Published in final edited form as:

Org Biomol Chem. 2017 May 03; 15(17): 3603–3610. doi:10.1039/c7ob00582b.

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,*}

^aDepartment of Chemistry, University of Pennsylvania, 213 South 34th Street, Philadelphia, PA 19104, USA

^bDepartment of Medicine, Perelman School of Medicine, University of Pennsylvania, 3610 Hamilton Walk, Philadelphia, PA 19104

^cOregon State University, Department of Biochemistry and Biophysics, 2011 Ag Life Sciences Building, Corvallis, Oregon 97331, USA

^dDepartment of Chemistry, Portland State University, P.O. Box 751, Portland, Oregon 97207, USA

^eDepartment of Biochemistry & Molecular Biology, Oregon Health & Sciences University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97239, USA

Abstract

The amino acid acridon-2-ylalanine (Acd) can be a valuable probe of protein dynamics, either alone or as part of a Förster resonance energy transfer (FRET) or photo-induced electron transfer (eT) probe pair. We have previously reported the genetic incorporation of Acd by an aminoacyl tRNA synthetase (RS). However, this RS, developed from a library of permissive RSs, also incorporates *N*-phenyl-amino-phenylalanine (Npf), a trace byproduct of one Acd synthetic route. We have performed negative selections in the presence of Npf and analyzed the selectivity of the resulting AcdRSs by *in vivo* protein expression and detailed kinetic analyses of the purified RSs. We find that selection conferred a ~50-fold increase in selectivity for Acd over Npf, eliminating incorporation of Npf contaminants, and allowing one to use a high yielding Acd synthetic route for improved overall expression of Acd-containing proteins. More generally, our report also provides a cautionary tale on the use of permissive RSs, as well as a strategy for improving selectivity for the target amino acid.

E. James Petersson: ejpetersson@sas.upenn.edu; Tel: 1-215-746-2221.

[†]These authors contributed equally to this work.

Author Contributions

I.S. and L.C.S. synthesized amino acids and performed *in vivo* protein expression and MALDI MS analysis. Z.M.H. cloned and expressed LexA and AcdRSs. R.J.B. and J.J.Po. performed AcdRS selections. C.M.D. and L.Z.M. performed AcdRS kinetic measurements. J.A.V. generated computational models. J.G.S. supervised J.A.V.; J.J.Pe. supervised. C.M.D. and L.Z.M.; R.M.K. supervised Z.M.H.; R.A.M. supervised R.J.B. and J.J.Po.; and E.J.P. supervised I.S. and L.C.S. All authors contributed to the writing of the manuscript.

Electronic Supplementary Information (ESI) available: Procedures for the expression and characterization of proteins, analysis of enzyme kinetics, and structural modeling. See DOI: 10.1039/x0xx00000x

TOC image

Introduction

It is now well-established that protein folding and dynamics play essential roles in health and disease. For example, the small protein calmodulin (CaM) undergoes a dramatic conformational rearrangement to carry out its calcium sensor function in eukaryotic cells.^{1, 2} In a second example, the bacterial repressor-protease LexA uses a complex sequence of RecA-induced structural change, self-proteolysis, and dissociation of subunits to sense DNA damage and activate genes that ultimately lead to antibiotic resistance.^{3, 4} Finally, the conformational flexibility of the neuronal protein α -synuclein (α S) is a liability, as it leads α S to misfold and form amyloid fibrils that contribute to the pathogenesis of Parkinson's disease.^{5, 6} Fluorescence spectroscopy is a powerful tool for studying such processes, as it allows one to observe protein motions in real time under physiological conditions, including measurements in live cells.^{7, 8} One can even obtain low resolution structural information using distance-dependent chromophore interactions such as Förster resonance energy transfer (FRET) and quenching by photo-induced electron transfer (eT).⁹ To appropriately model protein motions, one needs a set of probes that are capable of accurately reporting on distance changes without disrupting the fold and function of the protein of interest.¹⁰

Recent developments in genetic code expansion and biorthogonal chemistry have made the site-specific incorporation of non-canonical amino acids (ncAAs) and installation of fluorophores through post-translational modification straightforward, even in cells or lysates.¹¹ Unfortunately, the fluorophores used in these approaches are often relatively large and are attached by long flexible linkers, which have a non-trivial impact on the observed FRET measurements. Additionally, larger fluorophores cannot be introduced co-translationally to allow packing into the interior of a folded protein. Thus, they will be restricted to surface-accessible positions, limiting the regions of the protein for which conformational changes can be studied.

Smaller probes that are more closely tied to the backbone are better able to report on conformational changes of the protein. The Petersson laboratory has developed small fluorescent probes and quenchers that should be non-perturbing to proteins and which are closely tied to the protein backbone. These include thioamide substitutions of the backbone itself and intrinsically fluorescent ncAAs such as acridon-2-ylalanine (Acridon-2-ylalanine, Acridon-2-ylalanine, Acridon-2-ylalanine) or **1**.¹² These new approaches are significant because, of the >100 ncAAs that have been genetically encoded in *E. coli*, only four are fluorescent.^{12–19}

Acid is a blue-wavelength fluorescent amino acid that is a useful fluorophore because of its small size (222 Å³), near unity quantum yield in water ($\phi = 0.95$), unusually long lifetime ($\tau \sim 15$ ns) and high photostability (<5% degradation after 3 h irradiation).^{20–22} Previous work in the Petersson laboratory has shown that Acid can be efficiently quenched by a thioamide through an eT mechanism.²³ We have also shown that it can be a valuable intermolecular FRET acceptor from Trp or methoxycoumarin, and a donor to more red-shifted dyes such as nitrobenzoxadiazole (NBD) or fluorescein.¹² Further examples of the utility of Acid in intramolecular FRET with Mcm and in fluorescence polarization studies of protein binding are shown in Figs. S8 and S9. (see ESI)

Previously, the Mehl and Petersson laboratories developed an *in vivo* system for Acid incorporation using methods pioneered by Schultz.²⁴ These methods require the generation of an aminoacyl tRNA synthetase (RS) that is selective for the ncAA and a tRNA that can be selectively charged by the ncAARS to deliver the tRNA to an unassigned codon, typically the amber stop codon UAG (tRNA_{CUA}).²⁵ An RS was selected from a library of permissive *M. janaschii* (*Mj*) tyrosyl RS mutants that had previously been shown to incorporate bulky aromatic amino acids such as 4-(2'-bromoisobutyramido) phenylalanine (Brb) and *p*-benzoyl phenylalanine.^{26–28} The most active mutant from this library (G2) was used to express Acid-containing variants of CaM, triose phosphate isomerase, and α S.¹² Here, we will refer to this mutant as AcdRS1.

At the same time, we also developed an efficient synthetic route, starting from the natural amino acid Tyr (**2**) and using a Buchwald-Hartwig coupling to *O*-methyl anthranilate (**4**) to form **5**.¹² (See Scheme 1) We obtained an 86% overall yield from Tyr using the route shown in Scheme 1 with H₂SO₄ used in the Friedel-Crafts cyclization step (no racemization observed by high performance liquid chromatography, HPLC). Unfortunately, this route also produced a decarbonylated by-product, *N*-phenyl-aminophenyl alanine (Npf or **6**) in trace amounts. We found that Npf was in fact incorporated much more efficiently by AcdRS1 than Acid. We were able to initially address this problem by converting **5** to Acid using LiOH deprotection followed by PPA cyclization. This eliminated Npf formation, but limited us to a 44% overall yield from Tyr with complete racemization (a 22% yield of the requisite L-Acid form). While we were able to express proteins containing exclusively Acid, this was not a very satisfactory solution, and we sought to obtain an AcdRS that was sufficiently selective against Npf that we could use the higher yielding H₂SO₄ synthetic route.

A typical selection experiment to evolve RSs for ncAAs consists of rounds of positive selection (performed in the presence of the ncAA and the 20 canonical amino acids) and negative selection (performed in the presence of only the 20 canonical amino acids). In this work, we sought to obtain an Acid-selective RS by performing these standard selections, as well as negative selections where Npf was included in the selection media to eliminate those AcdRS mutants charging tRNA with Npf. This resulted in two additional RSs, AcdRS2a (clone G11, from traditional negative selection) and AcdRS2b (clone A9, from Npf counter-selection), which both showed good selectivity in an initial screen in which the ncAAs were incorporated into green fluorescent protein (GFP).

We have performed detailed studies of their *in vivo* selectivities in expressions of CaM, α S, and LexA and found that AcdRS2b has superior selectivity for Acd when compared to AcdRS2a. We have also expressed and purified AcdRS1 and AcdRS2b to measure their Acd and Npf activation kinetics. We are able to rationalize their selectivities in terms of the X-ray crystal structure of AcdRS1 and a homology model of AcdRS2b. Our study provides an improved Acd incorporation method for fluorescent labeling of proteins, and it also validates a general strategy for how one may optimize a permissive RS to eliminate incorporation of an unwanted contaminant.

Results and Discussion

CaM, LexA and α S constructs with UAG mutations were expressed in *E. coli* along with plasmids encoding AcdRS1 and its cognate tRNA_{CUA} species. We analyzed the selectivity of AcdRS1 based on matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS) data of both intact and trypsin-digested, purified proteins, including CaM, LexA, and α S. Data for incorporation at position 113 in CaM are shown in Fig. 2 and 3; additional data for LexA and α S are shown in the ESI. In all cases, when the proteins are expressed using media containing pure Npf or Acd produced using the PPA route, a single peak for the intended product is obtained in the MALDI spectra. When the proteins are expressed in media containing Acd produced using the H₂SO₄ route, we observe a roughly 45:55 Npf/Acd ratio of the two CaM species, even though Npf is present only in trace quantities. We also considered the possibility that Npf could be generated *in vivo* by several possible routes. For example, a carboxy-lyase such as YigC could convert **7** (also a <1% contaminant in the H₂SO₄ Acd synthesis) into Npf,²⁹ or some fraction of Acd could be converted to Npf. Our PPA Acd expression data allow us to exclude metabolic processing of Acd, since we see no Npf incorporation in this case. When we use media in which we intentionally include 1% or 10% Npf with Acd, we detect products containing Npf/Acd ratios of 65:35 and >95:5, respectively. (ESI, Fig. S2) These data and similar data for other proteins (see ESI) indicated to us that the selectivity of our nominal AcdRS—designed to incorporate Acd specifically—in fact favored Npf incorporation by approximately 100-fold. Despite the incongruity between its intended use and its actual fidelity, one should keep in mind that AcdRS1 was selected based on expression yields of GFP in the presence of media containing Acd from the H₂SO₄ route (and therefore containing ~1% Npf). Similar contamination, barely detectable by HPLC analysis, may be present in other stocks of home-synthesized or commercial batches of ncAAs used in RS selection.

AcdRS Optimization

RS Selection to Remove Npf Activity—To reduce the incorporation of Npf, we screened mutant *Mj*TyrRS libraries with positions in the amino acid binding pocket randomized using a GFP expression screen common to the Mehl laboratory.^{27, 30} Two rounds of positive and negative selection were performed according to standard protocols, with the desired ncAA (i.e., Acd) present in the media for rounds of positive selection and only the 20 canonical amino acids present in the media for rounds of negative selection. In parallel, we also performed a similar two round selection experiment with a novel negative selection step that included the undesired Npf in the media. Thus, in this second “counter-

selection” protocol, we explicitly selected against Npf incorporation rather than relying on high activity for Acd to be mutually exclusive of activity for Npf. The *Mj*RS libraries were based around the G2 (AcdRS1) and F9 clones, both of which incorporate Acd, and have very similar sequences. The level of Npf misincorporation is noticeably higher for F9 than for G2, highlighting the idea that small sequence changes can have a large impact on selectivity.

From these two selection protocols, eleven RS clones were identified that showed high levels of Acd incorporation. Of these clones, only A9 derived from the Npf counter-selection. Among the ten clones from the standard negative selection, the G11 clone demonstrated the highest level of selectivity as measured by the GFP fluorescence of cell suspensions expressed in media containing either 1 mM Acd or 1 mM Npf. (Fig. 1 and ESI Fig. S1) Therefore, the G11 and A9 RSs were cloned into the pDule2 vector for expression of other proteins, and are referred to as AcdRS2a and AcdRS2b, respectively.

Comparison of AcdRS Active Sites—The availability of an X-ray crystal structure of AcdRS1 (G2) allowed us to examine the sequences of various library members to understand how changes might confer increased selectivity for Acd relative to Npf. Although we were never able to obtain suitably diffracting crystals of an Acd complex with AcdRS1, the Mehl group has previously published a structure of this RS with Brb bound. (Fig. 1, Middle) Recall that “AcdRS1” is a permissive RS, as can be seen in an analysis of its activation of a variety of aromatic ncAAs in Cooley *et al.*³⁰ Brb, Npf, and Acd share the feature of a nitrogen atom in the *para* position of the phenylalanine ring, which seems to be a key recognition determinant as it can hydrogen bond with Glu₆₅. Both of the RS clones that give the highest selectivity for Acd, AcdRS2a (G11) and AcdRS2b (A9), have a mutation at this position, but one that maintains the potential hydrogen bond acceptor functionality. Other library members with Pro, Trp, Ile, or Val at position 65 had less than ten-fold selectivity for Acd over Npf in the GFP assay. (See ESI for details) Further analysis of AcdRS1 (G2), AcdRS2a (G11), and AcdRS2b (A9) selectivity based on homology modelling and docking studies is given below.

Analysis of AcdRS Selectivity

In Vivo Characterization of AcdRS Selectivity—To more rigorously investigate the selectivity of AcdRSs 2a and 2b, we expressed proteins in *E. coli*, purified them, and analyzed ncAA incorporation by MALDI MS of whole proteins and trypsin digests. As above, each experiment was carried out under five media conditions, varying in the amino acid provided and/or the synthetic route by which it was obtained: with H₂SO₄ Acd, PPA Acd, Npf, 1% Npf/Acd, and 10% Npf/Acd. Data for incorporation at position 113 in CaM are shown in Fig. 2 and Fig. 3; data for αS and LexA are shown in the ESI. As anticipated from the GFP screening data, we found that both AcdRS2a and 2b have improved selectivity against Npf. However, AcdRS2a still gives a mixture of Npf- and Acd-containing protein when H₂SO₄ Acd is used in the growth media. In contrast, proteins expressed using AcdRS2b contain only Acd, even when challenged with 10% Npf in the media. This level of selectivity was observed in CaM, αS, and LexA. (Fig. 2 and ESI Figs. S2–S4) It should also be noted that AcdRS2b selectivity is observed in a variety of media. CaM and αS are expressed in minimal media or Luria broth (LB) with isopropyl β-D-1-thiogalactopyranoside

(IPTG) induction, while LexA is expressed in richer, lactose auto-induction media. Based on these data, we selected the AcdRS2b (A9) for further characterization.

For a more rigorous, quantitative analysis of selectivity, we analyzed the CaM trypsin digest data by normalizing the intensity of the peak for the 108–116 fragment, containing Acd or Npf at position 113, to the intensity of the peak for the 117–127 fragment. (Fig. 3) CaM₁₁₇₋₁₂₇ should be produced in a 1:1 ratio with CaM₁₀₈₋₁₁₆ when the protein is completely digested by trypsin, and this can be confirmed by varying the digest time and observing that the intensity ratios do not change (data not shown). Normalization using an internal standard is essential to interpreting the intensity data correctly. The CaM₁₀₈₋₁₁₆Acd₁₁₃ fragment (A in Fig. 3) ionizes 5.5-fold better than the CaM₁₁₇₋₁₂₇ fragment (* in Fig. 3), while the CaM₁₀₈₋₁₁₆Npf₁₁₃ fragment (N in Fig. 3) ionizes 6.1-fold worse than the CaM₁₁₇₋₁₂₇ fragment. This can be seen by examining the peak ratios for PPA Acd (i.e., Acd only) and Npf expressions. After peak scaling, one obtains an MS-based Acd/Npf selectivity ratio (MS Sel) of 2.6×10^{-3} for AcdRS1 and 0.20 for AcdRS2b, calculated as follows:

$$\text{MS Sel} = (\text{Scaled Acd} / \text{Scaled Npf}) / (\text{Acd/Npf ratio in media})$$

The AcdRS1 MS Sel value is in good agreement with our estimate of 100-fold selectivity for Npf based on the less quantitative whole protein MALDI MS data. While the AcdRS2b MS Sel value may seem surprisingly low given the absence of any obvious CaM₁₀₈₋₁₁₆Npf₁₁₃ peak in the MALDI spectra, it is important to keep in mind that the RS is only being challenged with at most 10% Npf in the media and that CaM₁₀₈₋₁₁₆Npf₁₁₃ ionizes 30-fold worse than CaM₁₀₈₋₁₁₆Acd₁₁₃. Using these Acd/Npf selectivity ratios, we determined that selection resulted in a 76-fold improvement in selectivity for AcdRS2b relative to AcdRS1. We note that the effective *in vivo* Acd selectivities for both AcdRS1 and AcdRS2b are less than one and are probably influenced by preferential uptake of Npf into cells.

***In Vitro* Characterization of RS Activity**—To better understand AcdRS1 and 2b selectivity, we generated His-tagged variants of the enzymes, then expressed and purified them for *in vitro* activity assays. Charging of tRNA by RSs is a two-step process, where the first step is the RS-catalyzed reaction of amino acid with ATP to form an aminoacyl-adenylate intermediate (Aa-AMP), releasing inorganic pyrophosphate (PP_i); and the second step is the reaction of this enzyme-bound adenylate with the 2' or 3' hydroxyl group on A76 at the 3' end of the tRNA.³¹

While some prior studies have used assays that measure only the first step of the aminoacylation reaction to demonstrate that *in vitro* activities are consistent with ncAA

incorporation,^{24, 32, 33} it has been shown by the Perona laboratory that full tRNA aminoacylation assays correlate well with *in vivo* observations of RS activity.³⁴ This is expected, since amino acid incorporation into protein *in vivo* can only occur upon aminoacylation. Thus, aminoacylation of an *in vitro* transcribed ³²P-labeled tRNA_{CUA} was measured under single-turnover conditions at a variety of Acd or Npf concentrations for AcdRS1 and AcdRS2b. Plots of the first-order rate constants as a function of amino acid concentration were used to determine k_{obs} and K_{d} for each amino acid and RS combination, as previously described.³⁵ (Fig. 4 and Table 1) For these measurements, k_{obs} corresponds to the microscopic rate constant for the chemical steps of aminoacylation, or to a closely linked first-order rearrangement that follows tRNA binding and precedes aminoacylation.

Comparative kinetic analysis of AcdRS1 and AcdRS2b reveal catalytic preferences that correlate well with measurements of amino acid incorporation *in vivo* and mass spectrometry of proteins incorporating Acd and Npf. AcdRS1 aminoacylates tRNA_{CUA} with Acd three-fold slower than with Npf, while AcdRS2b aminoacylates tRNA_{CUA} with Acd five-fold faster than with Npf (Table 1). These changes in k_{obs} are primarily responsible for the change in selectivity between the RSs, as there are only relatively minor differences in K_{d} for the ncAAs. The enzymological selectivity (Enz Sel), calculated as

$$\text{Enz Sel} = [k_{\text{obs}}(\text{Acd}) / K_{\text{d}}(\text{Acd})] / [k_{\text{obs}}(\text{Npf}) / K_{\text{d}}(\text{Npf})]$$

changes from about a two-fold Npf preference by AcdRS1 to 16-fold Acd selectivity for AcdRS2b. Overall, AcdRS2b is improved in selectivity for Acd relative to Npf by 34-fold, in reasonable agreement with the 76-fold specificity shift derived from the *in vivo* measurements made based on trypsin digest peak intensities. Note that while the relative change in selectivity is consistent between the *in vitro* and *in vivo* data, the actual selectivities are quite different, presumably because the *in vivo* MS Sel is the product of Enz Sel and factors affecting the availabilities of the amino acids inside the cells. We estimate that the effective Npf concentration in cells is about 100-fold greater than the Acd concentration. This could be the result of preferential uptake of Npf, or sequestration of Acd by binding to other targets. Such targets could include DNA or RNA, both of which are known targets of acridone-based intercalator molecules.³⁶ Thus, to prevent Npf misincorporation, AcdRS2b must be significantly Acd-selective, even though Npf contaminants are only present at about 1% (10 μM) in the media.

Modeling of AcdRSs complexed with Npf and Acd—To inform the differences in Acd and Npf charging activity, models were built based upon the crystallographic structure of AcdRS1 complexed with Brb (PDB ID: 4PBR).³⁰ In generating a model of AcdRS1 (G2), the phenyl ring on the Brb ligand guided the placement of Acd and Npf within the binding pocket. Hydrogen atoms were added to the protein, and the ligand structure was energy minimized within the fixed protein structure. For Acd, a void volume remains between the molecule and the G₃₂-G₃₄ β -strand. In each G2 complex, Glu₆₅ is poised to accept a hydrogen-bond from the N-H of the ligand's side chain. For Npf, the additional free volume allows the molecule to relax to a twisted conformation, alleviating repulsive internal steric contacts between the ligand's two phenyl rings. Using the same protein structure,

computational protein design methods developed in the Saven laboratory were used to generate a model of A9,^{37–40} followed by energy minimization of the atomic coordinates of the ligands and the mutated side chains. In A9, the Gly₃₂Ala mutation forces Npf to adopt a more planar conformation of the ligand's two phenyl rings in order to form a hydrogen bond with Asp₆₅. Instead, Npf relaxes to avoid this high-energy distortion, which not only prevents hydrogen bonding of the Npf aniline N-H with Asp₆₅, but also forces the Asp₆₅ sidechain to move. While it is possible that the actual conformations adopted in the active site are different in the presence of tRNA, these models are consistent with AcdRS2b's increased activity toward Acd relative to Npf. Thus, it appears that the combination of the Gly₃₂Ala and Glu₆₅Asp mutations provides the increase in Acd selectivity for AcdRS2b (A9). Further experimental studies of single point mutations are required to validate this hypothesis, and will be reported when complete.

Conclusions

Our elucidation of the process by which trace Npf in the Acd media was incorporated into expressed proteins and our subsequent optimization of AcdRS1 provide a cautionary tale to the field of genetic code expansion, as well as a potentially general solution to the problem. Explicit inclusion of the unwanted Npf in the counter-selection step allowed us to obtain AcdRS2b, which had much higher selectivity toward its cognate ncAA than AcdRS2a, obtained through standard rounds of positive and negative selection (against only the 20 canonical amino acids). Trace contaminants in batches of ncAAs, synthesized either in-house or bought from commercial vendors, may indeed be utilized by many RSs in the literature which have never been purposefully selected for or against charging of these ncAAs. Moreover, variations of *in vivo* availability (resulting from differential uptake, for example) may exacerbate these problems, as appears to be the case with Npf.

Permissive RSs in libraries, which provide a very efficient way of obtaining RSs for new ncAAs, are likely to be susceptible to this problem. Indeed, RSs with open binding sites that accept bulky amino acids such as Acd and Brb may be particularly prone to this difficulty. However, using such a permissive RS with an open pocket is sometimes necessary to finding an RS for an amino acid such as Acd. In some cases, RS selection for bulky ncAAs has required an initial round of selection using a more moderate sized amino acid to expand the binding pocket before selection for charging of a bulky aromatic ncAA.¹⁶ Such RSs may be prone to misincorporation of amino acids similar to that intermediate amino acid. Here, we have seen that small sequence changes (i.e. Gly->Ala and Glu->Asp) can result in large increases in *in vitro* and *in vivo* selectivity for permissive RSs. Thus, we expect that the counter-selection strategy used here will be generally useful to the genetic code expansion community.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by funding from the National Science Foundation (NSF CHE-1150351 to E.J.P., NSF MCB-1518265 to R.A.M., NSF CHE-1508318 to J.G.S.), National Institutes of Health (NIH DP2-GM105444 to R.M.K.) and the Searle Scholars Program (10-SSP-214 to E.J.P.). Instruments supported by the National Science Foundation and National Institutes of Health include: LCMS/HRMS (NIH RR-023444), MALDI MS (NSF MRI-0820996), NMR (NIH RR-022442), and a stopped flow fluorometer (NSF CHE-1337449). J.G.S. acknowledges infrastructural support from the Penn Laboratory for Research on the Structure of Matter (NSF DMR 1120901) and use of the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF grant ACI-1053575 under grant TG-CHE 110041. I.S. thanks the Royal Thai Foundation for fellowship support. Z.M.H. and J.A.V. thank the the NIH for funding through the Chemistry Biology Interface Training Program (T32 GM071399). J.A.V. and J.G.S. acknowledge assistance and preliminary computational studies from Lu Gao.

Notes and references

- Chin D, Means AR. Trends Cell Biol. 2000; 10:322–328. [PubMed: 10884684]
- Vetter SW, Leclerc E. Eur J Biochem. 2003; 270:404–414. [PubMed: 12542690]
- Butala M, Zgur-Bertok D, Busby SJW. Cell Mol Life Sci. 2009; 66:82–93. [PubMed: 18726173]
- Culyba MJ, Mo CY, Kohli RM. Biochemistry. 2015; 54:3573–3582. [PubMed: 26016604]
- Lashuel HA, Overk CR, Oueslati A, Masliah E. Nat Rev Neurosci. 2013; 14:38–48. [PubMed: 23254192]
- Spillantini MG, Schmidt ML, Lee VMY, Trojanowski JQ, Jakes R, Goedert M. Nature. 1997; 388:839–840. [PubMed: 9278044]
- Haney CM, Wissner RF, Petersson EJ. Curr Opin Chem Biol. 2015; 28:123–130. [PubMed: 26253346]
- Giepmans BNG, Adams SR, Ellisman MH, Tsien RY. Science. 2006; 312:217–224. [PubMed: 16614209]
- Lakowicz, JR. Principles of Fluorescence Spectroscopy. 3. Springer; US: 2006.
- Speight LC, Samanta M, Petersson EJ. Aust J Chem. 2014; 67:686–700.
- Lang K, Chin JW. Chem Rev. 2014; 114:4764–4806. [PubMed: 24655057]
- Speight LC, Muthusamy AK, Goldberg JM, Warner JB, Wissner RF, Willi TS, Woodman BF, Mehl RA, Petersson EJ. J Am Chem Soc. 2013; 135:18806–18814. [PubMed: 24303933]
- Dumas A, Lercher L, Spicer CD, Davis BG. Chem Sci. 2015; 6:50–69. [PubMed: 28553457]
- Summerer D, Chen S, Wu N, Deiters A, Chin JW, Schultz PG. Proc Natl Acad Sci U S A. 2006; 103:9785–9789. [PubMed: 16785423]
- Wang J, Xie J, Schultz PG. J Am Chem Soc. 2006; 128:8738–8739. [PubMed: 16819861]
- Lee HS, Guo J, Lemke EA, Dimla RD, Schultz PG. J Am Chem Soc. 2009; 131:12921–+. [PubMed: 19702307]
- Chatterjee A, Guo JT, Lee HS, Schultz PG. J Am Chem Soc. 2013; 135:12540–12543. [PubMed: 23924161]
- Harkiss AH, Sutherland A. Org Biomol Chem. 2016; 14:8911–8921. [PubMed: 27714204]
- Luo J, Uprety R, Naro Y, Chou CJ, Nguyen DP, Chin JW, Deiters A. J Am Chem Soc. 2014; 136:15551–15558. [PubMed: 25341086]
- Szymanska A, Wegner K, Lankiewicz L. Helv Chim Acta. 2003; 86:3326–3331.
- Taki M, Yamazaki Y, Suzuki Y, Sisido M. Chem Lett. 2010; 39:818–819.
- Hamada H, Kameshima N, Szymanska A, Wegner K, Lankiewicz L, Shinohara H, Taki M, Sisido M. Bioorg Med Chem. 2005; 13:3379–3384. [PubMed: 15848750]
- Goldberg JM, Speight LC, Fegley MW, Petersson EJ. J Am Chem Soc. 2012; 134:6088–6091. [PubMed: 22471784]
- Wang L, Brock A, Herberich B, Schultz PG. Science. 2001; 292:498–500. [PubMed: 11313494]
- Wang L, Schultz PG. Angew Chem Int Ed. 2005; 44:34–66.
- Miyake-Stoner SJ, Refakis CA, Hammill JT, Lusic H, Hazen JL, Deiters A, Mehl RA. Biochemistry. 2010; 49:1667–1677. [PubMed: 20082521]

27. Peeler JC, Woodman BF, Averick S, Miyake-Stoner SJ, Stokes AL, Hess KR, Matyjaszewski K, Mehl RA. *J Am Chem Soc.* 2010; 132:13575–13577. [PubMed: 20839808]
28. Stokes AL, Miyake-Stoner SJ, Peeler JC, Nguyen DP, Hammer RP, Mehl RA. *Mol Biosys.* 2009; 5:1032–1038.
29. Díaz E, Ferrández A, Prieto MA, García JL. *Microbiol Mol Biol Rev.* 2001; 65:523–569. [PubMed: 11729263]
30. Cooley RB, Karplus PA, Mehl RA. *ChemBioChem.* 2014; 15:1810–1819. [PubMed: 25044993]
31. Ibba M, Soll D. *Ann Rev Biochem.* 2000; 69:617–650. [PubMed: 10966471]
32. Francklyn CS, First EA, Perona JJ, Hou YM. *Methods.* 2008; 44:100–118. [PubMed: 18241792]
33. Nehring S, Budisa N, Wiltshi B. *PLOS ONE.* 2012; 7
34. Rauch BJ, Porter JJ, Mehl RA, Perona JJ. *Biochemistry.* 2016; 55:618–628. [PubMed: 26694948]
35. Rodríguez-Hernández A, Perona John J. *Structure.* 19:386–396.
36. Belmont P, Bosson J, Godet T, Tiano M. *Anticancer Agents Med Chem.* 2007; 7:139–169. [PubMed: 17348825]
37. Kono H, Saven JG. *J Mol Biol.* 2001; 306:607–628. [PubMed: 11178917]
38. Calhoun JR, Kono H, Lahr S, Wang W, DeGrado WF, Saven JG. *J Mol Biol.* 2003; 334:1101–1115. [PubMed: 14643669]
39. Bender GM, Lehmann A, Zou H, Cheng H, Fry HC, Engel D, Therien MJ, Blasie JK, Roder H, Saven JG, DeGrado WF. *J Am Chem Soc.* 2007; 129:10732–10740. [PubMed: 17691729]
40. Fry HC, Lehmann A, Sinks LE, Asselberghs I, Tronin A, Krishnan V, Blasie JK, Clays K, DeGrado WF, Saven JG, Therien MJ. *J Am Chem Soc.* 2013; 135:13914–13926. [PubMed: 23931685]

Fig. 1. AcdRS Selection. Top Left: Images of *E. coli* agar plates used in rounds of positive (+ Acd) and negative (+ Npf) selection. Media for both plates also contain sources of the 20 canonical amino acids. Top Right: Fluorescence of suspensions of *E. coli* cells expressing GFP with a TAG codon at position 150 using the indicated RS clone and amino acid mixture. Emission was measured at 528 nm with excitation at 485 nm. Middle: Image of the AcdRS1 (G2) active site with radical polymerization initiator Brb bound. (PDB ID: 4PBR) A favourable hydrogen bond between the carboxylate of Glu₆₅ and the aniline N-H of Brb can be seen. Bottom Sequences of *Mj* RS clones used for incorporation of Acd. Additional sequences of clones from GFP-based screening are given in the ESI.

Fig. 2. In Vivo

AcdRS Selectivity. CaM (UAG codon at 113) was expressed in minimal media containing 1 mM ncAA: Acd, synthesized either using the H₂SO₄ route or PPA route, or Npf. Expression was performed with one of the three AcdRSs indicated. Significant incorporation of Npf is seen for AcdRS2a (G11) when using H₂SO₄ Acid, but only Acd-containing protein is seen with AcdRS2b (A9). Experimental details are given in the ESI.

Fig. 3.

CaM₁₁₃ Trypsin Digest Data for AcRS Selectivity Analysis. CaM (UAG codon at 113) was expressed in minimal media containing 1 mM ncAA: Ac, synthesized either using the H₂SO₄ route or PPA route, Npf, or either a 99:1 or 90:10 mixture of PPA Ac and Npf. Expression was performed with one of the three AcRSs indicated. The peaks for the (M + H)⁺ masses of the CaM₁₀₈₋₁₁₆Ac₁₁₃ fragment (A, 1179.8 Da), CaM₁₀₈₋₁₁₆Npf₁₁₃ fragment (N, 1153.8 Da), and CaM₁₁₇₋₁₂₇ fragment (*, 1349.9 Da) are indicated. The intensities of the CaM₁₀₈₋₁₁₆Ac₁₁₃ and CaM₁₀₈₋₁₁₆Npf₁₁₃ fragment peaks were normalized using the intensity of the CaM₁₁₇₋₁₂₇ fragment peak.

Fig. 4. tRNA Aminoacylation Kinetics. Plot of k_{obs} for reactions of AcRS1 and AcRS2b as a function of Acd or Npf concentration. Reaction rates were determined using thin layer chromatography following nuclease digestion of aminoacylated tRNA as described in the ESI. Sample primary data are shown in Fig. S7. Rate constants were determined from progress curves for each enzyme at varying amino acid concentrations.

Fig. 5. AcdRS Homology Models. Acd or Npf were energy minimized in the active site of AcdRS1 (G2), taken directly from PDB structure file 4PBR, or a model of AcdRS2b (A9). Npf is accommodated in the AcdRS1 active site where the phenyl rings can become non-planar with respect to each other. In the A9 active site, Npf twisting is restricted by the Ala₃₂ sidechain, and a hydrogen bond cannot be made with Asp₆₅.

Scheme 1.

Synthesis of Acids by H_2SO_4 or Polyphosphoric Acid (PPA) Routes. (a) 1. SOCl_2 , MeOH 2. Boc_2O , Na_2CO_3 , THF/ H_2O (3:1) 3. PhNTf_2 , DMAP, NEt_3 , CH_2Cl_2 ; 88% yield over 3 steps; b) $\text{Pd}(\text{OAc})_2$, *rac*-BINAP, Cs_2CO_3 ; 92% yield; c) H_2SO_4 , reflux; 98% yield; d) 1. LiOH, 2. PPA; 53% yield over two steps (racemic mixture)

Table 1AcdRS *In Vitro* Enzymology and *In Vivo* MS Selectivity Parameters

AcidRS	$k_{\text{obs}}(\text{Acd}) \times 10^{-3} \text{ s}^{-1}$	$K_d(\text{Acd}) \mu\text{M}$	$k_{\text{obs}}(\text{Npf}) \times 10^{-3} \text{ s}^{-1}$	$K_d(\text{Npf}) \mu\text{M}$	Enz Sel	MS Sel
1 (G2)	8.0 ± 2.0	45.1 ± 18.8	24.0 ± 1.0	64.1 ± 10.8	0.47	2.6×10^{-3}
2b (A9)	28.0 ± 2.0	39.9 ± 6.8	5.0 ± 1.0	114 ± 10.0	16	0.20