Improving Target Amino Acid Selectivity in a Permissive Aminoacyl tRNA Synthetase Through Counter-Selection

Itthipol Sungwienwong,[a,†] Zachary M. Hostetler,[b,†] Robert J. Blizzard,[c] Joseph J. Porter,[c] Camden M. Driggers,[d] Lea Z. Mbengi,[d] José A. Villegas,[a] Lee C. Speight,[a] Jeffery G. Saven,[a] John J. Perona,[d,e] Rahul M. Kohli,[b] Ryan A. Mehl,[c] and E. James Petersson[a,*]

a. Department of Chemistry, University of Pennsylvania, 213 South 34th Street, Philadelphia, PA 19104, USA. Email: ejpetersson@sas.upenn.edu; Tel: 1-215-746-2221

b. Department of Medicine, Perelman School of Medicine, University of Pennsylvania, 3610 Hamilton Walk, Philadelphia, PA 19104

c. Oregon State University, Department of Biochemistry and Biophysics, 2011 Agriculture and Life Sciences Building, Corvallis, Oregon 97331, USA

d. Department of Chemistry, Portland State University, P.O. Box 751, Portland, Oregon 97207, USA

e. Department of Biochemistry & Molecular Biology, Oregon Health & Sciences University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97239, USA

† These authors contributed equally to this work.

Table of Contents

General Information

Materials. L-Tyrosine, thionyl chloride, di-*tert*-butyl dicarbonate (Boc anhydride), methyl 2-aminobenzoate, and phenylsepharose CL-4B resin were purchased from Sigma-Aldrich (St. Louis, MO, USA). *N*-Phenyl-bis(trifluoromethane sulfonimide) was purchased from Oakwood Chemical (West Columbia, SC, USA). *E. coli* BL21(DE3) cells were purchased from Stratagene (La Jolla, CA, USA). *E. coli* ElectroMAX DH10B cells were purchased from Invitrogen (Grand Island, NY, USA). Milli-Q filtered (18 MΩ) water was used for all solutions (Millipore; Billerica, MA, USA). Bradford reagent assay kits were purchased from BioRAD (Hercules, CA, USA). Amicon Ultra centrifugal filter units (3 kDa MWCO) were purchased from EMD Millipore. All other reagents and solvents were purchased from Fisher Scientific or Sigma-Aldrich unless otherwise specified. DNA sequencing was performed at the University of Pennsylvania DNA sequencing facility.

Instruments. Low resolution electrospray ionization mass spectra (ESI-LRMS) were obtained on a Waters Acquity Ultra Performance LC connected to a single quadrupole detector (SQD) mass spectrometer. UV-Vis absorption spectra were acquired on a Hewlett-Packard 8452A diode array spectrophotometer (currently Agilent Technologies; Santa Clara, CA, USA). Nuclear magnetic resonance (NMR) spectra were obtained on a Bruker DRX 500 MHz instrument (Billerica, MA, USA). Matrix assisted laser desorption/ionization with time-of-flight detector (MALDI-TOF) mass spectra were acquired on a Bruker Ultraflex III instrument. Analytical HPLC was performed on an Agilent 1100 Series HPLC system. Preparative HPLC was performed on a Varian Prostar HPLC system (currently Agilent Technologies). HPLC columns were purchased from W. R. Grace & Compnay (Columbia, MD, USA).

Chemical Synthesis

Methyl (S)-2-((tert-butoxycarbonyl)amino)-3-(4-((((trifluoromethyl)sulfonyl)oxy)phenyl) propanoate (Boc-L-Tyr(OTf)-OMe, 3). L-Tyrosine (**2**) (5.07 g, 27.99 mmol) was reacted with SOCl₂ 10.21 mL (139.94 mmol) in 50 mL of cooled MeOH. The reaction was allowed to stir overnight at room temperature under Ar. The reaction mixture was concentrated under vacuum and washed with MeOH. 6.44 g of crude L-tyrosine methyl ester (99% yield) was obtained after drying under vacuum overnight. The crude product was further dissolved in 120 mL tetrahydrofuran (THF)/H₂O (3:1). Na₂CO₃ 3.37 g (31.80 mmol) was then added, and the reaction was stirred for 10 min. Then 6.98 g Boc₂O was added, and the solution was stirred at room temperature overnight. 100 mL H₂O was added to the solution and then the reaction was acidified with 3 M HCl to pH 3. The organic layer was extracted with ethyl acetate (EtOAc), dried with MgSO₄, and concentrated under reduced pressure. The product can be recrystallized by dissolution in THF/hexane (3:1). *N*-Boc-L-tyrosine methyl ester (4.50 g, 15.24 mmol) was further dissolved in 30 mL CH₂Cl₂. Triethylamine (6.38 mL, 45.72 mmol), 4-(dimethylamino) pyridine (0.186 g, 1.524 mmol), and *N*-phenyl bis(trifluoro-methanesulfonimide) (7.62g, 21.34mmol) were then added, respectively. The reaction was stirred overnight. The reaction mixture was concentrated under reduced pressure and worked up with NH₄Cl and then extracted by EtOAc. The crude product was purified by flash chromatography, eluting with 8:2 hexanes/EtOAc. (R_f = 0.26). The purified product Boc-L-Tyr-OTf-OMe (**3**) was obtained as colorless oil 5.77 g (88.62% yield). ESI-LRMS and NMR spectra matched previous reports.[1]

(S)-Methyl 2-((4-(2-((tert-butoxycarbonyl)amino)-3-methoxy-3-oxopropyl)phenyl)amino) benzoate (5). 75 mL degassed toluene was added to Boc-L-Tyr-OTf-OMe (**3**) (3.000 g, 7.02 mmol) in a dried round-bottom flask, followed by methyl 2-aminobenzoate (1200 µL, 9.27

mmol). The solution was degassed with Ar for 5 min. Then palladium(II) acetate (0.082 g, 0.365 mmol), and racemic 2,2′- bis(diphenyl-phosphino)-1,1′-binaphthyl (0.054 g, 0.087 mmol) were added to the flask. Cesium carbonate (6.88 g, 21.1 mmol) was ground and added to the flask. The flask was then fitted with a reflux condenser and heated to 135 °C for 23 h. After the solution was allowed to cool to ambient temperature, the contents were filtered through a short plug of silica gel using CH_2Cl_2 to transfer the material to the silica (200 mL), and then ethyl acetate (300 mL) was used to elute the product. The clarified solution was then concentrated under reduced pressure and further purified by flash column chromatography (9:1 Hexane/EtOAc) afforded 2.759 g of compound **5** (6.44 mmol, 91.7%) as a yellow oil. ($R_f = 0.2$ in 85:15 Hexane/EtOAc) ESI-LRMS and NMR spectra matched previous reports.[1]

Acridon-2-ylalanine (Acd, 1) (H_2SO_4 route). A solution of 13.5 M sulfuric acid (12 mL) was added to a flask containing **5** (1.02 g, 2.38 mmol). The flask was then fitted with a reflux condenser and heated to 115 °C for 16 h in an oil bath. 80 mL water was then added to the flask and allowed to stir for 15 min. The reaction was then removed from the hot oil bath and allowed to cool down. Upon reaching ambient temperature, the solution was cooled to 4 °C and allowed to stand for 2 h. 100 g of ion-exchange resin (Dowex® 50WX8 hydrogen form, strongly acidic cation exchange resin) was made into a slurry with 1.8 M aqueous H_2SO_4 and applied to a flash chromatography column. The resin was washed with 350 mL 1.8 M aqueous H_2SO_4, 2 L of water, 1 L of 1.5 M aqueous NH_4OH, and 4 L of water. Following these washes, the resin was dried by passing air through the column. The cooled Acd solution was then vacuum filtered on a Büchner funnel to remove precipitated material, and the clarified solution was applied to the washed and dried ion-exchange resin.   The resulting resin slurry was shaken in the chromatography column for 5 min before the solution was drained. This solution was then

reapplied to the dried resin and shaken for an additional 5 min. The twice-passed solution was then set aside. The loaded resin was washed with 4 L water before the compound of interest was eluted with 1.45 L of 1.5 M NH_4OH. The solution was concentrated to 50 mL by rotary evaporation, and then lyophilized to dryness, yielding a crop of Acd (1) as a yellow powder (0.6375 g 2.26 mmol 94.9%). The ion-exchange resin was recycled by washing with 4 L of water and dried until further use. To maximize yield, the twice-passed solution was reapplied to washed and dried ion- exchange resin, and the process was repeated to yield a second crop of Acd (0.0232 g of 0.082 mmol 3.4%).

Acridon-2-ylalanine (Acd, 1) (PPA route). Compound **5** 2.16 g (5.04 mmol) was added to a round-bottom flask. Then 50 mL THF was added and the solution cooled to 4 °C. LiOH (3.31 g, 138.15 mmol) was dissolved with 150 mL water. The LiOH solution was then slowly added to the reaction. The mixture was stirred under Ar at 4 °C overnight. The pH of the reaction mixture was adjusted to 3 with 6 M HCl. Crude product was extracted with CH_2Cl_2 (3 x 100 mL), dried with $MgSO_4$, and concentrated under reduced pressure. The resulting crude oil was re-dissolved in 25 mL CH_2Cl_2 and cooled to 4 °C for 16 h. A first crop of recrystallized product was collected by vacuum filtration and washed with 100 mL of cold CH_2Cl_2 (1.63 g, 4.08 mmol, 80.8 %). The product was obtained as a white powder (82.6 % yield). The product was further reacted with polyphosphoric acid (PPA).

66.62 g of polyphosphoric acid (PPA) was added into a round-bottom flask with a stir bar. A round-bottom flask was heated to 135 °C in an oil bath. The depotected compound **7** (1.43 g, 3.57 mmol) was added to the flask and stirred for 2 h. Then 50 mL water was slowly added (2 mL portions over 10 min) and the reaction allowed to cool to 60 °C. After stirring for 1 h at 60 °C, the reaction was cooled to ambient temperature. Insoluble impurities were removed

by vacuum filtration and the clarified solution was adjusted the pH to 4 by addition of 8 M NaOH. Then the solution was cooled to 4 °C for precipitation about 16 h. Crude product was collected by vacuum filtration. After drying, the crude material was re-suspended in 50 mL water and brought into solution by adjusting the pH to 9.0. Insoluble impurities were removed by vacuum filtration. The pH of the clarified solution was then adjusted to 5.5 with 6 M HCl. The yellow precipitate was collected by vacuum filtration and dried overnight affording 0.66 g of Acd **(1)** (1.86 mmol, 65.3% yield).

Synthetase Selection

Positive and Negative Selection for Acd Specifity. A standard double sieve selection was performed in DH10B *Escherichia coli* (*E. coli*) with *Methanococcus jannaschii* (Mj) aminoacyl tRNA synthetase (RS) library 3D (3D-Lib).[2] The first positive and negative rounds of selection containted 1 mM Acd and no ncAA in the selection media, respectively. The second positive selection contained 1 mM Acd, while the second negative selection contained 1 mM Npf in the media. One microliter of the remaining library resulting from the second positive and the second negative rounds of selection was transformed with 60 μL of pALS plasmid containing DH10B cells.[2] The pALS plasmid contains a sfGFP reporter with a TAG codon at residue 150 as well as tyrosyl-tRNA$_{CUA}$. The cells were rescued for 1 h in 1 mL of SOC media (37 °C, 250 rpm). A 250 μL or 50 μL volume of cells from each library were plated on autoinducing media agar with 25 μg/mL kanamycin (Kn), 50 μg/mL tetracycline (Tet), and 1 mM Acd. Plates were grown at 37 °C for 24 h, and then grown on the bench top, at room temperature, for an additional 24 h. Autoinducing agar plates were prepared by combining the reagents with an autoclaved solution

of 40 g of agar in 400 mL water.[2] Sterile water was added to a final volume of 500 mL. Antibiotics were added to a final concentration of 25 μg/mL Tet and 50 μg/mL Kn.

A total of 96 visually green colonies were selected from the 1 mM Acd plates and used to inoculate a 96-well plate containing 0.5 mL per well non-inducing media with 25 μg/mL Kn and 25 μg/mL Tet.[2] After 24 h of growth (37 °C, 250 rpm), 5 μL of these non-inducing samples were used to inoculate three 96-well plates with 0.5 mL autoinduction media containing 25 μg/mL Kn and 25 μg/mL Tet. For the three 96-well plates, one plate contained 1 mM Acd, the second contained 1 mM Npf and the third contained no ncAA.

Fluorescence measurements of the cultures were collected 36 h after inoculation using a BIOTEK® Synergy 2 Microplate Reader. The emission was measured at 528 nm (20 nm bandwidth) with excitation at 485 nm (20 nm bandwidth). Samples were prepared by diluting suspended cells directly from culture 2-fold with phosphate buffer saline (PBS, 10 mM PO_4^{-3}, 137 mM NaCl, 2.7 mM KCl pH 7.4). Eleven unique RSs showed high efficiency with Acd, good fidelity without ncAA, and reduced ability to incorporate Npf. Fluorescence data from the G2, F9, G11, and A9 clones are shown in Fig. 1 in the main text, and the sequences of all 11 RS variants are shown in Table S1.

Quantitative Analysis of Highest-Fluorescing Clones. Non-inducing cultures (3 mL) with 25 μg/mL Kn and 25 μg/mL Tet were grown to saturation (37 °C with shaking at 250 rpm) from the top eleven expressions in the 96 well plate analysis. Autoinduction cultures (5 mL) with 25 μg/mL Kn and 25 μg/mL Tet were inoculated with 30 μL of non-inducing cultures and grown with 1 mM Acd, 1 mM Npf, or without ncAA at 37 °C with shaking at 250 rpm. All cultures were grown in triplicate. After approximately 40 hours, fluorescence was assessed as described (Fig. S1).
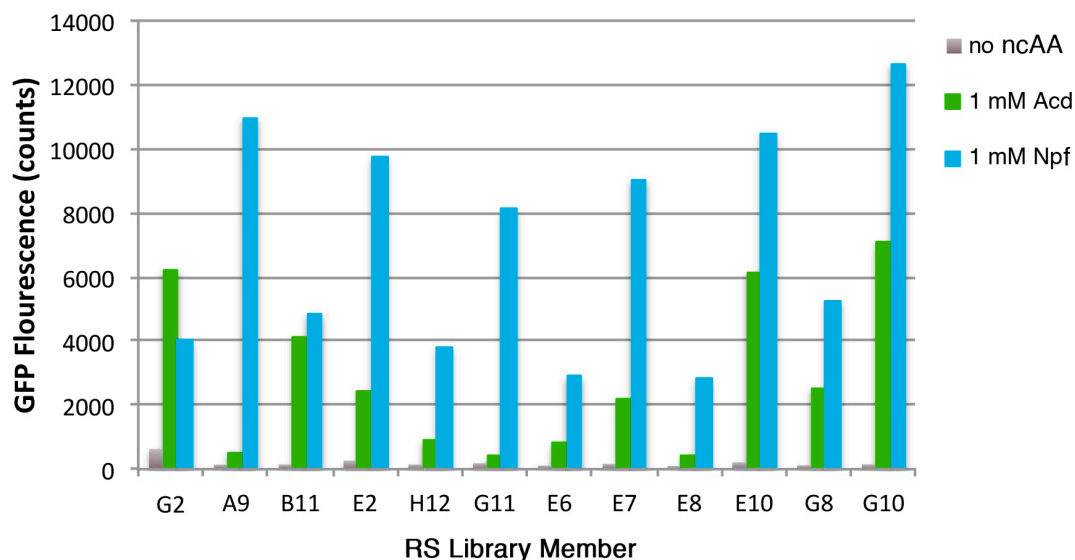
Fig. S1. Fluorescence measurements of RSs with GFP reporter. Grey, green, and blue represent fluorescence from colonies induced in media containing no ncAA, 1 mM Acd, or 1 mM Npf, respectively. Expressions of 500 µL were grown for 48 hours before 2-fold dilution of suspended cells directly from culture with PBS. Fluorescence measurements were collected using a BIOTEK® Synergy 2 Microplate Reader.

Table S1. Sequence of top performing Acd-RSs. The A9 and G11 clones were moved into the pDule2 plasmid for protein expression.

Residue number

Variant	32	65	70	108	109	155	158	159	162
WT	Y	L	H	F	Q	Q	D	I	L
G2	G	E	H	W	M	Q	S	I	K
A9	A	D	H	L	E	Q	S	I	T
B11	G	V	H	L	L	Q	E	I	T
E2	G	V	H	A	S	Q	S	I	E
H12	A	W	H	T	E	Q	G	I	A
G11	A	Y	H	H	E	Q	S	I	S
E6	G	P	H	E	S	Q	G	I	H
E7	G	P	H	E	S	Q	G	I	H
E10	G	I	H	E	S	Q	G	I	T
G8	A	P	H	S	Q	Q	D	I	A
G10	A	W	H	R	E	Q	G	I	S

S8

In Vivo **Selectivity Anaylsis**

Acd Mutant Calmodulin Protein Expression. For calmodulin (CaM) mutant expression, pCaM-L$_{113}$TAG and pDule2-AcdRS plasmids was used to transform *E. coli* BL21(DE3) cells.[1] The pDule2-AcdRS1, pDule2-AcdRS2a, and pDule2-AcdRS2b plasmids encode AcdRS1(G2), AcdRS2a(G11) and AcdRS2b(A9), respectively. The pCaM-L$_{113}$TAG plasmid confers ampicillin (Amp) resistance and the pDule2-AcdRS plasmids confer streptomycin (Strep) and spectinomycin (Spec) resistance. Transformed cells were selected on the basis of Amp and Strep or Amp and Spec resistance. Single colonies were used to inoculate 5 mL of LB media supplemented with Amp and Strep (100 µg/mL each). To an autoclaved solution containing 42 mM Na$_2$HPO$_4$, 22 mM KH$_2$PO$_4$, 19 mM NH$_4$Cl, and 86 mM NaCl (M9 salts), the following autoclaved solutions were added per liter of M9 salts: 1 mL of 2 M MgSO$_4$, 1 mL of 15 mg/mL FeCl$_2$ (in 1.0 M HCl), 1 mL of 15 mg/mL ZnCl$_2$ (in acidified H$_2$O), 6.25 mL of 40% glucose, 100 µL of 1 M CaCl$_2$ and 2 mL of 10% BactoTM Yeast Extract. The primary 5 mL culture was incubated at 37 °C with shaking at 250 rpm for 4 h. Upon reaching saturation, the primary culture was added to 1 L of M9 minimal media supplemented with Amp and Strep (or Spec). The 100 mL culture was incubated at 37 °C with shaking at 250 rpm until the OD$_{600}$ reached 0.7. At this point, the different amounts of unnatural amino acid were added to the culture depending on the following conditions:

A: a solution of 28.23 mg Acd (**1**) from H$_2$SO$_4$ route in 5 mL sterile water

B: a solution of 28.23 mg Acd (**1**) from PPA route

C: a solution of 25.63 mg Npf (**6**)

D: a solution of 0.26 mg Npf (**6**) and 27.95 mg Acd (**1**)

E: a solution of 2.56 mg Npf (**6**) and 25.41 mg Acd (**1**)

Acd was solubilized with 5 drops 10 M NaOH. The protein expression was then induced with IPTG. The culture was incubated at 37 °C for an additional 16 h. The cells were harvested at 5000 x g for 15 min, and the resulting pellet was suspended in 15 mL of 3-(N-morpholino) propanesulfonic acid (MOPS) resuspension buffer (50 mM MOPS, 100 mM KCl, 1 mM EDTA, pH 7.5). Following sonication, the cell lysate was allowed to cool on ice for 5 min. CaCl$_2$ was added to the sonicated lysate to a final concentration of 5 mM prior to centrifugation for 20 min at 30,000 x g, 4 °C.

CaM was purified from the cleared cell lysate using a phenyl-sepharose (PhS) CL-4B column with EDTA as eluent. Using a total resin bed volume of 10 mL, the column was first equilibrated with 4 column volumes of PhS Buffer A (50 mM Tris base, 1 mM CaCl$_2$, pH 7.5). After the cleared cell lysate was loaded and allowed to pass through the resin, the column was washed with 4 column volumes of PhS Buffer A, 4 column volumes of high-salt PhS Buffer B (50 mM Tris base, 0.5 M NaCl, 0.1 mM CaCl$_2$, pH 7.5), and an additional 2 column volume washes of PhS Buffer A to restore low-salt conditions. CaM was eluted with PhS Buffer C (10 mM Tris base, 10 mM EDTA, pH 7.5) and collected in 1 mL fractions. The presence of protein was detected by SDS-PAGE. Fractions containing protein were combined and dialyzed against water for 16 h at 4 °C.
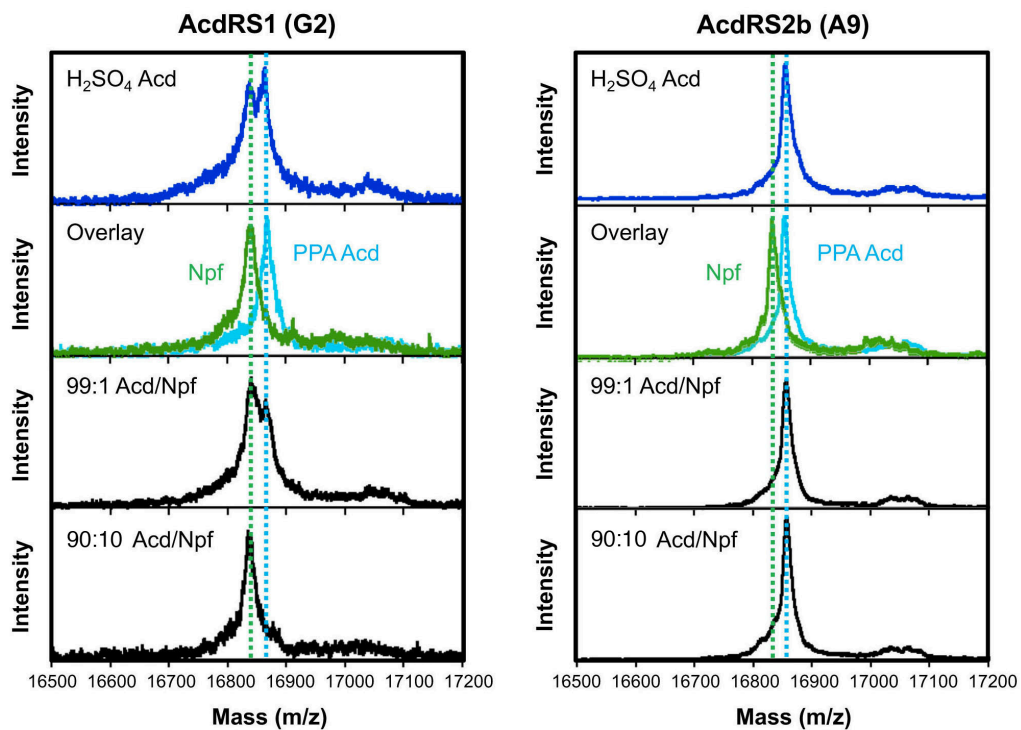
Fig. S2. *In Vivo* AcdRS Selectivity. CaM (UAG codon at 113) was expressed in minimal media containing 1 mM ncAA: Acd, synthesized either using the H_2SO_4 route or PPA route, Npf, or either a 99:1 or 90:1 mixture of PPA Acd and Npf. Expression was performed with either AcdRS1 (G2) or AcdRS2b (A9). AcdRS2b (A9) can tolerate at least 10% Npf in the growth media with no apparent Npf incorporation.

Acd Mutant α-Synuclein Protein Expression. For α-synuclein (αS) mutant expression, *E. coli* BL21-Gold (DE3) cells were transformed with the αS-F$_{94}$TAG plasmid and a pDule2 plasmid containing the AcdRS and tRNA$_{CUA}$ pair. Cells were selected for resistance to both Amp (100 µg/mL) and Strep (100 µg/mL) or Spec (50 µg/mL). Single colonies were used to inoculate 4 mL of LB media. The primary culture was grown at 37 °C with shaking at 250 rpm for 4 h. The 100 mL culture was incubated at 37 °C with shaking at 250 rpm until the OD$_{600}$ reached 0.8 AU. At this point, the different amounts of unnatural amino acid were added to the culture grown under conditions A-E as in the CaM protocol above. The cells were harvested at 5000 x g for 15 min and the resulting pellet was resuspended in 20 mM Tris, pH 8.0 with 1 mM PMSF and sonicated. Following sonication, the cell lysate was boiled for 15 min prior to centrifugation for 20 min at 13,200 x g, 4 °C. The cleared lysate was dialyzed overnight against 20 mM Tris, pH 8.0 at 4 °C prior to purified using a HiTrap Q HP column (GE Healthcare). FPLC fractions were dialyzed against 20 mM Tris, pH 8.0 and stored at 4 °C.
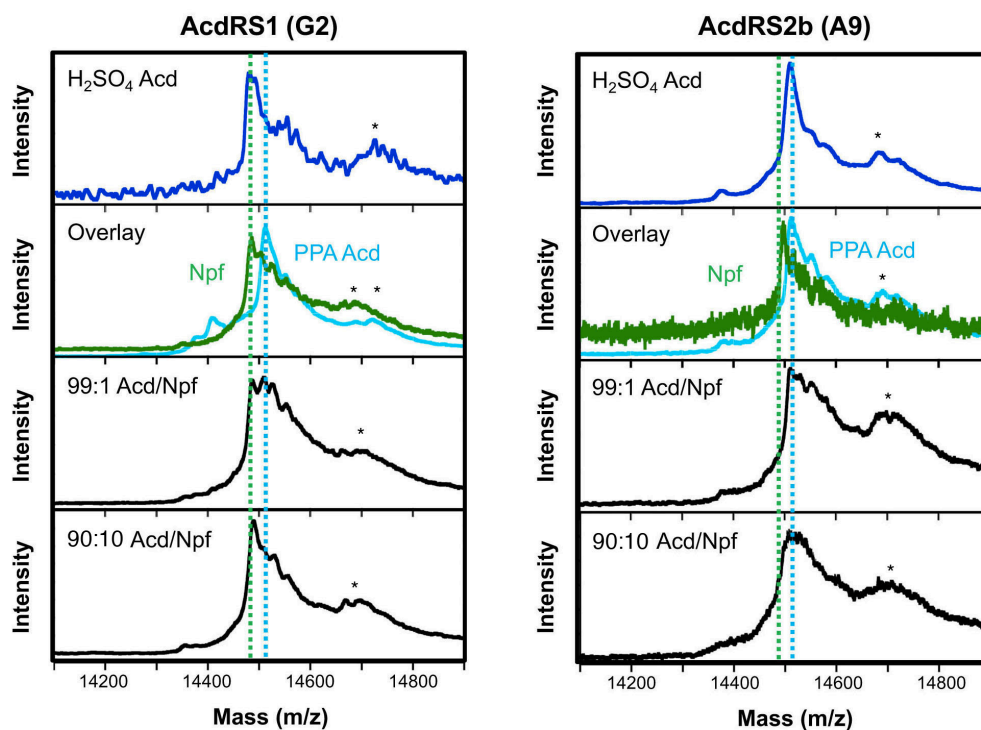
Fig. S3. *In Vivo* AcdRS Selectivity. αS (UAG codon at 94) was expressed in LB media containing 1 mM ncAA: Acd, synthesized either using the H_2SO_4 route or PPA route, Npf, or either a 99:1 or 90:1 mixture of PPA Acd and Npf. Expression was performed with either AcdRS1 (G2) or AcdRS2b (A9). AcdRS2b (A9) can tolerate at least 10% Npf in the growth media with no apparent Npf incorporation. * indicates MALDI matrix adduct. αSNpf$_{94}$, Calc'd: 14482; αSAcd$_{94}$, Calc'd: 14508.

Trypsin Digest Analysis of Acd Mutants. Protein Acd mutants were precipitated using 1:4 8.75 M trichloroacetic acid/protein sample and incubated at 4 °C for 15 minutes. The precipitate was centrifuged for 15 min at 13,200 rpm to pellet protein. The protein pellet was then washed three times with cold acetone to remove trace trichloroacetic acid. Trace acetone was removed by incubating protein pellets in a 95 °C water bath for 5 min open to the atmosphere. Protein pellets were then re-suspended in 6 M guanidinium hydrochloride with 50 mM Tris pH 8.0, and denatured by boiling at 95 °C for 10 minutes. Protein samples were then diluted to 0.75 M guanidinium hydrochloride with 50 mM Tris pH 7.6 and 1 mM calcium chloride. Sequencing grade modified trypsin (0.6 μg, Promega) was used to digest samples for 24 hours at 37 °C. Trypsin digest aliquots (1 μL) were combined with α-cyano-4- hydroxycinnamic acid (1 μL of a saturated solution in 1:1 H_2O/CH_3CN with 1 % TFA) and analyzed by MALDI-MS.

Peak intensity scaling to determine Acd/Npf selectivity was performed as follows: For PPA Acd expressions and Npf expressions, which should have only Acd or Npf in the $CaM_{108-116}$ trypsin fragment, a ratio of the peak for this fragment ($CaM_{108-116}Npf_{113}$: 1153.8 m/z, $CaM_{108-116}Acd_{113}$: 1179.8 m/z) was calculated relative to the peak for $CaM_{117-127}$ (1349.9 m/z). This ratio represensts the relative ease of ionization of these fragments. Similarly, the intensities of the peaks for the absent fragment were used to determine a level of background counts at this mass. These values were used to correct the data to give the scaled peak intensities in Table S2. The scaled intensities were then used to calculate an Acd/Npf MS incorporation selectivity factor using the following equation:

MS Sel = (Scaled Acd Peak/ Scaled Npf Peak)/(Acd/Npf ratio in media)

For each AcdRS, MS Sel values were averaged between 1% Npf and 10% Npf data sets. LexA and αS trypsin digests did not consistently show complete digestion, preventing such analysis.

Table S2. CaM 113 Tyrpsin Digest Intensity Scaling.

ncAAs	G2			A9		
	m/z	Peak Height	Scaled Peak	m/z	Peak Height	Scaled Peak
Acd (H_2SO_4)	1153.8	391	2277	1153.7	13	57.6
	1179.8	8166	1458	1179.7	5878	1083.1
	1349.9	5661		1349.8	1348	
Acd (PPA)	1153.7	24		1153.7	38	
	1179.8	7574		1179.7	15487	
	1349.9	1091		1349.8	4132	
Npf	1153.8	518		1153.7	90	
	1179.8	15		1179.5	10	
	1349.9	3500		1349.8	493	
1% Npf/Acd	1153.8	1224	7281	1153.4	17	73
	1179.8	20839	3801	1179.7	5386	984
	1349.9	7968		1349.8	1922	
10% Npf/Acd	1153.8	3034	18197	1153.5	22	121
	1179.8	754	15	1179.7	1621	294
	1350.0	10266		1349.8	790	

$CaM_{108-116}Acd_{113}$/ $CaM_{117-127}$ Scaling Factor: 5.47
$CaM_{108-116}Npf_{113}$/ $CaM_{117-127}$ Scaling Factor: 0.165

$CaM_{108-116}Npf_{113}$/ $CaM_{117-127}$ Background Factor: 0.0156
$CaM_{108-116}Acd_{113}$/ $CaM_{117-127}$ Background Factor: 0.0122

LexA Cloning/Expression/Purification. To create a C-terminal His-tagged LexA expression construct, a recombinant *E. coli lexA* gene containing the $S_{119}A$ catalytically-inactive mutation was PCR amplified with forward[3]

A: 5'-GGCAGCCATATGAAAGCGTTAACGG-3'

and reverse

B: 5'-AATCTCGAGCCAGTCGCCGTTGC-3'

primers. This amplified *lexA* gene was sub-cloned into the NdeI and XhoI sites in pET-41a(+) plasmid (EMD Millipore) to produce pET41-LexA-S_{119}A-HIS. To generate an amber stop codon in the S_{60} position, the pET41-LexA-S_{119}A-HIS plasmid was amplified with forward

C: 5'-ATTGTTTAGGGCGCATCACGCGGAATTCGTC-3'

and reverse

D: 5'-GTGATGCGCCCTAAACAATTTCAATAACGCC-3'

primers designed for site-directed mutagenesis with Phusion polymerase, creating pET41-LexA-S_{60}TAG-S119A-HIS.[4]

Chemically competent BLR(DE3) cells (EMD Millipore) were transformed with either of the above pET41-LexA plasmids (Kn resistance) and either of the pDule2-AcdRS1a or pDule2-AcdRS2b plasmids (streptomycin resistance) and selected on LB + Kn (30 µg/mL) + Strep (100 µg/mL) plates. Isolated colonies were grown overnight to saturation in liquid LB + Kn + Strep. Overnight cultures were inoculated at a 1:100 ratio into 25 mL MD-5051 auto-inducing medium containing 150 µg/mL Kn, 100 µg/mL Strep, and 1 mM acridone (H_2SO_4 route).[5] Cultures were grown for 24 hours in sterile 250 mL polypropylene centrifuge tubes (Corning) in a shaking 37 °C incubator. Cells were harvested at 4000 x g for 30 min at 4 °C.

Cell pellets were resuspended in 4 mL of wash buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.1) containing 1x BugBuster Protein Extraction Reagent (EMD Millipore), 0.5 mg/mL lysozyme, 25 U/µL benzonase, and 1x cOmplete EDTA-free protease inhibitor cocktail (Roche). After incubation at room temperature for 30 min, lysed cells were centrifuged at 20000 x g for 30 min at 4 °C. LexA proteins were purified from the clarified cell lysate using 0.5 mL HisPur cobalt spin columns per manufacturer's instructions (Thermo Scientific), using 300 mM imidazole as eluent. LexA proteins were detected by running purification samples on 15% SDS-PAGE gels. Elutions containing LexA were pooled and buffer-exchanged into low salt buffer (20 mM Tris-HCl, 25 mM NaCl, pH 7.6) using Vivaspin 500 concentration devices.
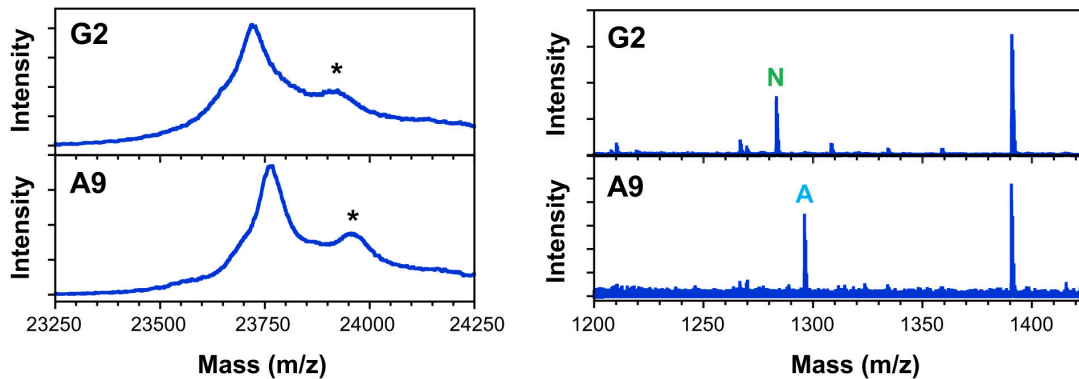

Fig. S4. *In Vivo* AcdRS Selectivity. LexA (UAG codon at 60) was expressed in auto-induction media containing 1 mM Acd, synthesized using the H_2SO_4 route. Expression was performed with either AcdRS1 (G2) or AcdRS2b (A9). Although the whole protein peaks (left) are very broad, trypsin digest producing the $LexA_{53-64}$ fragment (right) indicates the presence of only Npf in the G2 sample and only Acd in the A9 sample. N denotes $LexA_{53-64}Npf_{60}$. A denotes $LexA_{53-64}Acd_{60}$. * indicates MALDI matrix adduct. $LexA_{53-64}Npf_{60}$, Calc'd: 1366.5, Obsv'd: 1366.6; $LexA_{53-64}Acd_{60}$, Calc'd: 1392.5, Obsv'd: 1392.4

Cloning and Purification of Synthetases

AcdRS Cloning/Expression/Purification. Synthetic genes (IDT) for AcdRS1 and AcdRS2b were amplified with forward

    E: 5'-GAGCGGATAACAATTCCCCTCTAG-3'

and reverse

    F: 5'-GTGGTGGTGCTCGAGTCTCTTTC-3'

primers. The resulting PCR product was digested with NdeI and XhoI and sub-cloned into a pBAD24 plasmid modified with a C-terminal His tag, resulting in the three pBAD24-AcdRS-HIS constructs. Chemically competent BL21(DE3) cells were transformed with one of the three above pBAD24-AcdRS-HIS plasmids (Amp resistance) and selected on LB + Amp (100 µg/mL) plates. Isolated colonies were inoculated into 200 mL arabinose auto-inducing medium containing 100 µg/mL carbenicillin (Carb).[6] Cultures were grown for 24 hours in 2 L flasks in a shaking 37° incubator. Cells were harvested at 4000g for 15 min at 4°.

Cell pellets were resuspended in 30 mL of wash buffer (50 mM HEPES-KOH, 500 mM NaCl, 10 mM imidazole, pH 7.9) containing 1 mM PMSF and 1x cOmplete EDTA-free protease inhibitor cocktail (Roche). Cell suspensions were lysed using a French pressure cell (FA-032, Thermo Electron Corp) and then centrifuged at 20000g for 30 min at 4 °C (Sorvall RC 6+). Clarified lysates were loaded with a BioLogic LP System (Bio-Rad) onto a column containing 4 mL of packed HisPur resin (Thermo Scientific), and proteins were eluted with a linear imidazole gradient. AcdRS proteins were detected by running elution fractions on 12% SDS-PAGE gels. Elutions containing AcdRS were pooled and dialyzed three times against 2 L of 20 mM Tris-HCl pH 8.5, 50 mM NaCl, 10 mM 2-mercaptoethanol, and 20% glycerol. Dialyzed samples were loaded onto a 5 mL HiTrap Heparin HP column (GE Healthcare) and proteins were separated

with a linear NaCl gradient. FPLC fractions were analyzed by 12% SDS-PAGE, and those containing AcdRS proteins were concentrated using Amicon Ultra centrifugal filter units (Sigma), buffer-exchanged (50 mM HEPES-KOH pH 7.5, 20 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, and 50% glycerol), and stored at -80 °C.

AcdRS Activity Test. The activity of each AcdRS enzyme was verified by monitoring pyrophosphate release during the initial condensation reaction of Acd or Npf with ATP. To measure pyrophosphate, an enzyme-coupled, colorimetric assay recently described in the literature was used.[7] Briefly, pyrophosphate generated during each reaction is hydrolyzed by excess inorganic pyrophosphatase (PPiase), and free orthophosphate is incubated with malachite green under acidic conditions (Malachite Green Phosphate Assay Kit, Sigma) to generate a malachite green phosphomolybdate complex (600-660 nm absorbance). Several concentrations of each AcdRS enzyme (0, 0.5, 1.0, and 2.0 μM) were incubated in reaction buffer (50 mM HEPES-KOH pH 7.5, 20 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT) with 2 U/ml PPiase (NEB), 1 mM of either Acd or Npf, and with or without 0.2 mM ATP for 35 min at 37 °C. Reactions were quenched with chilled EDTA to a final concentration of 10 mM. Samples were developed per manufacturer's instructions in a clear 96-well plate (Corning Costar), and absorbances at 620 nm were read in an Infinite F200 plate reader (Tecan). Standard curves were used to interpolate observed phosphate concentrations, and the data describing calculated pyrophosphate release rate versus AcdRS concentration were fit to a straight line.
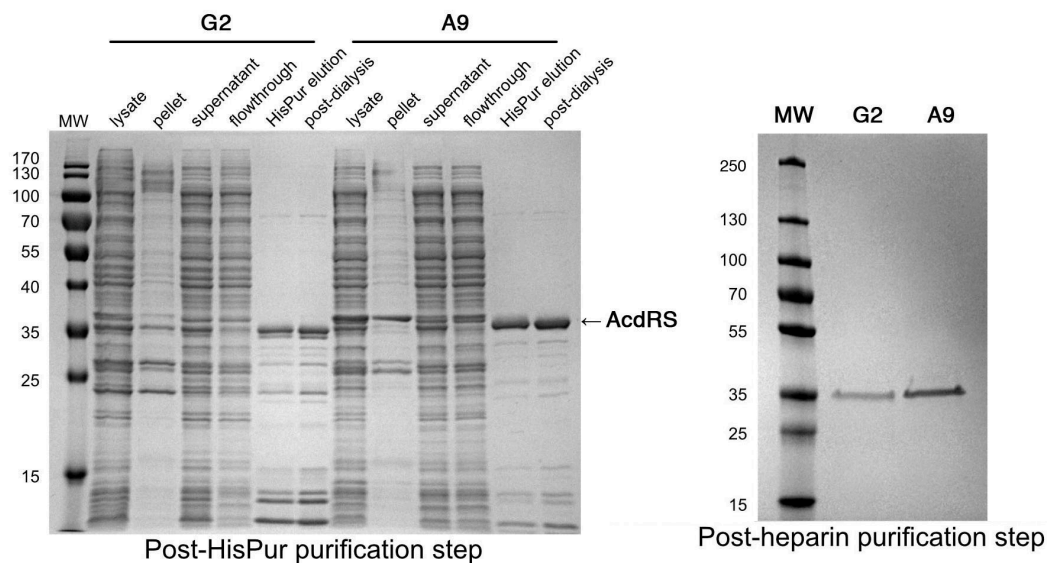
Fig. S5. AcdRS Purification. Left: SDS-PAGE gel showing purification samples of AcdRS1 (G2, lanes 2-7) and AcdRS2b (A9, lanes 8-13) taken throughout the cell lysis and HisPur affinity resin steps. Right: SDS-PAGE gel showing final, purified samples of both AcdRS enzymes following the purification scheme described in the supplemental text. For the lanes in each gel, an equal volume of sample and Laemmli buffer (Bio-Rad) were mixed, boiled at 95°, centrifuged, and loaded onto 12% SDS-PAGE gels, which were run for 50 min at 200 V.
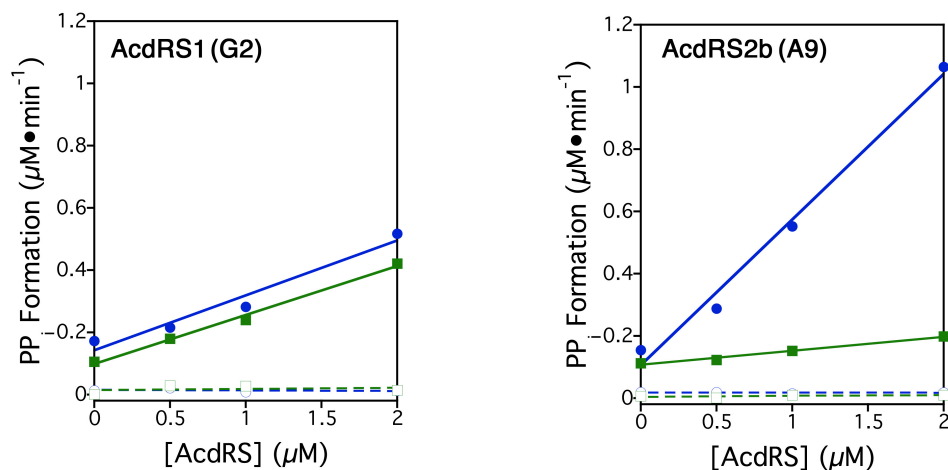

Fig. S6. AcdRS Malachite Green Activity Assay. Plots showing rate of inorganic pyrophosphate formation versus AcdRS concentration for AcdRS1 (G2, left) and AcdRS2b (A9, right). Data are from reactions at 37° containing either 1 mM Acd (blue) or 1 mM Npf (green) and in the presence (solid points) or absence (open points) of 0.2 mM ATP. The linear dependence of activity on AcdRS concentration was demonstrated by fitting a straight line through the data points.

In Vitro **Aminoacylation Analysis**

Preparation and Labeling of tRNA Transcripts for Aminoacylation. The tRNA substrate for aminoacylation by the AcdRS1 (G2) and AcdRS2b (A9) isolates was prepared by *in vitro* transcription with T7 RNA polymerase,[8, 9] using DNA templates encoding a self-cleaving hammerhead ribozyme immediately upstream of the T7 promoter.[10] DNA templates were assembled from four overlapping synthetic oligonucleotides, listed below (overlapping regions are underlined). The tRNA[Tyr] species used is the orthogonalized tRNA (Orthog-tRNA) isolated by genetic selections when the *Mj* TyrRS-tRNA[Tyr] system was first developed for ncAA incorporation.[11]

   G: 5'-AATTCCTGCAGTAATACGACTCACTATA<u>GGGAGACCGGCTGATGAGTC</u>-3'

   H: 5'-<u>CCGGGACGGTACCGGG</u>TACCGTTTCGTCCTCACG<u>GACTCATCAGCCGGTCTCCC</u>-3'

   I: 5'-<u>CCCGGTACCGTCCCGGC</u>GGTAGTTCAGCCT<u>GGTAGAACGGCGGACTGTAG</u>-3'

   J: 5'-TGGTCCGGCGGGCCGGATTTGAACCAGCGACATGCGGAT<u>CTACAGTCCGCCGTTCTACC</u>-3'

Oligonucleotide assembly by PCR was performed as described.[9] A fifth oligonucleotide was used, along with oligonucleotide **G**, to amplify the assembled template:

   K: 5'-[2'-OMe]U[2'-OMe]GGTCCGGCGGGCCGG-3'

The 2'-*O*-methyl modifications inhibit run-over transcription by T7 RNA polymerase.[12] Amplification of the full-length template, template purification, and transcription reactions were performed as described.[9] The enzyme for transcription was the T7 RNA polymerase$_{\Delta172\text{-}173}$ variant, which was expressed and purified in the laboratory as described.[13] The transcription reaction mixture was incubated at 37 °C for 16-24 h.

     Cleavage of the hammerhead ribozyme was performed as described.[9] After the reaction, the mature tRNA was purified from uncleaved transcripts and free hammerhead ribozyme by

electrophoresis through a denaturing 15 cm gel containing 15% polyacrylamide:bisacrylamide (29:1), 8 M urea and 1X TBE. tRNA was visualized by UV shadowing and excised from the gel. tRNA was extracted from the gel band with 10 volumes of TE6 buffer (10 mM BisTris-HCl and 1 mM EDTA, pH 6.0), shaking for 16-24 h at room temperature. Insoluble gel debris were removed by centrifugation at 5000 rcf for 10 min. The supernatant was passed though a 0.45 μm filter and washed with TE6 buffer using a centrifugal filtration device (Amicon; 10 kDa MWCO) such that urea was diluted to a concentration below 1 μM.

The tRNA substrates for aminoacylation were radiolabeled at the 3' internucleotide linkage in accordance with established methods,[9, 14] using *E. coli* nucleotidyltransferase, purified in the Perona laboratory as described.[15] Prior to labeling, tRNA was denatured by incubating for 10 min at 65° C in ~50 μL water, and refolded by slow-cooling to room temperature over the course of one hour.[9] Labeled tRNA was purified by phenol:chloroform:isoamyl alcohol extraction (Sigma; 25:24:1; pH 7.8) followed by polyacylamide gel electrophoresis. tRNA was then extracted from the gel band with 2 volumes of TE6 buffer without shaking, in a 16-24 h incubation at ambient temperature.[9]

Aminoacylation Reaction Assays. Aminoacylation assays were conducted at 37° C in siliconized conical tubes containing 20 ul of 50 mM HEPES KOH (pH 7.5), 20 mM NaCl, 10 mM MgCl$_2$, 1 mM DTT, 2.5 mM MgATP and various concentrations of amino acid. All reactions were conducted under single-turnover conditions, and control experiments were performed to ascertain saturating concentrations for tRNA. For every combination of enzyme (AcdRS1 or AcdRS2b) and amino acid (Acd or Npf), these experiments established that 1.0 μM enzyme with 0.1 μM tRNA allowed full tRNA binding. These concentrations were therefore employed in all measurements.

Methods used were similar to those employed in the Perona laboratory for the study of other RSs.[9, 16] tRNAs were refolded in the presence of 33 mM $MgCl_2$ prior to aminoacylation, with refolding performed as described.[9] Time points in aminoacylation assays were quenched by diluting 1:3 into a solution containing 200 mM NaOAc (pH 5.2), 0.2% SDS, and subsequently digested for 15 minutes at room temperature with *Penicillium citrinum* P1 nuclease (Sigma) at a concentration of 0.0125 U/μL.[17] Free nucleotides were separated by thin layer chromatography across 10 cm PEI cellulose sheets (Sigma) with a solvent containing 1 M ammonium acetate and 5% acetic acid. Spots corresponding to AMP and aminoacyl-AMP were visualized by phosphorimaging and quantified with the software package *ImageQuant* 5.2. These data were used to determine single-turnover rate constants for a given AcdRS and amino concentration as shown in Fig S7. To extract $k_{obs,}$ the timecourses were fit to the single exponential equation $Y = A_1 \cdot e^{-k_{obs} \cdot t}$. To derive K_d, k_{obs} was determined at several amino acid concentrations and the plot of k_{obs} versus initial amino acid concentration (S_0) was fit to the hyperbolic binding curve $Y = S_0 * k_{chem}/S_0 + K_d$. Prism 5 was used for all data fitting.
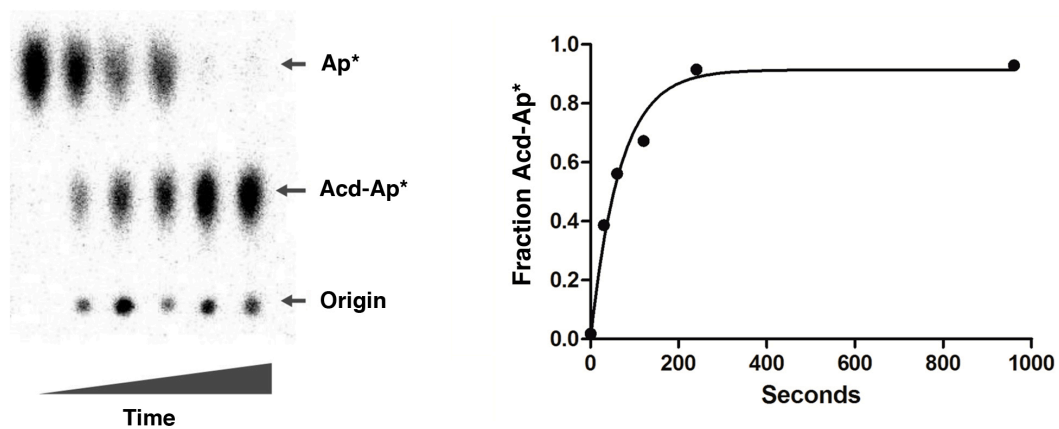
Fig. S7. Representative primary data for aminoacylation. Left: TLC plate showing reactions of the AcdRS2b(A9) enzyme with saturating levels of orthog-tRNA and Acd at varying times. There is almost complete aminoacylation by the end of the timecourse (about 5 min). Right: Timecourse for the AcdRS2b(A9) reaction fit to derive k_{obs} (see the text).

Computational Modeling

Geometry Optimization of Ligands. Structures of Acd and Npf were constructed in ChemDraw and saved in Molfile format. Acd and Npf Molfiles were submitted to the PRODRG webserver for energy minimization.[18] Coordinates of Npf were further optimized using density functional theory (DFT) with the B3LYP hybrid functional and the 6-311+G** basis set in Gaussian09.[19] The UltraFine integration grid was used with tight convergence criteria.

Molecular Mechanics Parameterization of Ligands. Force field parameters and partial charges for Acd and Npf were obtained from the CGenFF webserver.[20-22] Bond orders were imposed based upon the connectivity. The partial charges of the main chain and beta carbon atoms were changed to those of the free tyrosine amino acid in CHARMM22.[23] An adjustment of ±0.002 was made to the partial charge of the gamma carbon to make the molecule entirely neutral.

Ligand Modeling in AcdRS1 (G2). Atomic coordinates for G2 were obtained from the PDB structure 4PBR. Amino acid sidechain coordinates for sites with partial occupancy were taken from the A conformations. Alignments by superimposition of the first benzyl rings of Acd and Npf with the Brb benzyl ring were carried out in PyMol,[24] and atomic coordinates for the equivalent amino acid atoms of the ligands were adjusted to those of Brb. PSF files and PDB structure files with hydrogen atoms were generated using the AutomaticPSF builder tool in VMD.[25] 10,000 steps of energy minimization of ligand atoms and all hydrogen atoms were carried out using the method of conjugated gradients in NAMD.[26] The CHARMM22 force field was used for all protein atoms, and the CGenFF force field for all interactions involving the ligands.

Generation of the AcdRS2b (A9) Structure. The statistical, computationally assisted design strategy[27] was used to determine atomic coordinates for the sidechains of mutated residues.

Atomic coordinates of the protein backbone, wild-type residues and the Acd ligand, along with hydrogen atoms, were obtained from energy-minimized structure of Acd within G2. Alllowed conformations of the mutated resideus were those in a rotamer library.[28] Sidechain bond angles and bond lengths used were the average values collated from the HiQ54 dataset of crystal structures.[29] Sidechain modeling conformations were performed with an effective inverse temperature of $\beta = 0.5$ mol/kcal. This procedure yielded a model of A9 having the most probable sidechain conformation of each mutated amino acid.

Ligand Modeling in AcdRS2b (A9). Atomic coordinates for Acd and Npf were obtained from the minimization calculations in G2. PSF files and PDB structure files with hydrogen atoms were generated using the AutomaticPSF builder tool in VMD. 10,000 steps of energy minimization of ligand atoms, mutated amino acid sidechains, and all hydrogen atoms were carried out using the conjugate gradient method in NAMD. The CHARMM22 force field was used for all protein atoms, and the CGenFF force field for all interactions involving the ligands.
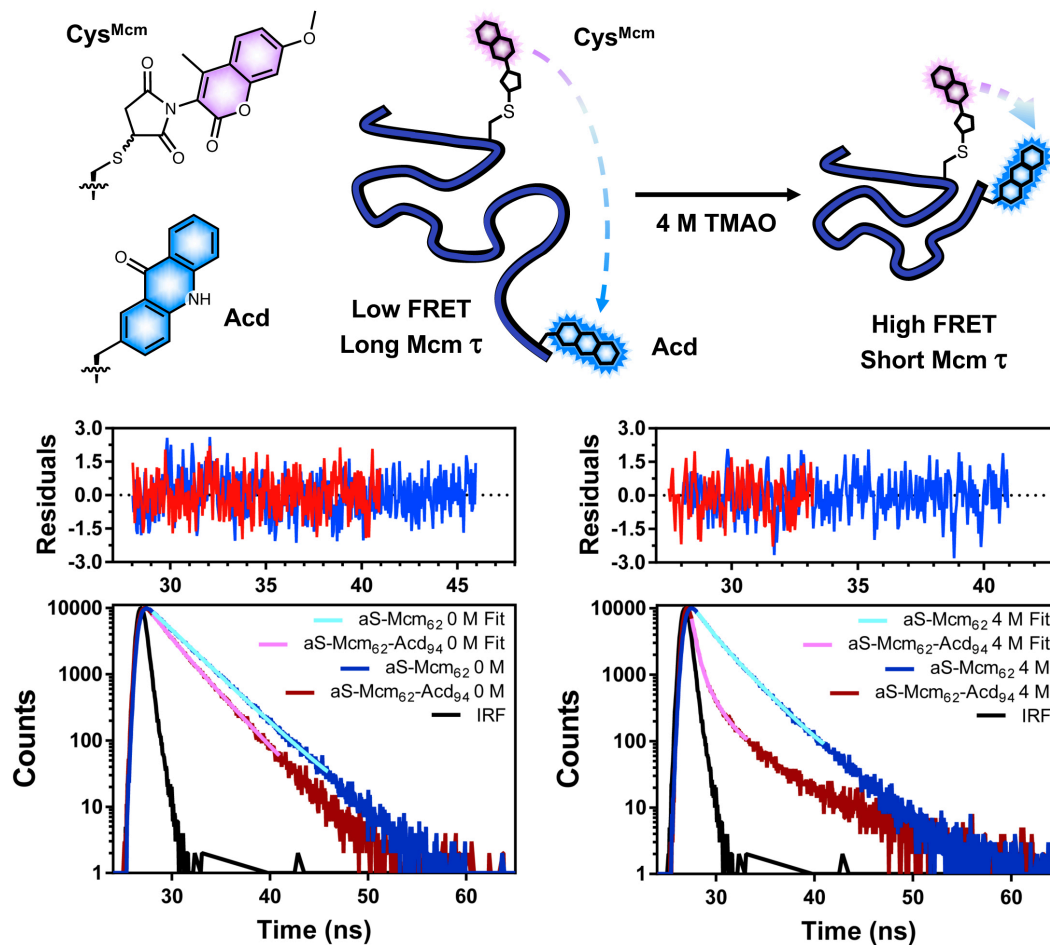
Fig. S8. Monitoring a Conformational Change in αS with a Methoxycoumarin(Mcm)/Acd FRET Pair. A Cys$_{62}$Acd$_{94}$ αS mutant was expressed and labeled with Mcm-maleimide to give αS- Cys$^{Mcm}_{62}$Acd$_{94}$. Mcm/Acd FRET was measured in buffer (0 M) and in a solution of 4 M trimethylamine *N*-oxide (TMAO), which causes the protein to compact. The Mcm fluorescence lifetime (τ) was measured with excitation at 340 nm and emission at 380 nm using a PTI Quantamaster 40 fluorometer with time correlated single photon counting detection. A decrease in the the Mcm lifetime indicates increased FRET with Acd.
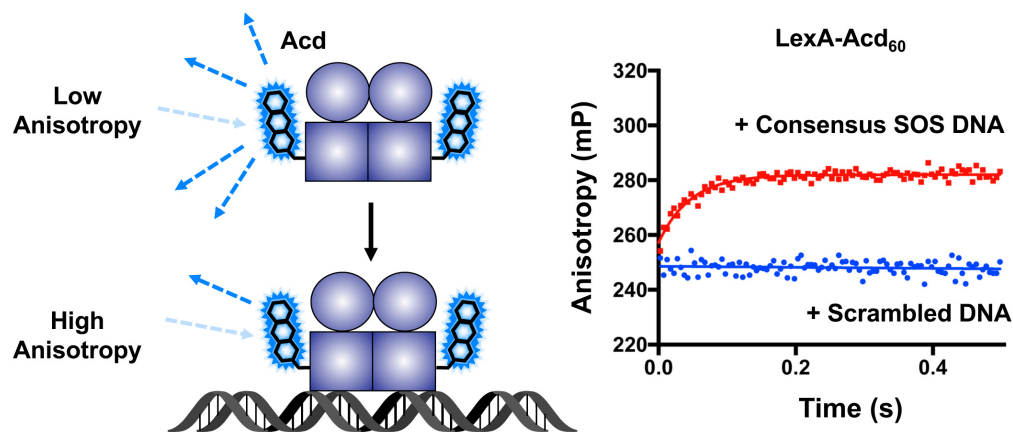
Fig. S9. Measuring the Kinetics of LexA-DNA Binding Using Rapid Mixing and Fluorescence Anisotropy. LexA-Acd$_{60}$ was rapidly mixed with double-stranded 44-mer DNA substrates that contained either the "consensus" LexA binding site or a "scrambled" LexA binding site using a KinTek Auto SF-120 stopped-flow instrument to monitor Acd fluorescence anisotropy. An increase in anisotropy was observed upon specific binding of LexA-Acd$_{60}$ to consensus operator DNA (but not scrambled DNA). Polarized light at 386 nm was used to excite the mixed sample, and emitted signals in parallel and perpendicular polarized directions were measured at 440 nm. Anisotropy values were calculated automatically using the KinTek Auto SF-120 instrument software. Five replicates were collected, timepoints were averaged, and the data were fit to single exponential (for the consensus DNA) or straight line (for the scrambled DNA) models.

References

1. L. C. Speight, A. K. Muthusamy, J. M. Goldberg, J. B. Warner, R. F. Wissner, T. S. Willi, B. F. Woodman, R. A. Mehl and E. J. Petersson, *J. Am. Chem. Soc.*, 2013, **135**, 18806-18814.

2. R. J. Blizzard, D. R. Backus, W. Brown, C. G. Bazewicz, Y. Li and R. A. Mehl, *J. Am. Chem. Soc.*, 2015, **137**, 10044-10047.

3. C. Y. Mo, S. A. Manning, M. Roggiani, M. J. Culyba, A. N. Samuels, P. D. Sniegowski, M. Goulian and R. M. Kohli, *Msphere*, 2016, **1**.

4. L. Zheng, U. Baumann and J.-L. Reymond, *Nucleic Acids Res.*, 2004, **32**, e115-e115.

5. F. W. Studier, in *Structural Genomics: General Applications*, ed. Y. W. Chen, Humana Press, Totowa, NJ, 2014, DOI: 10.1007/978-1-62703-691-7_2, pp. 17-32.

6. J. T. Hammill, S. Miyake-Stoner, J. L. Hazen, J. C. Jackson and R. A. Mehl, *Nat. Protocols*, 2007, **2**, 2601-2607.

7. I. Cestari and K. Stuart, *J. Biomol. Screen.*, 2013, **18**, 490-497.

8. L. D. Sherlin, T. L. Bullock, T. A. Nissan, J. J. Perona, F. J. Lariviere, O. C. Uhlenbeck and S. A. Scaringe, *RNA*, 2001, **7**, 1671-1678.

9. B. J. Rauch, J. J. Porter, R. A. Mehl and J. J. Perona, *Biochemistry*, 2016, **55**, 618-628.

10. S. R. Price, N. Ito, C. Oubridge, J. M. Avis and K. Nagai, *J. Mol. Biol.*, 1995, **249**, 398-408.

11. L. Wang, A. Brock, B. Herberich and P. G. Schultz, *Science*, 2001, **292**, 498-500.

12. C. Kao, M. Zheng and S. Rudisser, *RNA*, 1999, **5**, 1268-1272.

13. D. L. Lyakhov, B. He, X. Zhang, F. W. Studier, J. J. Dunn and W. T. McAllister, *J. Mol. Biol.*, 1997, **269**, 28-40.

14. A. D. Wolfson and O. C. Uhlenbeck, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 5965-5970.

15. R. G. McGann and M. P. Deutscher, *Eur. J. Biochem.*, 1980, **106**, 321-328.

16. A. Rodriguez-Hernandez and J. J. Perona, *Structure*, 2011, **19**, 386-396.

17. A. Rodriguez-Hernandez, J. L. Spears, K. W. Gaston, P. A. Limbach, H. Gamper, Y.-M. Hou, R. Kaiser, P. F. Agris and J. J. Perona, *J. Mol. Biol.*, 2013, **425**, 3888-3906.

18. A. W. Schuttelkopf and D. M. F. van Aalten, *Acta Crys. D*, 2004, **60**, 1355-1363.

19. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Journal*, 2016.

20. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. MacKerell, *J. Comput. Chem.*, 2010, **31**, 671-690.

21. K. Vanommeslaeghe and A. D. MacKerell, *J. Chem. Inf. Model.*, 2012, **52**, 3144-3154.

22. K. Vanommeslaeghe, E. P. Raman and A. D. MacKerell, *J. Chem. Inf. Model.*, 2012, **52**, 3155-3168.

23. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586-3616.

24. Schrödinger, The PyMOL Molecular Graphics System, Version 1.3, 2015

25. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33-38.

26. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781-1802.

27. H. Kono and J. G. Saven, *J. Mol. Biol.*, 2001, **306**, 607-628.

28. M. V. Shapovalov and R. L. Dunbrack, *Structure*, 2011, **19**, 844-858.

29. A. Leaver-Fay, M. J. O'Meara, M. Tyka, R. Jacak, Y. F. Song, E. H. Kellogg, J. Thompson, I. W. Davis, R. A. Pache, S. Lyskov, J. J. Gray, T. Kortemme, J. S. Richardson, J. J. Havranek, J. Snoeyink, D. Baker and B. Kuhlman, in *Methods in Protein Design*, ed. A. E. Keating, Elsevier Academic Press Inc, San Diego, 2013, vol. 523, pp. 109-143.