. The sample was eluted in 1.5 column volumes of SEC Buffer and analyzed by UV absorbance at 280 nm.

II. Transcription and purification of tRNAs

DNA templates. The DNA template used for transcribing *M. alvus* $tRNA^{Pyl}$ (*Ma*-$tRNA^{Pyl}$) (*36*) was prepared by annealing and extending the ssDNA oligonucleotides *Ma*-PylT-F and *Ma*-PylT-R (2 mM, Table S1) using OneTaq 2x Master Mix (NEB). The annealing and extension used the following protocol on a thermocycler (BioRad C1000 Touch™): 94 ºC for 30 s, 30 cycles of [94 ºC for 20 s, 53 ºC for 30 s, 68 ºC for 60 s], 68 º C for 300 s. Following extension, the reaction mixture was supplemented with sodium acetate (pH 5.2) to a final concentration of 300 mM, washed once with 1:1 (*v*/*v*) acid phenol:chloroform, twice with chloroform, and the dsDNA product precipitated upon addition of ethanol to a final concentration of 71%. The pellet was resuspended in water and the concentration of dsDNA determined using a NanoDrop ND-1000 (Thermo Scientific). The template begins with a single C preceding the T7 promoter, which increases yields of T7 transcripts (*74*). The penultimate residue of *Ma*-PylT-R carries a 2'-methoxy modification, which reduces non-templated nucleotide addition by T7 RNA polymerase during *in vitro* transcription (*75*).

Transcription and purification. *Ma*-$tRNA^{Pyl}$ was transcribed *in vitro* using a modified version of a published procedure (*76*). Transcription reactions (25 µL) contained the following components: 40 mM Tris-HCl (pH 8.0), 100 mM NaCl, 20 mM DTT, 2 mM spermidine, 5 mM adenosine triphosphate (ATP), 5 mM cytidine triphosphate (CTP), 5 mM guanosine triphosphate (GTP), 5 mM uridine triphosphate (UTP), 20 mM guanosine monophosphate (GMP), 0.2 mg/mL bovine serum albumin, 20 mM $MgCl_2$, 12.5 ng/µL DNA template, 0.025 mg/mL T7 RNA polymerase. These reactions were incubated at 37 ºC in a thermocycler for 3 h. Four 25 µL reactions were pooled, and sodium acetate (pH 5.2) was added to a final concentration of 300 mM in a volume of 200 µL. The transcription reactions were extracted once with 1:1 (*v*/*v*) acid phenol:chloroform, washed twice with chloroform, and the tRNA product precipitated by adding ethanol to a final concentration of 71%. After precipitation, the tRNA pellet was resuspended in water and incubated with 8 U of RQ1 RNAse-free DNAse (Promega) at 37 ºC for 30 min according to the manufacturer's protocol. The tRNA was then washed with phenol:chloroform and chloroform as described above, precipitated, and resuspended in water. To remove small molecules, the tRNA was further purified using a Micro Bio-Spin™ P-30 Gel Column, Tris Buffer RNase-free (Bio-Rad) after first exchanging the column buffer to water according to the

manufacturer's protocol. The tRNA was precipitated once more, resuspended in water, quantified using a NanoDrop ND-1000, aliquoted, and stored at -20 ºC.

Urea-PAGE analysis. tRNA was analyzed by Urea-PAGE (fig. S2D**)** using a 10% Mini-PROTEAN® TBE-Urea Gel (BioRad). The gels were run at 120 V for 30 min then stained with SYBR-Safe gel stain (Thermo-Fisher) for 5 minutes before imaging.

LC-MS analysis. *Ma*-tRNA^{Pyl} was analyzed by LC-MS to confirm its identity. Samples were resolved on a ACQUITY UPLC BEH C18 Column (130 Å, 1.7 µm, 2.1 mm X 50 mm, Waters part # 186002350, 60 ºC) using an ACQUITY UPLC I-Class PLUS (Waters part # 186015082). The mobile phases used were (A) 8 mM triethylamine (TEA), 80 mM hexafluoroisopropanol (HFIP), 5 µM ethylenediaminetetraacetic acid (EDTA, free acid) in 100% MilliQ water; and (B) 4 mM TEA, 40 mM HFIP, 5 µM EDTA (free acid) in 50% MilliQ water/50% methanol. The method used a flow rate of 0.3 mL/min and began with Mobile Phase B at 22% that increased linearly to 40 % B over 10 min, followed by a linear gradient from 40 to 60% B for 1 min, a hold at 60% B for 1 min, a linear gradient from 60 to 22% B over 0.1 min, then a hold at 22% B for 2.9 min. The mass of the RNA was analyzed using LC-HRMS with a Waters Xevo G2-XS Tof (Waters part #186010532) in negative ion mode with the following parameters: capillary voltage: 2000 V, sampling cone: 40, source offset: 40, source temperature: 140 ºC, desolvation temperature: 20 ºC, cone gas flow: 10 L/h, desolvation gas flow: 800 L/h, 1 spectrum/s. Expected masses of oligonucleotide products were calculated using the AAT Bioquest RNA Molecular Weight Calculator (*77*). Deconvoluted mass spectra were obtained using the MaxEnt software (Waters Corporation).

III. **Procedure for RNAse A assays**

tRNA acylation. Reaction mixtures (25 µL) used to acylate tRNA contained the following components: 100 mM Hepes-K (pH 7.5), 4 mM DTT, 10 mM $MgCl_2$, 10 mM ATP, 0 - 10 mM substrate, 0.1 U *E. coli* inorganic pyrophosphatase (NEB), 25 µM *Ma*-tRNA^{Pyl}, and 2.5 µM enzyme (*Ma*PylRS, *Ma*FRS1, *Ma*FRS2, or *Ma*FRSA). Reaction mixtures were incubated at 37 °C in a dry-air incubator for 2 h.

Digestion of tRNA products with RNAse A. tRNA samples from enzymatic acylation reactions were quenched with 27.5 µL of RNAse A solution (1.5 U/µL RNAse A (Millipore-Sigma), 200 mM sodium acetate, pH 5.2) and incubated for 5 min at room temperature. Proteins were then precipitated upon addition of 50% trichloroacetic acid (TCA, Sigma-Aldrich) to a final concentration of 5%. After precipitating protein at -80 ºC for 30 min, insoluble material was removed by centrifugation at 21,300 × g for 10 min at 4 °C. The soluble fraction was then transferred to autosampler vials, kept on ice until immediately before LC-MS analysis, and returned to ice immediately afterwards.

Analysis via mass spectrometry. Samples analyzed by mass spectrometry were resolved using a Zorbax Eclipse XDB-C18 RRHD column (2.1 × 50 mm, 1.8 µm, room temperature, Agilent Technologies part # 981757-902) fitted with a guard column (Zorbax Eclipse XDB-C18, 2.1 x 5 mm 1.8 µm, Agilent part # 821725-903) using a 1290 Infinity II UHPLC (G7120AR, Agilent). The mobile phases used were (A) 0.1% formic acid in water; and (B) 100% acetonitrile. The method used a flow rate of 0.7 mL/min and began with Mobile Phase B held at 4% for 1.35 min, followed by a linear gradient from 4 to 40% B over 1.25 min, a linear gradient from 40 to 100% B over 0.4 min, a linear gradient from 100 to 4% B over 0.7 min, then finally B held at 4% for 0.8 min. Acylation was confirmed by correctly identifying the exact mass of the 2' and 3' acyl-adenosine product corresponding to the substrate tested in the extracted ion chromatogram by LC-HRMS with an Agilent 6530 Q-TOF AJS-ESI (G6530BAR). The following parameters were used: fragmentor voltage of 175 V, gas temperature of 300ºC, gas flow of 12 L/min, sheath gas temperature of 350ºC, sheath gas flow of 12 L/min, nebulizer pressure of 35 psi, skimmer voltage of 65 V, Vcap of 3500 V, and collection rate of 3 spectra/s. Expected exact masses of acyl-adenosine nucleosides (Table S2) were calculated using ChemDraw 19.0 and extracted from the total ion chromatograms ± 100 ppm.

IV. Malachite green assay to monitor adenylation

Procedure. Enzymatic adenylation reactions were monitored using malachite green using a previous protocol with modifications (*50*). Each adenylation reaction (60 µL) contained the following components: 200 mM HEPES-K (pH 7.5), 4 mM DTT, 10 mM $MgCl_2$, 0.2 mM ATP, 0 - 10 mM substrate, 4 U/mL *E. coli* inorganic pyrophosphatase (NEB), and 2.5 µM enzyme

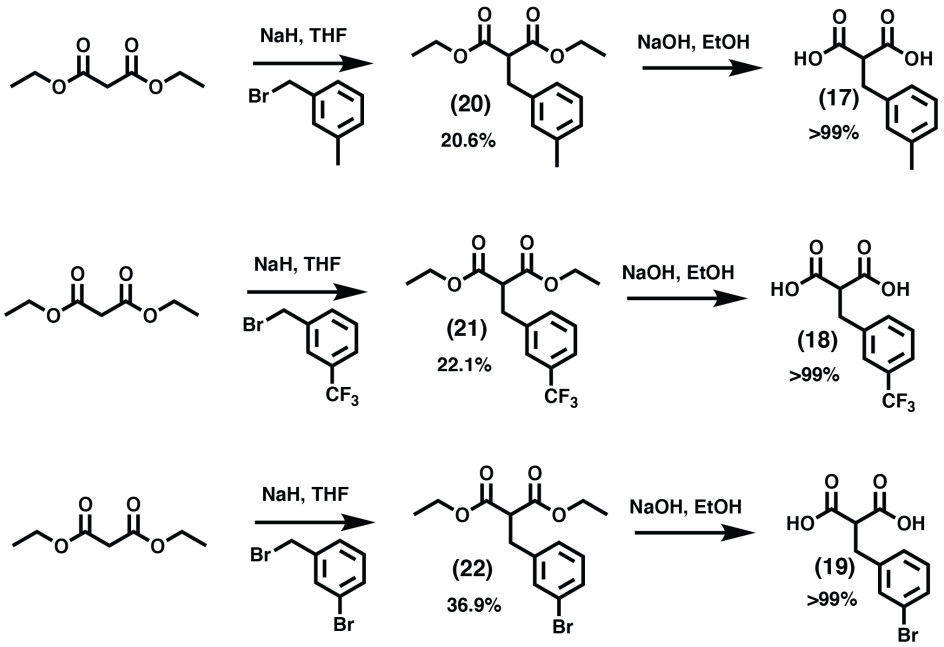
(*Ma*FRS1 or *Ma*FRSA). Adenylation reactions were incubated at 37 °C in a dry-air incubator. Aliquots (10 µL) were withdrawn after 0, 5, 10, 20, and 30 min and quenched upon addition to an equal volume of 20 mM EDTA (pH 8.0) on ice. Once all aliquots were withdrawn, 80 µL of Malachite Green Solution (Echelon Biosciences) was added to each aliquot and the mixture incubated at RT for 30 min. After shaking for 30 sec to remove bubbles, the absorbance at 620 nm was measured on a Synergy HTX plate reader (BioTek). The absorbance was then converted to phosphate concentration using a phosphate standard curve (0 - 100 µM) and plotted over time to determine turnover numbers.

V. Synthesis of *meta*-substituted 2-benzylmalonates 17-19 (Scheme 1) and malonate 16 (Scheme 2).

General notes. Alkylation reactions to synthesize **20-23** as well as hydrolysis reactions to synthesize **16-19** were based on published methods (*79*, *80*). All reagents and solvents were used as received from commercial suppliers, unless indicated otherwise. Alkylation reactions were carried out with exclusion of air and moisture. Room temperature is considered 20–23 °C. Stirring was achieved with Teflon-coated magnetic stir bars. TLC was performed on glass-backed silica gel plates (median pore size 60 Å) and visualized using UV light at 254 nm or staining with iodine. Column chromatography was performed on an Isco Teledyne Combiflash Nextgen 300+ instrument using pre-packed Redi-sep Gold silica gel cartridges (particle diameter 20-40 µM, pore diameter 60 Å). The eluents are given in brackets. Mass spectrometry was performed on an LTQ FT-ICR mass spectrometer equipped with an electrospray ionization source (Finnigan LTQ FT, Thermo Fisher Scientific, Waltham, MA) operated in either positive or negative ion mode. ^{1}H NMR spectra NMR data were acquired at 298 K using a 500 MHz Bruker Avance Neo NMR spectrometer that was equipped with a 5 mm iProbe or a 400 MHz Bruker Avance I spectrometer equipped with a 5 mm BBO Smart Probe. The experiments were conducted using the default Bruker NMR parameters and data was time-averaged until a sufficient level of sensitivity was achieved. ^{1}H NMR data was calibrated by using the residual peak of the solvent as the internal standard ($CDCl_3$: δ_H = 7.26 ppm; CD_3OD: δ_H = 3.31 ppm). All coupling constants are recorded in Hz. NMR spectra were processed with MestReNova v14.1.2-25024 software using the baseline and phasing correction features. Multiplicities and coupling

constants were calculated using the multiplet analysis feature with manual intervention as necessary.

Diethyl 2-(3-methylbenzyl)malonate (20) Diethyl malonate (500.57 mg, 3.125 mmol, 1.05 equiv.) was added dropwise to a suspension of 60% NaH on mineral oil (125 mg, 3.125 mmol, 1.05 equiv.) in 6 mL dry THF at 0 °C. After 20 min, 3-methylbenzyl bromide (550.86 mg, 2.97 mmol, 1 equiv.) was added in one portion and the reaction mixture was refluxed overnight. The next day, the reaction was cooled and quenched by the addition of H_2O. Et_2O was added and the aqueous layer was extracted three times with Et_2O. The combined organic layers were dried over Na_2SO_4 then evaporated to dryness under reduced pressure. The crude product was purified by flash chromatography on SiO_2 [eluent: EtOAc/hexane (5% then 10% then 15% then 20%)] to obtain pure diethyl 2-(3-methylbenzyl)malonate **20** as a clear liquid. Yield 20.6%. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.19 (t, J = 7.8 Hz, 1H), 7.07 – 7.00 (m, 3H), 4.19 (qd, J = 7.1, 2.8 Hz, 4H), 3.65 (t, J = 7.8 Hz, 1H), 3.20 (d, J = 7.8 Hz, 2H), 2.34 (s, 3H), 1.24 (t, J = 7.1 Hz, 6H). HR-EI-MS [M+H]+: calculated for C15H21O4+, m/z 265.1434, found m/z 265.1395.



Scheme 1

2-(3-methylbenzyl)malonic acid (17) Diethyl 2-(3-methylbenzyl)malonate **20** (100 mg, 0.362 mmol) was dissolved in 1 mL of ethanol then added dropwise to 375 µL of 6.67 M NaOH. The

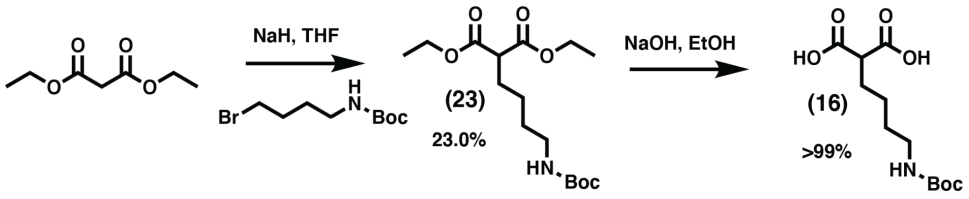
mixture was stirred for 5 h at 60 °C. The solution was then cooled to 0 °C, carefully acidified to pH 1 with 1 N HCl, and extracted with 5 portions of Et_2O. The combined extracts were washed with a saturated aqueous solution of NaCl, dried over Na_2SO_4 and evaporated to dryness. 2-(3-methylbenzyl)malonic acid was dissolved in 1:1 H_2O:MeCN and lyophilized to give a white solid. Yield >99%. ^{1}H NMR (500 MHz, MeOD) δ 7.04 (t, J = 7.6 Hz, 1H), 6.95 (s, 1H), 6.94 – 6.88 (m, 2H), 3.49 (t, J = 7.8 Hz, 1H), 3.01 (d, J = 7.7 Hz, 2H), 2.20 (s, 3H). HR-ESI-MS [M-H]-: calculated for C11H11O4, m/z 207.0663, found m/z 207.0661.

Diethyl 2-(3-(trifluoromethyl)benzyl)malonate (21) Diethyl malonate (500.57 mg, 3.125 mmol, 1.05 equiv.) was added dropwise to a suspension of 60% NaH on mineral oil (125 mg, 3.125 mmol, 1.05 equiv.) in 6 mL dry THF at 0°C. After 20 min, 3-(trifluoromethyl)benzyl bromide (711.377 mg, 2.97 mmol, 1 equiv.) was added in one portion and the reaction mixture was refluxed overnight. The next day, the reaction was cooled and quenched by the addition of H_2O. Et_2O was added and the aqueous layer was extracted three times with Et_2O. The combined organic layers were then dried over Na_2SO_4 then evaporated to dryness under reduced pressure. The crude product was purified by flash chromatography on SiO_2 [eluent: EtOAc/hexane (5% then 10% then 15% then 20%)] to obtain pure diethyl 2-(3-(trifluoromethyl)benzyl)malonate as a clear liquid. Yield 22.1%. ^{1}H NMR (400 MHz, CDCl3) δ 7.56 – 7.49 (m, 2H), 7.45 (p, J = 2.1 Hz, 2H), 4.21 (qd, J = 7.2, 1.7 Hz, 4H), 3.68 (t, J = 7.8 Hz, 1H), 3.31 (d, J = 7.9 Hz, 2H), 1.25 (t, J = 7.1 Hz, 6H). HR-EI-MS [M+H]+: calculated for C15H18F3O4+, m/z 319.1152, found m/z 319.1112.

2-(3-(trifluoromethyl)benzyl)malonic acid (18) Diethyl 2-(3-(trifluoromethyl)benzyl)malonate (100 mg, 0.314 mmol) was dissolved in 1 mL of ethanol then added dropwise to 375 µL of 6.67 M NaOH. The mixture was stirred for 5 h at 60 °C. The solution was then cooled to 0°C, carefully acidified to pH 1 with 1 N HCl, and extracted with 5 portions of Et_2O. The combined extracts were washed with a saturated solution of NaCl, dried over Na_2SO_4, and evaporated to dryness. 2-(3-(trifluoromethyl)benzyl)malonic acid was dissolved in 1:1 H_2O:MeCN and lyophilized to give a white solid. Yield >99%. ^{1}H NMR (500 MHz, MeOD) δ 7.60 – 7.46 (m, 4H), 3.69 (t, J = 7.8 Hz, 1H), 3.26 (d, J = 7.7 Hz, 2H). HR-ESI-MS [M-H]-: calculated for C11H8F3O4, m/z 261.0380, found m/z 261.0377.

Diethyl 2-(3-bromobenzyl)malonate (22) Diethyl malonate (500.57 mg, 3.125 mmol, 1.05 equiv.) was added dropwise to a suspension of 60% NaH on mineral oil (125 mg, 3.125 mmol, 1.05 equiv.) in 6 mL dry THF at 0 °C. After 20 min, 3-bromobenzyl bromide (743.91 mg, 2.97 mmol, 1 equiv.) was added in one portion and the reaction mixture was refluxed overnight. The next day, the reaction was cooled and quenched by the addition of H_2O. Et_2O was added and the aqueous layer was extracted three times with Et_2O. The combined organic layers were then dried over Na_2SO_4 then evaporated to dryness under reduced pressure. The crude product was purified by flash chromatography on SiO_2 [eluent: EtOAc/hexane (5% then 10% then 15% then 20%)] to obtain pure diethyl 2-(3-(trifluoromethyl)benzyl)malonate as a clear liquid. Yield 22.1%. ^{1}H NMR (400 MHz, CDCl3) δ 7.39 – 7.32 (m, 2H), 7.19 – 7.10 (m, 2H), 4.17 (qd, J = 7.1, 1.3 Hz, 4H), 3.61 (t, J = 7.8 Hz, 1H), 3.18 (d, J = 7.9 Hz, 2H), 1.22 (t, J = 7.1 Hz, 6H). HR-EI-MS [M+H]+: calculated for C14H18BrO4, m/z 329.0383, found m/z 329.0343.

2-(3-bromobenzyl)malonic acid (19) Diethyl 2-(3-bromobenzyl)malonate **22** (100 mg, 0.305 mmol) was dissolved in 1 mL of ethanol then added dropwise to 375 µL of 6.67 M NaOH. The mixture was stirred for 5 h at 60 °C. The solution was then cooled to 0 °C, carefully acidified to pH 1 with 1 N HCl, and extracted with 5 portions of Et_2O. The combined extracts were washed with a saturated solution of NaCl, dried over Na_2SO_4, and evaporated to dryness. 2-(3-bromobenzyl)malonic acid was dissolved in 1:1 H_2O:MeCN and lyophilized to give a white solid. Yield >99%. ^{1}H NMR (500 MHz, MeOD) δ 7.33 (t, J = 1.9 Hz, 1H), 7.26 (dt, J = 7.7, 1.7 Hz, 1H), 7.16 – 7.06 (m, 2H), 3.52 (t, J = 7.8 Hz, 1H), 3.04 (d, J = 7.8 Hz, 2H). HR-ESI-MS [M-H]-: calculated for C10H8BrO4, m/z 270.9611, found m/z 270.9609.



Scheme 2

Diethyl 2-(4-((tert-butoxycarbonyl)amino)butyl)malonate (23) Diethyl malonate (166.86 mg, 1.04 mmol, 1.05 equiv.) was added dropwise to a suspension of 60% NaH on mineral oil (41.67 mg, 1.04 mmol, 1.05 equiv.) in 6 mL dry THF at 0 °C. After 20 min, 4-(Boc-amino)butyl

bromide (250.17 mg, 0.99 mmol, 1 equiv.) was added in one portion and the reaction mixture was refluxed overnight. The next day, the reaction was cooled and quenched by the addition of H_2O. Et_2O was added and the aqueous layer was extracted three times with Et_2O. The combined organic layers were then dried over Na_2SO_4 then evaporated to dryness under reduced pressure. The crude product was purified by flash chromatography on SiO_2 [eluent: EtOAc/hexane (5% then 10% then 15% then 20%)] to obtain pure diethyl 2-(4-((tert-butoxycarbonyl)amino)butyl)malonate as a clear liquid. Yield 22.1%. ^{1}H NMR (500 MHz, CDCl3) δ 4.44 (s, 1H), 4.13 (qd, J = 7.1, 2.0 Hz, 4H), 3.24 (t, J = 7.5 Hz, 1H), 3.04 (q, J = 6.7 Hz, 2H), 1.87 – 1.79 (m, 2H), 1.44 (p, J = 7.3 Hz, 2H), 1.37 (s, 9H), 1.28 (tt, J = 10.5, 6.3 Hz, 2H), 1.20 (t, J = 7.1 Hz, 6H).

2-(4-((tert-butoxycarbonyl)amino)butyl)malonic acid (16) Diethyl 2-(4-((tert-butoxycarbonyl)amino)butyl)malonate (72 mg, 0.22 mmol) was dissolved in 1 mL of ethanol then added dropwise to 375 µL of 6.67 M NaOH. The mixture was stirred for 5 h at 60 °C. The solution was then cooled to 0 °C, carefully acidified to pH 1 with 1 N HCl, and extracted with 5 portions of Et_2O. The combined extracts were washed with a saturated solution of NaCl, dried over Na_2SO_4, and evaporated to dryness. 2-(4-((tert-butoxycarbonyl)amino)butyl)malonic acid was dissolved in 1:1 H_2O:MeCN and lyophilized to give a clear serum. Yield >99%. ^{1}H NMR (500 MHz, CDCl3) δ 6.29 (s, 1H), 4.57 (s, 1H), 3.35 (t, J = 7.0 Hz, 1H), 3.03 (t, J = 6.7 Hz, 2H), 1.92 – 1.85 (m, 2H), 1.43 (p, J = 6.7 Hz, 2H), 1.36 (s, 11H). HR-ESI-MS [M-H]-: calculated for C12H20O6N1 m/z 274.1296, found m/z 274.1295.

(S)-6-(Boc-amino)-2-hydroxyhexanoic Acid (24) was synthesized by following a published method (*81*).

VI. Structure determination

Cloning. The following synthetic dsDNA sequence was cloned upstream of *Ma*FRSA (*Ma*PylRS N166A:V168A) into pET32a-*Ma*FRSA by Gibson assembly (*71*) and used for subsequent crystallographic studies: GSS linker - 6xHis - SSG linker - thrombin site - *Ma*FRSA (Table S1). The sequence of the pET32a-6xHis-thrombin-MaFRSA plasmid was confirmed with Sanger sequencing from Genewiz using primers T7 F and T7 R (Table S1).

Protein expression and purification for crystallography. The procedure used to express and purify *Ma*FRSA for crystallography using pET32a-6xHis-thrombin-*Ma*FRSA was adapted from a reported protocol used to express and purify wild-type *M. alvus* PylRS by Seki *et al.* (*55*). BL21(DE3) Gold competent cells (Agilent Technologies) were transformed with pET32a-6xHis-thrombin-*Ma*FRSA and grown in TB media at 37 °C. Protein expression was induced at an OD_{600} reading of 1.2 with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The temperature was lowered to 20 °C and growth continued overnight. Cells were pelleted for 1 h at 4,300 × g and resuspended in Lysis buffer (50 mM potassium phosphate (pH 7.4), 25 mM imidazole, 500 mM sodium chloride, 5 mM β-mercaptoethanol, 1 cOmplete Mini EDTA-free protease inhibitor tablet). Cells were lysed by homogenization (Avestin Emulsiflex C3). After centrifugation for 1 h at 10,000 × g, the clarified lysate was bound to TALON® Metal Affinity Resin (Takara Bio) for 1 h at 4 ºC, washed with additional lysis buffer, and eluted with Elution Buffer (50 mM potassium phosphate (pH 7.4), 500 mM imidazole, 500 mM sodium chloride, 5 mM β-mercaptoethanol). The eluate was dialyzed overnight at 4 °C into Cleavage buffer (40 mM potassium phosphate (pH 7.4), 100 mM NaCl, 1 mM dithiothreitol (DTT)) then incubated overnight at room temperature with thrombin protease on a solid agarose support (MilliporeSigma). Following cleavage, the protein was passed over additional TALON® resin to remove the 6xHis tag and dialyzed overnight at 4 °C into Sizing buffer (30 mM potassium phosphate (pH 7.4), 200 mM NaCl, 1 mM DTT). The protein was concentrated and loaded onto a HiLoad® 16/600 Superdex® 200 pg column (Cytiva Life Sciences) equilibrated with Sizing buffer on an ÄKTA Pure 25 fast-liquid chromatography machine. Purified *Ma*FRSA was dialyzed into Storage buffer (10 mM Tris-HCl (pH 8.0), 150 mM NaCl, 10 mM MgCl2, 10 mM β-mercaptoethanol), concentrated to 20 mg/mL, aliquoted, and flash-frozen for crystallography (fig. S2, F to G).

Co-crystallization of *Ma*FRSA with substrates. Initial crystallization screening conditions were adapted from Seki *et al.* (*55*). Crystals were grown by hanging drop vapor-diffusion in 24-well plates. 25 µL of 100 mM *meta*-trifluoromethyl-2-benzylmalonate (*meta*-CF_3-2-BMA) pH ~7 was added to 1.5 mL micrcentrifuge tubes and the water was removed by evaporation. The dried aliquots were then resuspended at a concentration of 100 mM with *Ma*FRSA in Storage buffer at three concentrations (6.9, 12.3, and 19.2 mg/mL) and 10 mM adenosine 5′-(β,γ-

imido)triphosphate lithium salt hydrate (AMP-PNP). The protein/substrate solution (1 µL) was mixed in a 1:1 ratio with the reservoir solution (1 µL) containing 10 mM Tris-HCl pH 7.4 and 26% polyethylene glycol 3350 and incubated over 1 mL of reservoir solution at 18 °C. Crystals with an octahedral shape appeared within one week. Crystals were plunged into liquid nitrogen to freeze with no cryoprotectant used.

X-ray data collection, structure solution, and refinement. Data were collected at the Advanced Light Source beamline 8.3.1 at 100 K with a wavelength of 1.11583 Å. Data collection and refinement statistics are presented in Table S4. Diffraction data were indexed and integrated with XDS (*82*), then merged and scaled with Pointless (*83*) and Aimless (*84*). The crystals were in the space group I4, and the unit cell dimensions were 108.958, 108.958, and 112.26 Å. The structure was solved by molecular replacement with Phaser (*85*) using a single chain of the wild-type apo structure of *M. alvus* PylRS (PDB code: 6JP2) (*55*) as the search model. There were two copies of *Ma*FRSA in the asymmetric unit. The model was improved with iterative cycles of manual model building in COOT (*86*) alternating with refinement in Phenix (*87*, *88*) using data up to 1.8 Å resolution. Structural analysis and figures were generated using Pymol version 2.4.2 (*89*).

VII. *In Vitro* Translation Initiation

tRNA preparation. The *Ma*-tRNA^{Pyl}-ACC dsDNA template was prepared as described in **II** using the primers *Ma*-PylT-ACC F and *Ma*-PylT-ACC R (Table S1). *Ma*-tRNA^{Pyl}-ACC was also transcribed, purified, and analyzed as described in **II**.

tRNA acylation and work-up. Enzymatic tRNA acylation reactions (150 µL) were performed as described in **III** with slight modifications. The enzyme concentration was increased to 12.5 µM (monomers **7**, **14**, and **15**) or 25 µM (monomer **13**) and the incubation time was increased to 3 hours at 37 °C. Sodium acetate (pH 5.2) was added to the acylation reactions to a final concentration of 300 mM in a volume of 200 µL. The reactions were then extracted once with a 1:1 (v/v) mixture of acidic phenol (pH 4.5):chloroform and washed twice with chloroform. After extraction, the acylated tRNA was precipitated by adding ethanol to a final concentration of 71% and incubation at -80 °C for 30 min followed by centrifugation at 21,300 × g for 30 min at 4 ºC.

Acylated tRNAs were resuspended in water to a concentration of 307 μM immediately before *in vitro* translation.

Synthesis and purification of DNA templates for translation. Templates for expression of MGVDYKDDDDK were prepared by annealing and extending the oligonucleotides MGVflag-1 and MGVflag-2 using Q5® High-Fidelity 2X Master Mix (NEB) (Table S1). The annealing and extension used the following protocol on a thermocycler (BioRad C1000 Touch™): 98 ºC for 30 s, 10 cycles of [98 ºC for 10 s, 55 ºC for 30 s, 72 ºC for 45 s], 10 cycles of [98 ºC for 10 s, 67 ºC for 30 s, 72 ºC for 45 s], and 72 º C for 300 s. Following extension, the reaction mixture was supplemented with sodium acetate (pH 5.2) to a final concentration of 300 mM, extracted once with a 1:1 (*v*/*v*) mixture of basic phenol (pH 8.0):chloroform, and washed twice with chloroform. The dsDNA product was precipitated upon addition of ethanol to a final concentration of 71% and incubation at -80 °C for 30 min followed by centrifugation at 21,300 × g for 30 min at 4 ºC. The dsDNA pellets were washed once with 70% (*v/v*) ethanol and resuspended in 10 mM Tris-HCl pH 8.0 to a concentration of 500 ng/μL and stored at -20 °C until use in translation.

***In vitro* translation reactions.** *In vitro* transcription/translation by codon skipping of the short FLAG tag-containing peptides X-Val-Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys (XV-Flag) where X = **7**, **13**, **14**, or **15** was carried out using the PURExpress® Δ (aa, tRNA) Kit (New England Biolabs, E6840S) based on a previous protocol with slight modifications (*16*). The XV-Flag peptides were produced with the following reactions (12.5 μL): Solution A (ΔtRNA, Δaa; 2.5 μL), amino acid stock mix (1.25 μL; 33 mM L-valine, 33 mM L-aspartic acid, 33 mM L-tyrosine, 33 mM L-lysine), tRNA solution (1.25 μL), Solution B (3.75 μL), 250 ng dsDNA MGVDYKDDDDK template (0.5 μL), and *Ma*-tRNA^{Pyl}-ACC acylated with **7**, **13**, **14**, or **15** (3.25 μL). The reactions were incubated in a thermocycler (BioRad C1000 Touch™) at 37 °C for 2 hours and quenched by placement on ice.

Anti-FLAG enrichment. Translated peptides were purified from *in vitro* translation reactions by enrichment using Anti-FLAG® M2 Magnetic Beads (Millipore Sigma) according to the manufacturer's protocol with slight modifications. For each peptide, 10 μL of a 50% (*v/v*) suspension of magnetic beads was used. The supernatant was pipetted from the beads on a magnetic manifold. The beads were then washed twice by incubating with 100 μL of TBS (150

mM NaCl, 50 mM Tris-HCl, pH 7.6) for 10 min at room temperature then removing the supernatant each time with a magnetic manifold. The *in vitro* translation reactions were added to the beads and incubated at RT for 30 min with periodic agitation. The beads were washed again three times with 100 µL of TBS as described above. Peptides were eluted by incubation with 12.5 µL of 0.1 M glycine-HCl pH 2.8 for 10 minutes. The supernatant was transferred to vials and kept on ice for analysis.

Analysis *via* mass spectrometry. The purified peptides were analyzed based on a previous protocol (*16*). The supernatant was analyzed on an ZORBAX Eclipse XDB-C18 column (1.8 µm, 2.1 × 50 mm, room temperature, Agilent) using an 1290 Infinity II UHPLC (G7120AR, Agilent). The following method was used for separation: an initial hold at 95% Solvent A (0.1% formic acid in water) and 5% Solvent B (acetonitrile) for 0.5 min followed by a linear gradient from 5 to 50% Solvent B over 4.5 min at flow rate of 0.7 mL/min. Peptides were identified using LC-HRMS with an Agilent 6530 Q-TOF AJS-ESI (G6230BAR). The following parameters were used: a fragmentor voltage of 175 V, gas temperature of 300 °C, gas flow rate of 12 L/min, sheath gas temperature of 350 °C, sheath gas flow rate of 11 L/min, nebulizer pressure of 35 psi, skimmer voltage of 75 V, Vcap of 3500 V, and collection rate of 3 spectra/s. Expected exact masses of the major charge state for each peptide were calculated using ChemDraw 19.0 and extracted from the total ion chromatograms ± 100 ppm.

Results

*Ma*PylRS retains much of the promiscuity of *Mm*PylRS

First we set out to establish whether the novel substrate scope of *Mm*PylRS reported for L-BocK analogs (Fig. 1B) (*20*) was retained by *Ma*PylRS, which offers advantages over *Mm*PylRS because it lacks the poorly soluble N-terminal tRNA-binding domain (*39*) and is easier to express and evaluate *in vitro* (*36*). The C-terminal catalytic domain of *Mm*PylRS (*34*) is 36% identical to *Ma*PylRS and the structures are largely superimposable (fig. S1, A to B). To evaluate whether D-BocK as well as L-BocK analogs with -OH or -H in place of the α-amino group were substrates for *Ma*PylRS, we made use of a validated RNAse A/LC-HRMS assay (*16*, *40*). This assay exploits RNAse A to cleave the phosphodiester bond of unpaired C and U residues to generate 2', 3'-cyclic phosphate products (*41*). As a result, the residue at the tRNA 3' terminus is

the only mononucleoside product lacking a phosphate (Fig. 1C). Incubation of L-BocK **1** (2 mM) with purified *Ma*PylRS (2.5 μM) and *Ma*-tRNA^{Pyl} (25 μM) (fig. S2) at 37 ºC for 2 hours led to a pair of RNAse A digestion products whose expected mass (496.25142 Da) corresponds to the adenosine nucleoside **6** as a mixture of 2'- and 3'-acylated species (Fig. 1D) (*42*). No products with this mass were observed when the reaction mixture lacked *Ma*-tRNA^{Pyl}, L-BocK, or *Ma*PylRS, and mass analysis of the intact tRNA product confirmed a 53% yield of acylated tRNA (**1-tRNA^{Pyl}**, fig. S3A). Under these conditions, L-BocK analogs with either -OH (**2**) or -H (**3**) in place of the α-amino group were also substrates for *Ma*PylRS as judged by RNAse A (Fig. 1D) and intact tRNA mass spectrometry assays with acylated tRNA yields of 89% (**2-tRNA^{Pyl}**) and 61% (**3-tRNA^{Pyl}**) (fig. S3, B to C, respectively). No reactivity was detected with D-BocK (figs. S3D and S7). We conclude that *Ma*PylRS retains much (although not all) of the previously reported (*20*) promiscuity of *Mm*PylRS. We note that certain non-natural monomers with relatively high activity, including α-hydroxy **2** and des-amino **3**, led to measurable levels (2-13%) of diacylated tRNA (Table S3). Diacylated tRNAs have been observed as products in cognate reactions of *T. thermophilus* PheRS (*43*) and are active in prokaryotic translation (*44*).

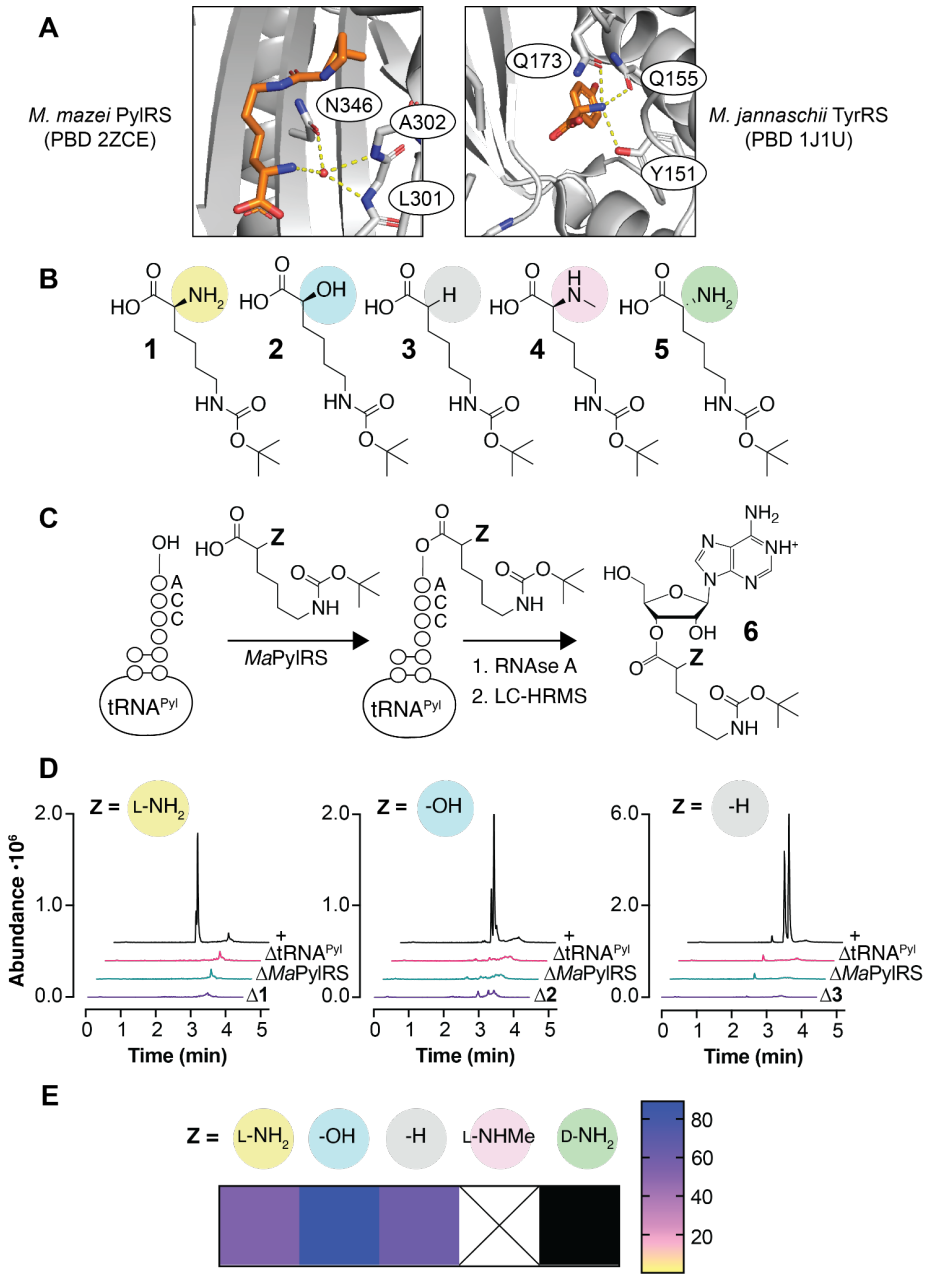


Fig. 1. **A.** The α-amines of L-α-amino acids are recognized differently by *M. mazei* PylRS (*Mm*PylRS, left) and *M. jannaschii* TyrRS (*Mj*TyrRS, right). **B.** N^{ε}-Boc-L-lysine (L-BocK, **1**) analogs evaluated as substrates for *Ma*PylRS. **C.** Ribonuclease A (RNAse A) assay used to detect acylation of *Ma*-tRNA^{Pyl} with BocK analogs shown in panel B. **D.** LC-HRMS analysis of *Ma*-tRNA^{Pyl} acylation reactions after RNAse A digestion. Peak masses correspond to adenosine nucleoside **6** acylated on the 2′ or 3′ hydroxyl with the indicated monomer. **E.** Heat map illustrating the relative activities of substrates **1-5** for *Ma*PylRS as determined by intact tRNA analysis (fig. S3) as described in **Supplementary Methods**. Reported yields are based on intact

tRNA analysis. Black indicates no reaction product detected; X indicates the substrate was not investigated.

_Ma_PylRS variants retain activity for phenylalanine derivatives with diverse α-amine substitutions

PylRS is a subclass IIc aaRS that evolved from PheRS (*45*). *Mm*PylRS variants with substitutions at two positions that alter the architecture of the side chain-binding pocket (N346, C348) (fig. S1A) accept L-phenylalanine and its derivatives in place of pyrrolysine (*37*, *45*, *26*). We integrated the mutations in two variants that accept unsubstituted L-Phe (*Mm*FRS1 and *Mm*FRS2) (*37*) into the *Ma*PylRS sequence to generate *Ma*FRS1 (N166A, V168L) and *Ma*FRS2 (N166A, V168K) (fig. S2). We then used RNAse A and intact tRNA mass spectrometry assays to determine if *Ma*FRS1 or *Ma*FRS2 retained activity for L-phenylalanine **7** and analogs in which the L-α-amino group was substituted by -OH **8**, -H **9**, -$NHCH_3$ **10,** or D-NH_2 **11** (Fig. 2A).

Incubation of L-Phe **7** (10 mM) with purified *Ma*FRS1 or *Ma*FRS2 (2.5 μM) and *Ma*-$tRNA^{Pyl}$ (25 μM) at 37 ºC for 2 hours led in both cases to a pair of products whose expected mass (415.17244 Da) corresponds to adenosine nucleoside **12** (Z = L-NH_2, Fig. 2B) as the expected mixture of 2'- and 3'-acylated products (Fig. 2C). No product with this mass was observed when the reaction mixture lacked *Ma*-$tRNA^{Pyl}$, L-Phe, or *Ma*FRS1 or *Ma*FRS2, and mass analysis of the intact tRNA product confirmed a 66% (*Ma*FRS1) and 65% (*Ma*FRS2) yield of acylated tRNA (**7-$tRNA^{Pyl}$**) (fig. S4A and S5A, respectively). L-Phe analogs **8-10** were all substrates for both *Ma*FRS1 and *Ma*FRS2 as judged by both RNAse A (Fig. 2C) and intact tRNA analysis (fig. S4, A to D and S5, A to D)**,** with reactivities in the order L-α-amino **7** > α-hydroxy **8** >> des-amino **9** ~ *N*-Me-L-α-amino **10** based on intact tRNA yields (Fig. 2D). Mono- and di-acylated tRNA products were observed for substrates **7** and **8** (Table S3). Interestingly, despite the fact that the des-amino-BocK analog **3** was a strong substrate for the wild-type *Ma*PylRS, the des-amino-Phe analog **9** had only modest activity with *Ma*FRS1 and *Ma*FRS2, as observed in the RNAse assay (Fig. 2C, figs. S4C and S5C). Again, no reactivity was detected with D-Phe (figs. S4E, S5E and S7).

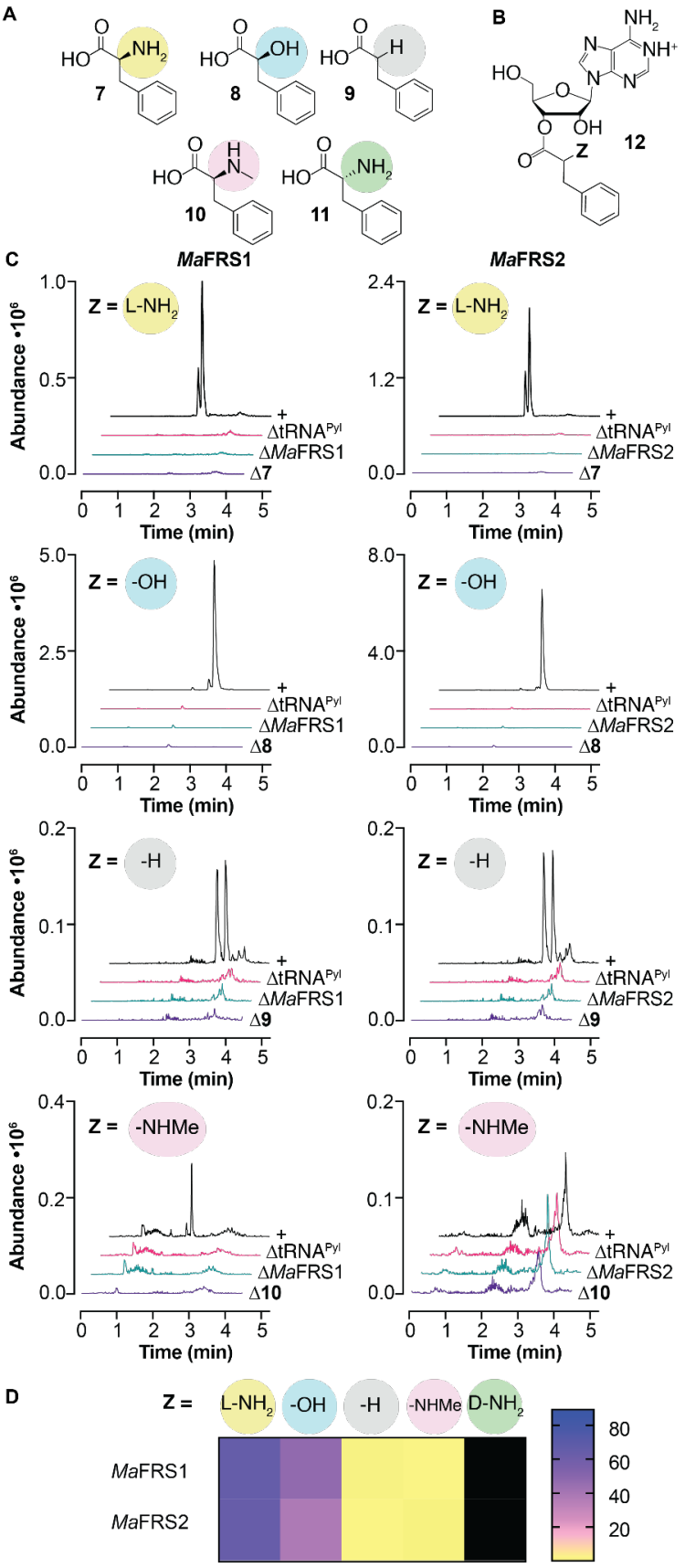

A
7
8
9
10
11
B
12
C
MaFRS1
MaFRS2
Z = L-NH2
Z = -OH
Z = -H
Z = -NHMe
+
ΔtRNAPyl
ΔMaFRS1
ΔMaFRS2
Δ7
Δ8
Δ9
Δ10
Abundance •10^6
Time (min)
D
Z =
L-NH2
-OH
-H
-NHMe
D-NH2
MaFRS1
MaFRS2
80
60
40
20

Fig. 2: **A.** Phenylalanine analogs evaluated as substrates for *Ma*FRS1 and *Ma*FRS2. **B.** Adenosine nucleoside formed during RNAse A digestion of acyl-tRNA. **C.** LC-HRMS analysis of *Ma*-tRNA^{Pyl} acylation reactions after RNAse A digestion. Adenine nucleoside **12** acylated on the 2′- or 3′-hydroxyl of the 3′ terminal ribose of *Ma*-tRNA^{Pyl} could be detected in *Ma*FRS1 and *Ma*FRS2 reactions with L-Phe **7** and substrate **8**; substrates **9** and **10** (Z = -H, -$NHCH_3$) showed more modest reactivity. **D.** Heat map illustrating the relative yields with substrates **7**-**11** for *Ma*FRS1 and *Ma*FRS2 as determined by intact tRNA analysis (fig. S4 and S5) as described in **Supplementary Methods**. Black indicates no reaction product detected.

_Ma_FRS1 & _Ma_FRS2 process substrates with novel α-substituents

We then began to explore phenylalanine analogs with larger and electrostatically distinct functional groups at the α-carbon: α-thio acids, *N*-formyl-L-α-amino acids, and α-carboxyl acids (malonates) (Fig. 3A). α-thio acids are substrates for extant *E. coli* ribosomes in analytical-scale *in vitro* translation reactions with yields as high as 87% of the corresponding α-amino acids (*14*), and thioesters can have a half-life in *E. coli* of more than 36 hours (*46*). Peptides and proteins containing thioesters could also act as substrates for PKS modules (*47*) to generate unique keto-peptide natural products. Formylation of methionyl-tRNA is important for initiation complex formation, and formylation could enhance initiation with non-methionyl α-amino acids *in vivo* (*48*, *49*). Moreover, *E. coli* ribosomes incorporate monomers containing a 1,3-dicarbonyl moiety at the peptide N-terminus to produce keto-peptide hybrids (*16*). To our knowledge, there are currently no aaRS enzymes, orthogonal or otherwise, that accept α-thio, *N*-formyl-L-α-amino, or α-carboxyl acid substrates to generate the acylated tRNAs required for *in vivo* translation (when extant ribosomes are compatible) or ribosome evolution (when extant ribosomes are incompatible).

We found that L-Phe analogs in which the α-amine was substituted with α-thiol (**13)**, α-carboxyl (**14**), or *N*-formyl-L-α-amine (**15**) moieties were all substrates for *Ma*FRS1 and *Ma*FRS2 (Fig. 3, A to C). In particular, α-carboxyl **14** and *N*-formyl-L-α-amine **15** were excellent substrates. Incubation of α-carboxyl acid **14** (10 mM) with *Ma*FRS1 or *Ma*FRS2 (2.5 μM) and *Ma*-tRNA^{Pyl} (25 μM) at 37 ºC for 2 hours followed by digestion by RNAse A led to formation of a pair of products whose expected mass (444.15137 Da) corresponds to the adenosine nucleoside **12** (Z = -COOH). LC-MS analysis of intact tRNA products confirmed a 24% (*Ma*FRS1) and 43%

(*Ma*FRS2) yield of acylated tRNA (**14-tRNA^{Pyl}**) (figs. S4H and S5H, respectively). Because the α-carbon of substrate **14** is prochiral, mono-acylation of *Ma*-tRNA^{Pyl} can generate two diastereomeric product pairs–one pair in which the 3'-hydroxyl group is acylated by the *pro*-S or *pro*-R carboxylate and another in which the 2'-hydroxyl group is acylated by the *pro*-S or *pro*-R carboxylate. These diastereomeric products result from alternative substrate orientations within the enzyme active site (*vide infra*).

Incubation of *N*-formyl-L-α-amine (**15**) under identical conditions led to the expected adenosine nucleoside **12** (Z = -NHCHO) and LC-MS analysis confirmed a 49% (*Ma*FRS1) and 31% (*Ma*FRS2) yield of acylated tRNA (**15-tRNA^{Pyl}**) (figs. S4I and S5H, respectively). The α-thio L-Phe analog **13** was also a substrate for *Ma*FRS1 and *Ma*FRS2, though higher concentrations of *Ma*FRS1 were necessary to observe the acyl-tRNA product (**13-tRNA^{Pyl}**) using intact tRNA LC-MS (figs. S4, F to G, and S5F). These results illustrate that the active site pocket that engages the α-amine in PylRS can accommodate substituents with significant differences in mass (-NHCHO) and charge (-COO$^-$). Despite these differences, 2-benzylmalonate **14** (2-BMA) is an excellent substrate for *Ma*FRS1 and *Ma*FRS2; kinetic analysis using the malachite green assay (*50*) revealed a rate of adenylation by *Ma*FRS1 that was 66% of the rate observed for L-Phe (Fig. 3D). The tolerance for malonate substrates extends to WT *Ma*PylRS itself: the α-carboxylate analog of L-BocK **16** was also a measurable substrate for WT *Ma*PylRS (figs. S3E and S7).

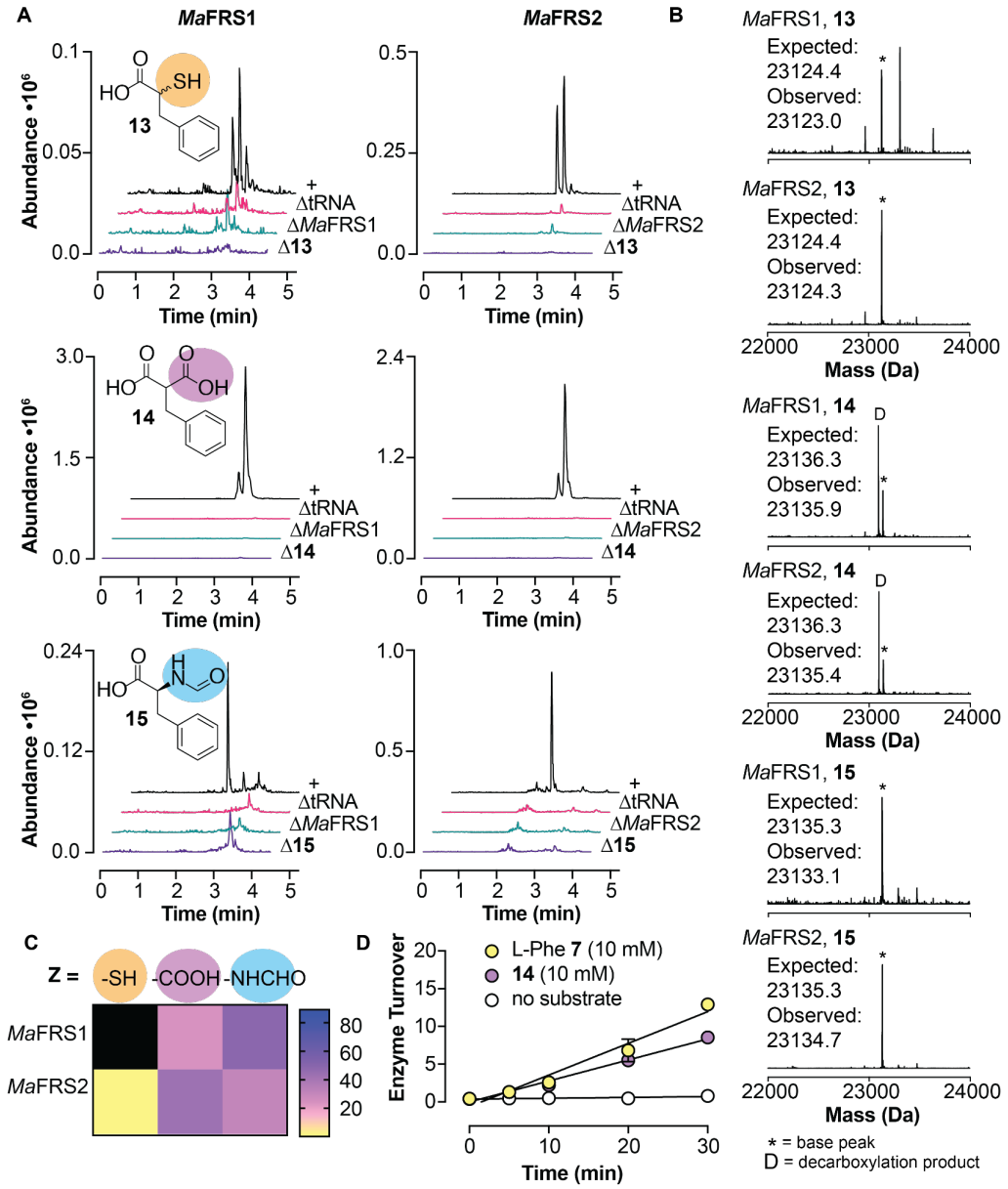


Fig. 3: **A.** LC-HRMS analysis of *Ma*-tRNA^{Pyl} acylation reactions using *Ma*FRS1 or *Ma*FRS2 following RNAse A digestion. Adenine nucleoside **12** acylated on the 2′- or 3′- hydroxyl of the 3′ terminal ribose of *Ma*-tRNA^{Pyl} could be detected in *Ma*FRS1 and *Ma*FRS2 reactions with α-thio acid **13**, α-carboxyl acid **14**, and *N*-formyl-L-Phe **15**. **B.** LC-MS analysis of intact tRNA products confirms that monomers **13-15** are substrates for *Ma*FRS1 and *Ma*FRS2. We note that intact tRNAs acylated with 2-benzylmalonate **14** showed evidence of decarboxylation (indicated by a D). No evidence for decarboxylation was observed when the same acyl-tRNAs were evaluated using the RNAse A assay (fig. S8), suggesting that decarboxylation occurs either during workup or during the LC-MS itself. **C.** Heat map illustrating the relative activities of

substrates **13**-**15** with *Ma*FRS1 and *Ma*FRS2 as determined by intact tRNA analysis (figs. S4 and S5) as described in **Supplementary Methods**. Black indicates no reaction product detected. Initially no acyl-tRNA was detected when *Ma*FRS1 was incubated with **13**, but when the enzyme concentration was increased 5-fold, acyl-tRNA was detected with mono- and diacyl yields of 0.7 and 9.7%, respectively. **D.** Turnover of *Ma*FRS1 over time with L-Phe **7** and 2-benzylmalonate **14** using the malachite green assay.

_Ma_FRSA processes _meta_-substituted 2-benzylmalonic acid substrates and is orthogonal to L-Phe

While *Ma*FRS1 and *Ma*FRS2 demonstrated the ability to process substrates with unusual α-substituents, they also process L-Phe with comparable efficiency (Fig. 2D), which would interfere with the selective charging of the non-L-α-amino acid. Variants of *Mm*PylRS that accept *para*-, *ortho-,* and *meta*-substituted L-Phe derivatives have been reported (*38*, *51–54*). In particular, *Mm*PylRS containing two active site mutations (N346A and C348A; henceforth referred to as FRSA) shows high activity for L-Phe analogs with bulky alkyl substituents and low activity towards L-Phe (*38*). We expressed and purified a variant of *Ma*PylRS containing these mutations (*Ma*FRSA: N166A, V168A) (fig. S2) and demonstrated that it shows high activity for derivatives of malonate **14** carrying *meta*-CH_3 (**17**), *meta*-CF_3 (**18**), and *meta*-Br (**19**) substituents and low activity for L-Phe using both RNAse A (Fig. 4A) and intact tRNA analysis (fig. S6, A to D). Of the *meta*-substituted 2-benzylmalonates, *Ma*FRSA shows the highest activity for *meta*-CF_3-2-benzylmalonate **18** (*meta*-CF_3-2-BMA). Kinetic analysis (*50*) revealed a rate of adenylation that was 36% of the rate observed for the L-α-amino acid counterpart *meta*-CF_3-L-Phe (Fig. 4D). Although derivatives of malonate **14** carrying *meta*-CH_3 (**17**), *meta*-CF_3 (**18**), and *meta*-Br (**19**) substituents are also excellent substrates for *Ma*FRS1 (fig. S4, J to L) and *Ma*FRS2 (fig. S5, I to K), *Ma*FRSA has significantly lower activity for L-Phe, allowing for the selective acylation of tRNA with *meta*-substituted 2-benzylmalonates without interference from L-Phe (Fig. 4C). Indeed, when *Ma*FRSA is incubated with an equal concentration (10 mM) of L-Phe **7** and *meta*-CF_3-2-BMA **18**, only the malonyl product (**18-tRNA^{Pyl}**) is observed (fig. S6E).

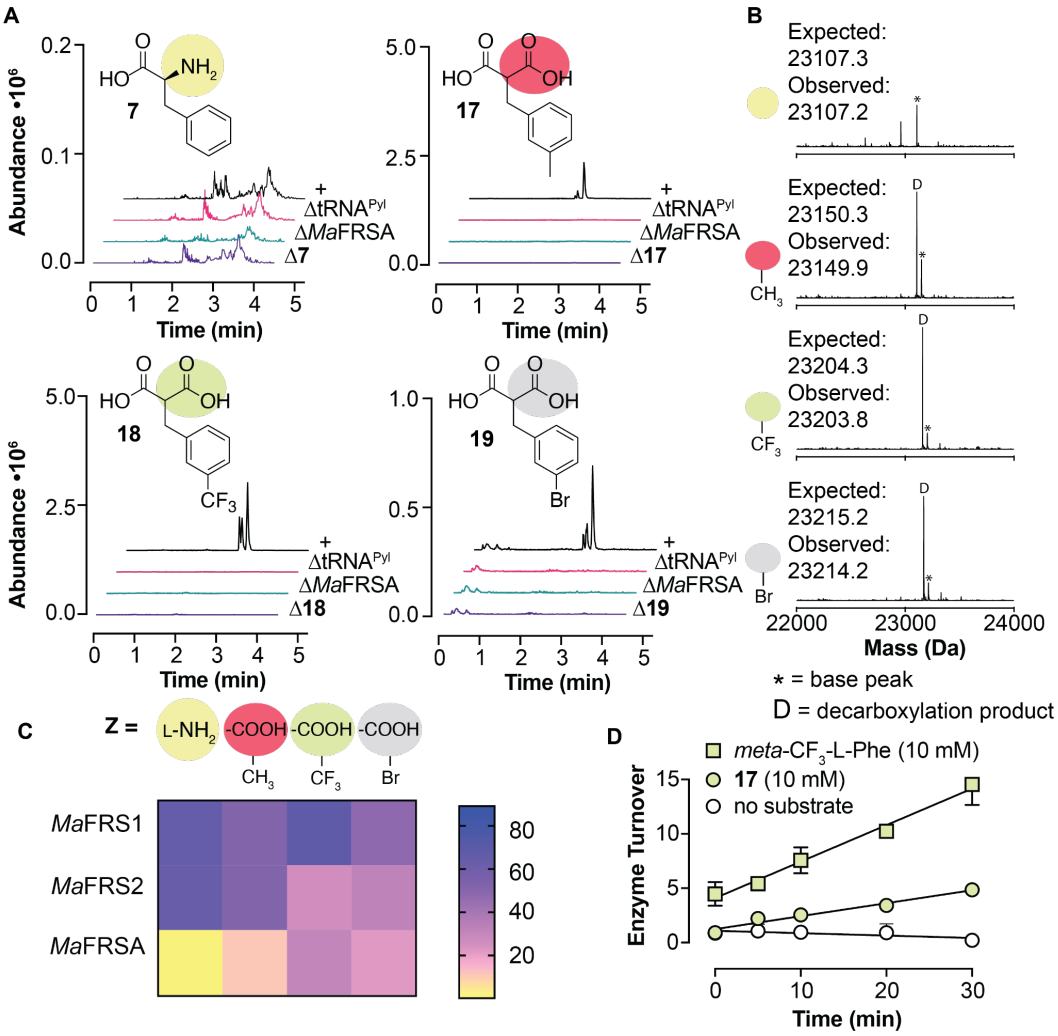


Figure 4. *Ma*FRSA selectively acylates *Ma*-tRNA^{Pyl} with *meta*-substituted 2-benzylmalonate derivatives. **A.** LC-HRMS analysis of *Ma*-tRNA^{Pyl} acylation products after digestion with RNAse A. **B.** LC-MS analysis of intact tRNA products confirms that *meta*-substituted 2-benzylmalonates **17-19** are substrates for *Ma*FRSA. We note that intact tRNAs acylated with *meta*-substituted 2-benzylmalonates **17-19** showed evidence of decarboxylation (indicated by a D). No evidence for decarboxylation was observed when the same acyl-tRNAs were evaluated using the RNAse A assay (fig. S8), suggesting that decarboxylation occurs either during workup or during the LC-MS itself. **C.** Heat map illustrating the relative activities of L-Phe **7** and substrates **17-19** with *Ma*FRS1, *Ma*FRS2, and *Ma*FRSA. Black indicates no reaction product detected. **D.** Turnover of *Ma*FRSA over time with *meta*-CF_3-L-Phe and *meta*-CF_3-2-BMA **18** using the malachite green assay.

Structural analysis of *Ma*FRSA-*meta*-CF_3-2-benzylmalonate complex reveals novel interactions

To better understand how 2-benzylmalonate substrates are accommodated by the *Ma*FRSA active site, we solved the crystal structure of *Ma*FRSA bound to both *meta*-CF_3-2-BMA and the non-hydrolyzable ATP analog adenosine 5′-(β,γ-imido)triphosphate (AMP-PNP). The structure was refined at 1.8 Å with clear substrate density for *meta*-CF_3-2-BMA visible at 1σ in the $2F_o$-F_c map (fig. S9, A to D). *Ma*FRSA crystallized with two protein chains in the asymmetric unit and an overall architecture resembling published PylRS structures (Fig. 5A) (*34*, *55*, *56*). The two protein chains in the asymmetric unit are not identical and interact with different orientations of *meta*-CF_3-2-BMA (Fig. 5B). One orientation of *meta*-CF_3-2-BMA (chain A, light purple) mimics that of L-pyrrolysine (Pyl) bound to *Mm*PylRS (*34*) and would result in adenylation of the *pro*-R carboxylate (Fig. 5C); the other orientation (chain B, dark purple) would result in adenylation of the *pro*-S carboxylate (Fig. 5D).

Regardless of orientation, discrete networks of hydrogen bonds are used by *Ma*FRSA to discriminate between the *pro*-R and *pro*-S carboxylates of *meta*-CF_3-2-BMA. In chain A, the *pro*-S carboxylate accepts a hydrogen bond from the backbone amides of L121 and A122 and the phenolic -OH of Y206 (Fig. 5C and fig. S9E). A hydrogen bond from the R150 guanidinium positions the *pro*-R carboxylate for nucleophilic attack with a distance of 2.7 Å between the carboxyl oxygen and the α-phosphorous of AMP-PNP. In chain B, the orientation of *meta*-CF_3-2-BMA is flipped relative to the conformation bound to chain A (Fig. 5D and fig. S9F). The *pro*-R carboxylate accepts a hydrogen bond from the backbone amides of L121 and A122 and the phenolic -OH of Y206 as seen for the *pro*-S carboxylate in chain A. However, in chain B the *pro*-S carboxylate is rotated away from AMP-PNP and towards Y206 resulting in loss of the hydrogen bond to R150 and a longer distance of 3.9 Å between the carboxyl oxygen and the α-phosphorous of AMP-PNP. RNAse A analysis of *Ma*-$tRNA^{Pyl}$ acylation by *meta*-CF_3-2-BMA shows more than two peaks of identical mass (Fig. 4A) that likely correspond to the two diastereomeric pairs formed from attack of the 2'- or 3'- tRNA hydroxyl group on the activated *pro*-R or *pro*-S carboxylate. More than two peaks with identical mass are also observed as RNAse A digestion products in *Ma*-$tRNA^{Pyl}$ acylation reactions of other *meta*-substituted 2-benzylmalonates (Fig. 4A). While the *meta*-CF_3-2-BMA conformation in chain A appears more favorably positioned for catalysis, suggesting that the *pro*-R carboxylate is acylated

preferentially, the appearance of more than two peaks in the RNAse A assay suggests that both conformations are catalytically competent.

As anticipated, the non-reactive, *pro*-S carboxylate of *meta*-CF_3-2-BMA is recognized by *Ma*FRSA chain A using interactions that are distinct from those used by *Mm*PylRS to recognize the Pyl α-amine. In the *Mm*PylRS:Pyl:AMP-PNP complex (PDB: 2ZCE) (*34*), the Pyl α-amine is recognized by water-mediated hydrogen bonds to the backbone amides of L301 and A302 and the side chain carbonyl of N346, rather than by direct hydrogen bonds to the backbone amides of L121 and A122 as seen for recognition of the non-reacting carboxylate of *meta*-CF_3-2-BMA by both chains of *Ma*FRSA (Fig. 5E). The interactions used to recognize the non-reacting carboxylate of *meta*-CF_3-2-BMA are, however, similar to those used to recognize the single carboxylate of diverse α-amino acids by other PylRS variants with mutations at N166/N346 (*Ma*/*Mm* numbering). For example, the structures of *Mm*PylRS variants with mutations at N346, such as *Mm*IFRS and *Mm*BtaRS bound to 3-iodo-L-phenylalanine (3-I-F, PDB: 4TQD) (*53*) and 3-benzothienyl-L-alanine (Bta, PDB: 4ZIB) (*57*), respectively, show the substrate bound with the carboxylate directly hydrogen bonded to the L301 and A302 backbone amides, just as seen for *meta*-CF_3-2-BMA bound to *Ma*FRSA (Fig. 5F and fig. S10A). In these cases, the bound water seen in the *Mm*PylRS:Pyl:AMP-PNP complex is either absent or displaced. Mutation of N166/N346 may destabilize the water-mediated hydrogen bonding between the substrate α-amine and backbone amides seen in wild-type PylRS and promote alternative direct hydrogen bonding of a substrate carboxylate to backbone amides as seen in *Ma*FRSA, *Mm*IFRS, and *Mm*BtaRS.

The largest differences between *Ma*FRSA and other reported PylRS structures are localized to a 6-residue loop that straddles β-strands β7 and β8 and contains the active site residue Y206. In the the *Mm*PylRS:Pyl:AMP-PNP structure (PDB: 2ZCE) (*34*) and the wild-type *Ma*PylRS apo structure (PDB: 6JP2) (*55*), the β7-β8 loop is either unstructured or in an open conformation positioning Y206/Y384 away from the active site, respectively (fig. S10B). Among wild-type PylRS structures, the β7-β8 loop exists in the closed conformation only in the structure of *Mm*PylRS bound to the reaction product, Pyl-adenylate (PDB: 2Q7H) (*56*). In this structure, Y384 accepts and donates a hydrogen bond to the Pyl-adenylate α-amine and pyrrole nitrogen, respectively, and forms a hydrophobic lid on the active site. In both chains of substrate-bound *Ma*FRSA, the non-reacting carboxylate of *meta*-CF_3-2-BMA forms similar hydrogen bonds to

Y206. These hydrogen bonds effectively close the β7-β8 loop to form a hydrophobic lid on the active site, which may contribute to the high acylation activity observed for *meta*-CF_3-2-BMA with *Ma*FRSA. We note that the β7-β8 loop in chain A exhibits lower B-factors than in chain B indicating tighter binding and providing further evidence that chain A represents the more active binding mode of *meta*-CF_3-2-BMA (fig. S10, C and D). The two alternative poses of *meta*-CF_3-2-BMA in chains A and B correspond to a ~120° degree rotation about the Cα-Cβ bond of the substrate that largely maintains interactions with the *meta*-CF_3-BMA side chain but alters the placement of the reacting carboxylate. This observation emphasizes the dominant stabilizing role of the main chain H-bonds provided by the backbone amides of L121 and A122.

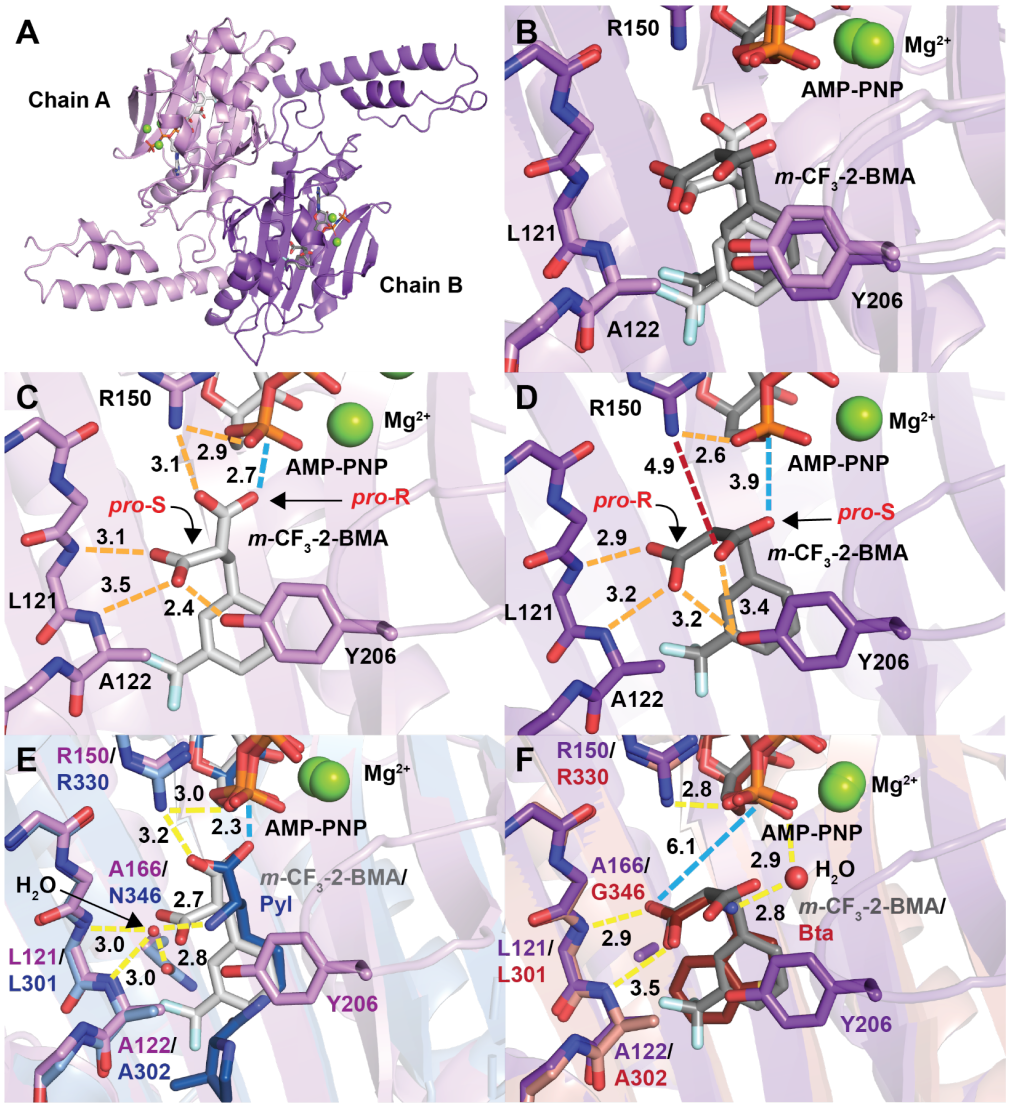


Figure 5. Structure of *Ma*FRSA bound to *meta*-CF_3-2-BMA and AMP-PNP reveals basis for distinct reactivity at *pro*-R and *pro*-S substrate carboxylates. **A.** *Ma*FRSA dimer containing two non-identical chains in the asymmetric unit. **B.** Alignment of the active sites of chains A (light

purple) and B (dark purple) reveals *meta*-CF_3-2-BMA (grays) bound in two alternate conformations. **C.** In chain A, *meta*-CF_3-2-BMA is coordinated by an extensive hydrogen bond network (orange dashes) that positions the *pro*-R carboxylate oxygen for nucleophilic attack (blue dashes); interatomic distances are shown over dashed lines in Å. **D.** In chain B, *meta*-CF_3-2-BMA is coordinated by similar hydrogen bonds, but in this case the *pro*-S carboxylate is rotated away from AMP-PNP with a loss of the hydrogen bond to R150 (red dashes) and a longer distance between the *pro*-S carboxylate nucleophile and the α-phosphate of AMP-PNP. **E.** Alignment of active site A with WT *Mm*PylRS bound to Pyl and AMP-PNP (PDB: 2ZCE, blue) illustrates the difference between the water-mediated hydrogen bonds (yellow dashes) to the α-amine of Pyl in PylRS *versus* the direct carboxyl to backbone hydrogen bonding of *m*-CF_3-2-BMA bound to *Ma*FRSA. **F.** Comparison of active site B with *Mm*BtaRS (N346G/C348Q) bound to Bta (PDB: 4ZIB, red) reveals similar binding modes with hydrogen bonds (yellow dashes) between the substrate carboxylate and amide backbone of the enzyme when N166/N346 (*Ma*/*Mm* numbering) is mutated.

In vitro **translation initiation with novel monomers** ***via*** **codon skipping**

We performed *in vitro* translation experiments to verify that the uniquely acylated derivatives of *Ma*-$tRNA^{Pyl}$ produced using *Ma*PylRS variants are effectively shuttled to and accommodated by the *E. coli* ribosome. Whereas the *E. coli* initiator $tRNA^{fMet}$ has been engineered into a substrate for *Mj*TyrRS variants to introduce non-canonical L-α-amino acids at the protein N-terminus (*49*), *Ma*-$tRNA^{Pyl}$ lacks the key sequence elements for recognition by *E. coli* initiation factors precluding its use for initiation *in vivo* (*36*, *58*). It has been reported (*59*) that in the absence of methionine, *in vitro* translation can begin at the second codon of the mRNA template, a phenomenon we refer to as "codon skipping". We took advantage of codon skipping and a commercial *in vitro* translation (IVT) kit to evaluate whether *Ma*-$tRNA^{Pyl}$ enzymatically charged with monomers **13**-**15** would support translation initiation.

To avoid competition with release factor 1 (RF1) at UAG codons, the anticodon of *Ma*-$tRNA^{Pyl}$ was mutated to ACC (*Ma*-$tRNA^{Pyl}$-ACC) to recode a glycine GGT codon at position 2 in the mRNA template. To maximize yields of acyl-tRNA, we increased the aaRS:tRNA ratio to 1:2 (monomers **7**, **14**, and **15**) or 1:1 (monomer **13**), extended the incubation time to 3 hours, and used the most active *Ma*FRSx variant for each monomer. These modifications resulted in acyl-

tRNA yields of 79% (**7**, *Ma*FRS1), 13% (**13**, *Ma*FRS2), 85% (**14**, *Ma*FRS2), and 82% (**15**, *Ma*FRS1). The acylated *Ma*-tRNA^{Pyl}-ACC was added with a DNA template encoding a short MGV-FLAG peptide (MGVDYKDDDDK) (Fig. 6A) to a commercial *in vitro* translation kit (PURExpress® Δ (aa, tRNA), NEB). When methionine was excluded from the reaction mixture, translation initiated at the second position by skipping the start codon to produce peptides with the sequence XVDYKDDDDK (X = **7**, **13**-**15**). Following FLAG-tag enrichment, LC-HRMS confirmed initiation with monomers **7** and **13**-**15** (Fig. 6, B to C). When the mass corresponding to the *m*/*z* = M+2H (**13**-**15**) or *m*/*z* = M+3H (**7**) charge state for each peptide was extracted from the total ion chromatogram, there was a clear peak for each peptide. No such peak was observed in reactions that lacked either the DNA template or acyl-tRNA, confirming templated ribosomal initiation with α-thio acid **13**, 2-benzylmalonic acid **14**, and *N*-formyl-L-α-amino acid **15**. Two peaks of identical mass are present in the EIC when translation was initiated with 2-benzylmalonic acid **14**, which we assign to diastereomeric peptides resulting from acylation at either the *pro*-S or *pro*-R carboxylate. Combined with the multiple peaks present in the RNAse A assay with 2-benzylmalonic acids **17, 18,** and **19** (Figs. 4A, S7), as well as the two *meta*-CF_3-2-BMA conformations observed in the structure of *Ma*FRSA (Fig. 5B), the two peptide products of identical mass generated in the IVT reactions imply that *Ma*FRSx enzymes can effectively acylate either of the two prochiral carboxylates of a malonic acid substrate.

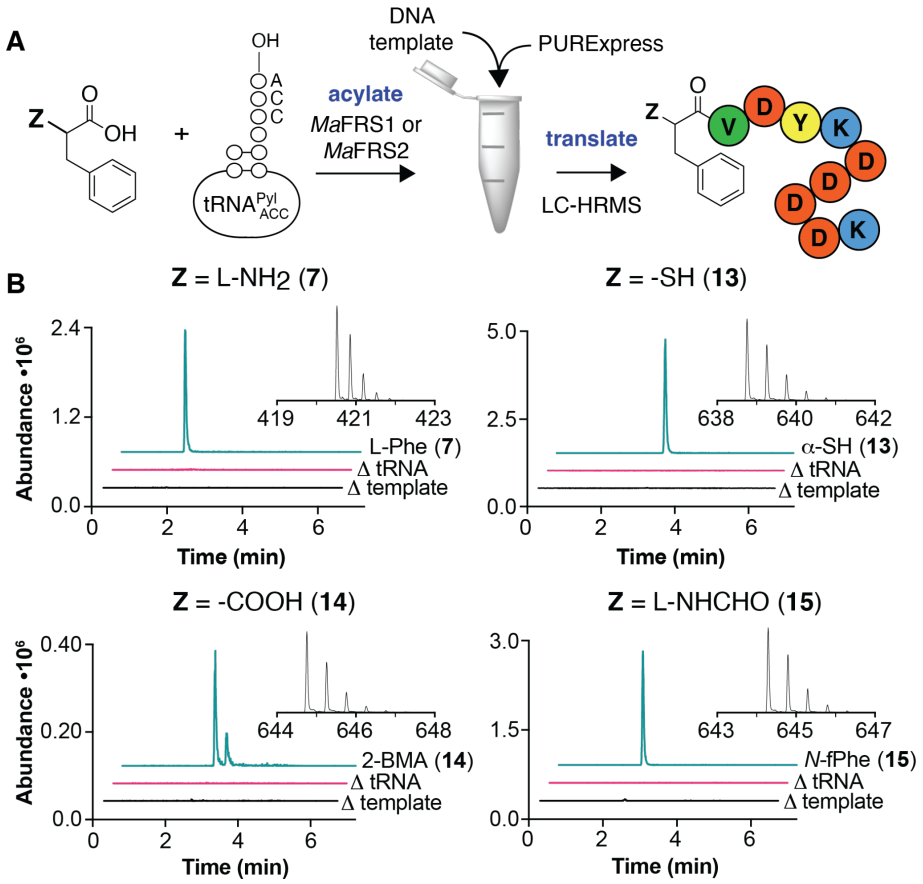


Figure 6. *In vitro* translation (IVT) of peptides initiated with monomers **7** (Z = L-NH_2), **13** (Z = -SH); **14** (Z = -COOH); and **15** (Z = L-NHCHO) *via* codon skipping. **A.** Workflow for *in vitro* translation. **B.** Extracted ion chromatograms (EICs) and mass spectra of peptide products obtained using *Ma*-tRNA^{Pyl}-ACC charged with monomers **7**, **13-15** by *Ma*FRS1 (**7** and **15**) or *Ma*FRS2 (**13** and **14**). Insets show mass spectra for major ions used to generate the EIC of the translated peptide initiated with the indicated monomer. Expected (exp) and observed (obs) *m/z* peaks in mass spectra are as follows: L-Phe **7** (M+3H), exp: 420.51906, obs: 420.52249; α-SH **13** (M+2H), exp: 638.75554, obs: 638.75614; *N*-fPhe **15** (M+2H), exp: 644.27242, obs: 644.27167; and 2-BMA **14** (M+2H), exp: 644.76442, obs: 644.76498.

Discussion

Expanding and reprogramming the genetic code for the templated biosynthesis of sequence-defined hetero-polymers demands orthogonal aminoacyl-tRNA synthetase/tRNA pairs that accept non-L-α-amino acid substrates. Fahnestock & Rich reported over fifty years ago that the peptidyl transferase center (PTC) of the *E. coli* ribosome tolerates an α-hydroxyl substituent in place of the α-amine of phenylalanine and described the *in vitro* biosynthesis of a polyester (*3*). More recent studies have broadened the scope of wild-type PTC reactivity to include other nucleophilic heteroatoms in place of the α-amine (*13–16*). However, with the exception of substrates carrying an α-hydroxyl substituent (*19–21*), none of these non-L-α-amino acid monomers are substrates for any known orthogonal aminoacyl-tRNA synthetase. The challenge is that most aminoacyl-tRNA synthetases with known structures simultaneously engage both the substrate α-amine and α-carboxylate moieties via multiple main-chain and side-chain hydrogen bonds to position the latter for adenylation and acylation. These well-conserved hydrogen bond networks complicate the engineering of new enzymes that accept substrates with alternative amine/carboxylate orientations or conformations (such as β-amino acids or aramids), or those whose α-substituents are large and/or electrostatically distinctive (such as malonates, α-thio acids, and *N*-acyl α-amino acids).

Yokoyama and coworkers demonstrated that the water-mediated hydrogen-bond network in the active site of *Methanosarcina mazei* pyrrolysyl-tRNA synthetase facilitated recognition of substrates with conservative substitutions (α-OH, α-H, α-$NHCH_3$) of the α-amino group (*20*). Here we expand the scope of monomers recognized and processed by PylRS variants to include α-thio acids and *N*-formyl-L-α-amino acids as well as those that carry an α-carboxyl functional group in place of the α-amine: pro-chiral malonic acids with protein-like side chains. Although thioesters and malonic acids are ubiquitous intermediates in polyketide and fatty acid biosynthesis (*60*, *61*), as far as we know, tRNAs enzymatically acylated with a polyketide precursor or an α-thio acid represent novel chemical species. Such tRNAs could forge a new synthetic link between ribosomal translation and assembly-line polyketide synthases (*62*), the two molecular machines responsible for protein and polyketide biosynthesis, respectively. Moreover, the availability of multiple mutually orthogonal synthetases that accept α-hydroxy acids provides a route to the *in vivo* production of mixed sequence polyesters.

Although many reactive nucleophiles support bond-forming reactions within the PTC of natural ribosomes, enolate ions are not, thus far, among this group. The reactive nucleophiles that operate within ribosomal PTC are all heteroatoms (N, O, S) that form new carbon-heteroatom bonds, whereas reaction of an enolate ion within the PTC would generate a carbon-carbon bond. Enolate ions are generated enzymatically under aqueous conditions upon decarboxylation of malonate half esters during both fatty acid and polyketide biosynthesis (*63–65*). In PKS active sites, decarboxylation is often facilitated by a pair of histidine side chains that coordinate the α-carboxylate and the enolate ion that results from decarboxylation (*66*). Although ribozymes have been reported to catalyze decarboxylative Claisen condensations in model systems (*67*), the ordered "proton wire" that promotes amide bond formation in the PTC (*68*) would likely need to reorganize to promote decarboxylation of a malonate precursor, as enolates and α-amino esters are conformationally distinct and not isosteric. Ribosomes with appropriately reorganized active sites are best selected for *in vivo*, but until now, no suitably active synthetase capable of supporting such selections was available. Combined with synthetic genomes (*69*, *70*) ribosomes capable of carbon-carbon bond formation would set the stage for the template-driven biosynthesis of unique hybrid biomaterials and sequence-defined polyketide-peptide oligomers, such as those produced by PKS-NRPS biosynthetic modules.

Acknowledgements: We thank Drs. Hasan Celik and Alicia Lund, and UC Berkeley's NMR facility in the College of Chemistry (CoC-NMR) for spectroscopic assistance, Prof. John Kuriyan for generously sharing crystallography equipment, Drs. James Holton and George Meigs for X-ray diffraction assistance, Dr. Shuai Zheng for advice on synthesis, and Isaac Knudson for advice on *in vitro* translation. The authors are grateful to Drs. Scott Miller, Jamie Cate, and Abhishek Chatterjee for meaningful discussion and feedback. **Funding:** This work was supported by the NSF Center for Genetically Encoded Materials (C-GEM), CHE 2002182. R.F. was supported by Agilent Technologies. C.S. was supported by a Berkeley Fellowship for Graduate Study. L.T.R. was supported by the NSF Graduate Research Fellowship Program under Grant No. 1752814. Beamline 8.3.1 at the Advanced Light Source is operated by the University of California at San Francisco with generous grants from the National Institutes of Health (R01 GM124149 and P30 GM124169), Plexxikon Inc., and the Integrated Diffraction Analysis Technologies program of the US Department of Energy Office of Biological and Environmental

Research. The Advanced Light Source (Berkeley, CA) is a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the US Department of Energy under contract number DE-AC02-05CH11231, Office of Basic Energy Sciences. **Author contributions:** R.F., C.S., O.A., and A.S. designed the project. R.F. led acylation experiments and assisted with crystallography and *in vitro* translation. C.S. led crystallography and *in vitro* translation and assisted with acylation experiments. L.T.R. synthesized materials and assisted with crystallography. N.H. assisted with acylation experiments. S.S. and C.G. assisted with crystallography. R.F., C.S., and A.S. prepared the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text and the supplementary materials. Structural model of *Ma*FRSA is available in the Protein Data Bank with code 7U0R.

References and notes

1. J. Lutz, M. Ouchi, D. Liu, M. Sawamoto, Sequence-Controlled Polymers. *Science*. **341**, 628- (2013).
2. J. Barnes, D. Ehrlich, A. Gao, F. Leibfarth, Y. Jiang, E. Zhou, T. Jamison, J. Johnson, Iterative exponential growth of stereo- and sequence-controlled polymers. *Nat. Chem.* **7**, 810–815 (2015).
3. S. Fahnestock, A. Rich, Ribosome-catalyzed polyester formation. *Science*. **173**, 340–343 (1971).
4. A. Ohta, H. Murakami, E. Higashimura, H. Suga, Synthesis of polyester by means of genetic code reprogramming. *Chem. Biol.* **14**, 1315–1322 (2007).
5. T. Katoh, Y. Iwane, H. Suga, Logical engineering of D-arm and T-stem of tRNA that enhances d-amino acid incorporation. *Nucleic Acids Res.* **45**, 12601–12610 (2017).
6. T. Fujino, Y. Goto, H. Suga, H. Murakami, Ribosomal Synthesis of Peptides with Multiple β-Amino Acids. *J. Am. Chem. Soc.* **138**, 1962–1969 (2016).
7. T. Katoh, H. Suga, Ribosomal Incorporation of Consecutive β-Amino Acids. *J. Am. Chem. Soc.* **140**, 12159–12167 (2018).
8. E. Adaligil, A. Song, K. K. Hallenbeck, C. N. Cunningham, W. J. Fairbrother, Ribosomal Synthesis of Macrocyclic Peptides with β^2 - and $\beta^{2,3}$ -Homo-Amino Acids for the Development of Natural Product-Like Combinatorial Libraries. *ACS Chem. Biol.* **16**, 1011–1018 (2021).

9. T. Katoh, T. Sengoku, K. Hirata, K. Ogata, H. Suga, Ribosomal synthesis and de novo discovery of bioactive foldamer peptides containing cyclic β-amino acids. *Nat. Chem.* **12**, 1081–1088 (2020).
10. T. Katoh, H. Suga, Ribosomal Elongation of Cyclic γ-Amino Acids using a Reprogrammed Genetic Code. *J. Am. Chem. Soc.* **142**, 4965–4969 (2020).
11. E. Adaligil, A. Song, C. N. Cunningham, W. J. Fairbrother, Ribosomal Synthesis of Macrocyclic Peptides with Linear γ^4 - and β-Hydroxy-γ^4 -amino Acids. *ACS Chem. Biol.* **16**, 1325–1331 (2021).
12. J. Lee, K. J. Schwarz, D. S. Kim, J. S. Moore, M. C. Jewett, Ribosome-mediated polymerization of long chain carbon and cyclic amino acids into peptides in vitro. *Nat. Commun.* **11**, 4304 (2020).
13. T. Katoh, H. Suga, Consecutive Ribosomal Incorporation of α-Aminoxy/α-Hydrazino Acids with l / d -Configurations into Nascent Peptide Chains. *J. Am. Chem. Soc.* **143**, 18844–18848 (2021).
14. R. Takatsuji, K. Shinbara, T. Katoh, Y. Goto, T. Passioura, R. Yajima, Y. Komatsu, H. Suga, Ribosomal Synthesis of Backbone-Cyclic Peptides Compatible with In Vitro Display. *J. Am. Chem. Soc.* **141**, 2279–2287 (2019).
15. T. Katoh, H. Suga, Ribosomal Elongation of Aminobenzoic Acid Derivatives. *J. Am. Chem. Soc.* **142**, 16518–16522 (2020).
16. O. Ad, K. S. Hoffman, A. G. Cairns, A. L. Featherston, S. J. Miller, D. Söll, A. Schepartz, Translation of Diverse Aramid- and 1,3-Dicarbonyl-peptides by Wild Type Ribosomes *in Vitro*. *ACS Cent. Sci.* **5**, 1289–1294 (2019).
17. A. Sievers, M. Beringer, M. V. Rodnina, R. Wolfenden, The ribosome as an entropy trap. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7897 (2004).
18. P. M. England, Y. Zhang, D. A. Dougherty, H. A. Lester, Backbone Mutations in Transmembrane Domains of a Ligand-Gated Ion Channel: Implications for the Mechanism of Gating. *Cell*. **96**, 89–98 (1999).
19. J. Guo, J. Wang, J. C. Anderson, P. G. Schultz, Addition of an α-Hydroxy Acid to the Genetic Code of Bacteria. *Angew. Chem.* **120**, 734–737 (2008).
20. T. Kobayashi, T. Yanagisawa, K. Sakamoto, S. Yokoyama, Recognition of Non-α-amino Substrates by Pyrrolysyl-tRNA Synthetase. *J. Mol. Biol.* **385**, 1352–1360 (2009).
21. Y.-M. Li, M.-Y. Yang, Y.-C. Huang, Y.-T. Li, P. R. Chen, L. Liu, Ligation of Expressed

Protein α-Hydrazides *via* Genetic Incorporation of an α-Hydroxy Acid. *ACS Chem. Biol.* **7**, 1015–1022 (2012).

22. C. Melo Czekster, W. E. Robertson, A. S. Walker, D. Söll, A. Schepartz, In Vivo Biosynthesis of a β-Amino Acid-Containing Protein. *J. Am. Chem. Soc.* **138**, 5194–5197 (2016).
23. S. Chen, X. Ji, M. Gao, L. M. Dedkova, S. M. Hecht, In Cellulo Synthesis of Proteins Containing a Fluorescent Oxazole Amino Acid. *J. Am. Chem. Soc.* **141**, 5597–5601 (2019).
24. B. I. Morinaka, E. Lakis, M. Verest, M. J. Helf, T. Scalvenzi, A. L. Vagstad, J. Sims, S. Sunagawa, M. Gugger, J. Piel, Natural noncanonical protein splicing yields products with diverse β-amino acid residues. *Science*. **359**, 779–782 (2018).
25. C. C. Liu, P. G. Schultz, Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
26. W. Wan, J. M. Tharp, W. R. Liu, Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta BBA - Proteins Proteomics*. **1844**, 1059–1070 (2014).
27. O. Vargas-Rodriguez, A. Sevostyanova, D. Söll, A. Crnković, Upgrading aminoacyl-tRNA synthetases for genetic code expansion. *Curr. Opin. Chem. Biol.* **46**, 115–122 (2018).
28. J. S. Italia, P. S. Addy, S. B. Erickson, J. C. Peeler, E. Weerapana, A. Chatterjee, Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites. *J. Am. Chem. Soc.* **141**, 6204–6212 (2019).
29. J. W. Chin, Expanding and reprogramming the genetic code. *Nature*. **550**, 53–60 (2017).
30. Srinivasan Gayathri, James Carey M., Krzycki Joseph A., Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA. *Science*. **296**, 1459–1462 (2002).
31. A. Ambrogelly, S. Palioura, D. Soll, Natural expansion of the genetic code. *Nat. Chem. Biol.* **3**, 29–35 (2007).
32. L. Wang, A. Brock, B. Herberich, P. G. Schultz, Expanding the Genetic Code of *Escherichia coli*. *Science*. **292**, 498–500 (2001).
33. T. Kobayashi, O. Nureki, R. Ishitani, A. Yaremchuk, M. Tukalo, S. Cusack, K. Sakamoto, S. Yokoyama, Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Biol.* **10**, 8 (2003).
34. T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama,

Crystallographic Studies on Multiple Conformational States of Active-site Loops in Pyrrolysyl-tRNA Synthetase. *J. Mol. Biol.* **378**, 634–652 (2008).

35. G. Borrel, H. M. B. Harris, W. Tottey, A. Mihajlovski, N. Parisot, E. Peyretaillade, P. Peyret, S. Gribaldo, P. W. O'Toole, J.-F. Brugere, Genome Sequence of "Candidatus Methanomethylophilus alvus" Mx1201, a Methanogenic Archaeon from the Human Gut Belonging to a Seventh Order of Methanogens. *J. Bacteriol.* **194**, 6944–6945 (2012).
36. J. C. W. Willis, J. W. Chin, Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat. Chem.* **10**, 831–837 (2018).
37. Y.-S. Wang, W. K. Russell, Z. Wang, W. Wan, L. E. Dodd, P.-J. Pai, D. H. Russell, W. R. Liu, The de novo engineering of pyrrolysyl-tRNA synthetase for genetic incorporation of l-phenylalanine and its derivatives. *Mol. Biosyst.* **7**, 714 (2011).
38. Y.-S. Wang, X. Fang, A. L. Wallace, B. Wu, W. R. Liu, A Rationally Designed Pyrrolysyl-tRNA Synthetase Mutant with a Broad Substrate Spectrum. *J. Am. Chem. Soc.* **134**, 2950–2953 (2012).
39. S. Herring, A. Ambrogelly, S. Gundllapalli, P. O'Donoghue, C. R. Polycarpo, D. Söll, The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity. *FEBS Lett.* **581**, 3197–3203 (2007).
40. J. L. McMurry, M. C. Y. Chang, Fluorothreonyl-tRNA deacylase prevents mistranslation in the organofluorine producer *Streptomyces cattleya*. *Proc. Natl. Acad. Sci.* **114**, 11920–11925 (2017).
41. D. Findly, D. G. Herries, A. P. Mathias, B. R. Rabin, C. A. Ross, The Active Site and Mechanism of Action of Bovine Pancreatic Ribonuclease. *Nature*. **190**, 781–784 (1961).
42. B. E. Griffin, M. Jarman, C. B. Reese, J. E. Sulston, D. R. Trentham, Some Observations Relating to Acyl Mobility in Aminoacyl Soluble Ribonucleic Acids*. *Biochemistry*. **5**, 3638–3649 (1966).
43. V. G. Stepanov, N. A. Moor, V. N. Ankilova, O. I. Lavrik, Phenylalanyl-tRNA synthetase from *Thermus thermophilus* can attach two molecules of phenylalanine to $tRNA^{Phe}$. *FEBS Lett.* **311**, 192–194 (1992).
44. B. Wang, J. Zhou, M. Lodder, R. D. Anderson, S. M. Hecht, Tandemly Activated tRNAs as Participants in Protein Synthesis. *J. Biol. Chem.* **281**, 13865–13868 (2006).
45. J. Ko, Y.-S. Wang, A. Nakamura, L.-T. Guo, D. Söll, T. Umehara, Pyrrolysyl-tRNA synthetase variants reveal ancestral aminoacylation function. *FEBS Lett.* **587**, 3243–3248

(2013).

46. W. Xuan, D. Collins, M. Koh, S. Shao, A. Yao, H. Xiao, P. Garner, P. G. Schultz, Site-Specific Incorporation of a Thioester Containing Amino Acid into Proteins. *ACS Chem. Biol.* **13**, 578–581 (2018).
47. S. Hwang, N. Lee, S. Cho, B. Palsson, B.-K. Cho, Repurposing Modular Polyketide Synthases and Non-ribosomal Peptide Synthetases for Novel Chemical Biosynthesis. *Front. Mol. Biosci.* **7** (2020), doi:10.3389/fmolb.2020.00087.
48. U. Varshney, C. P. Lee, B. L. Seong, U. L. RajBhandary, Mutants of initiator tRNA that function both as initiators and elongators. *J. Biol. Chem.* **266**, 18018–18024 (1991).
49. J. M. Tharp, O. Ad, K. Amikura, F. R. Ward, E. M. Garcia, J. H. D. Cate, A. Schepartz, D. Söll, Initiation of Protein Synthesis with Non-Canonical Amino Acids In Vivo. *Angew. Chem. Int. Ed.* **59**, 3122–3126 (2020).
50. I. Cestari, K. Stuart, A Spectrophotometric Assay for Quantitative Measurement of Aminoacyl-tRNA Synthetase Activity. *J. Biomol. Screen.* **18**, 490–497 (2013).
51. Y.-S. Wang, X. Fang, H.-Y. Chen, B. Wu, Z. U. Wang, C. Hilty, W. R. Liu, Genetic Incorporation of Twelve *meta* -Substituted Phenylalanine Derivatives Using a Single Pyrrolysyl-tRNA Synthetase Mutant. *ACS Chem. Biol.* **8**, 405–415 (2013).
52. Y.-J. Lee, M. J. Schmidt, J. M. Tharp, A. Weber, A. L. Koenig, H. Zheng, J. Gao, M. L. Waters, D. Summerer, W. R. Liu, Genetically encoded fluorophenylalanines enable insights into the recognition of lysine trimethylation by an epigenetic reader. *Chem. Commun.* **52**, 12606–12609 (2016).
53. L.-T. Guo, Y.-S. Wang, A. Nakamura, D. Eiler, J. M. Kavran, M. Wong, L. L. Kiessling, T. A. Steitz, P. O'Donoghue, D. Söll, Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. *Proc. Natl. Acad. Sci.* **111**, 16724–16729 (2014).
54. J. M. Tharp, Y.-S. Wang, Y.-J. Lee, Y. Yang, W. R. Liu, Genetic Incorporation of Seven *ortho* -Substituted Phenylalanine Derivatives. *ACS Chem. Biol.* **9**, 884–890 (2014).
55. E. Seki, T. Yanagisawa, M. Kuratani, K. Sakamoto, S. Yokoyama, Fully Productive Cell-Free Genetic Code Expansion by Structure-Based Engineering of *Methanomethylophilus alvus* Pyrrolysyl-tRNA Synthetase. *ACS Synth. Biol.* **9**, 718–732 (2020).
56. J. M. Kavran, S. Gundllapalli, P. O'Donoghue, M. Englert, D. Soll, T. A. Steitz, Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci.* **104**, 11268–11273 (2007).

57. M. Englert, A. Nakamura, Y.-S. Wang, D. Eiler, D. Söll, L.-T. Guo, Probing the active site tryptophan of *Staphylococcus aureus* thioredoxin with an analog. *Nucleic Acids Res.* **43**, 11061–11067 (2015).
58. B. S. Laursen, H. P. Sørensen, K. K. Mortensen, H. U. Sperling-Petersen, Initiation of Protein Synthesis in Bacteria. *Microbiol. Mol. Biol. Rev.* **69**, 101–123 (2005).
59. J. Lee, K. E. Schwieter, A. M. Watkins, D. S. Kim, H. Yu, K. J. Schwarz, J. Lim, J. Coronado, M. Byrom, E. V. Anslyn, A. D. Ellington, J. S. Moore, M. C. Jewett, Expanding the limits of the second genetic code with ribozymes. *Nat. Commun.* **10**, 5097 (2019).
60. A. Nivina, K. Yuet, J. Hsu, C. Khosla, Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem. Rev.* **119**, 12524–12547 (2019).
61. S. Tsai, in *Annual Review of Biochemistry*, R. Kornberg, Ed. (2018), vol. 87, pp. 503–531.
62. C. T. Walsh, R. V. O'Brien, C. Khosla, Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew. Chem. Int. Ed.* **52**, 7098–7124 (2013).
63. P. R. Vagelos, A. W. Alberts, Malonyl Coenzyme A-CO2 Exchange Reaction. *J. Biol. Chem.* **235**, 2786–2791 (1960).
64. R. J. Heath, C. O. Rock, The Claisen condensation in biology. *Nat Prod Rep*. **19**, 581–596 (2002).
65. C. Hertweck, The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed.* **48**, 4688–4716 (2009).
66. T. Robbins, J. Kapilivsky, D. Cane, C. Khosla, Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. *Biochemistry*. **55**, 4476–4484 (2016).
67. Y. Ryu, K.-J. Kim, C. A. Roessner, A. I. Scott, Decarboxylative Claisen condensation catalyzed by in vitro selected ribozymes. *Chem. Commun.*, 1439-1441 (2006).
68. Y. S. Polikanov, T. A. Steitz, C. A. Innis, A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome. *Nat. Struct. Mol. Biol.* **21**, 787–793 (2014).
69. N. Ostrov, M. Landon, M. Guell, G. Kuznetsov, J. Teramoto, N. Cervantes, M. Zhou, K. Singh, M. Napolitano, M. Moosburner, E. Shrock, B. Pruitt, N. Conway, D. Goodman, C. Gardner, G. Tyree, A. Gonzales, B. Wanner, J. Norville, M. Lajoie, G. Church, Design, synthesis, and testing toward a 57-codon genome. *Science*. **353**, 819–822 (2016).

70. J. Fredens, K. Wang, D. de la Torre, L. F. H. Funke, W. E. Robertson, Y. Christova, T. Chia, W. H. Schmied, D. L. Dunkelmann, V. Beranek, C. Uttamapinant, A. G. Llamazares, T. S. Elliott, J. W. Chin, Total synthesis of Escherichia coli with a recoded genome. *Nature*. **569**, 514–518 (2019).