1894 *Biochemistry* **2011,** *50,* 1894–1900
DOI: 10.1021/bi101929e

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

An Evolved Aminoacyl-tRNA Synthetase with Atypical Polysubstrate Specificity[†,‡]

Douglas D. Young,[§,⊥] Travis S. Young,[§,⊥] Michael Jahnz,[§] Insha Ahmad,[§] Glen Spraggon,[∥] and Peter G. Schultz*[,§]

[§]*Department of Chemistry and Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States, and* [∥]*Protein Science, Genomics Institute of the Novartis Research Foundation, San Diego, California 92121, United States.* [⊥]*These authors contributed equally to this work.*

Received December 2, 2010; Revised Manuscript Received January 11, 2011

ABSTRACT: We have employed a rapid fluorescence-based screen to assess the polyspecificity of several aminoacyl-tRNA synthetases (aaRSs) against an array of unnatural amino acids. We discovered that a *p*-cyanophenylalanine specific aminoacyl-tRNA synthetase (*p*CNF-RS) has high substrate permissivity for unnatural amino acids, while maintaining its ability to discriminate against the 20 canonical amino acids. This orthogonal *p*CNF-RS, together with its cognate amber nonsense suppressor tRNA, is able to selectively incorporate 18 unnatural amino acids into proteins, including trifluoroketone-, alkynyl-, and halogen-substituted amino acids. In an attempt to improve our understanding of this polyspecificity, the X-ray crystal structure of the aaRS−*p*-cyanophenylalanine complex was determined. A comparison of this structure with those of other mutant aaRSs showed that both binding site size and other more subtle features control substrate polyspecificity.

A wide variety of unnatural amino acids (UAAs) have been genetically encoded in bacteria, yeast, and mammalian cells with excellent fidelity and efficiency. These novel amino acids include nuclear magnetic resonance (NMR), infrared (IR), fluorescent, and crystallographic probes, orthogonal bioconjugation partners, metal ion chelators, redox active centers, and photo-cross-linking and photocaging agents (*1*, *2*). An orthogonal aminoacyl-tRNA synthetase/suppressor tRNA (aaRS/$tRNA_{CUA}$) pair (one that does not significantly cross-react with host tRNAs or aminoacyl-tRNA synthetases) is used to selectively incorporate the unnatural amino acid of interest in response to a nonsense or frameshift codon. To produce aaRSs that recognize an unnatural amino acid and no endogenous amino acid, previously we employed a double-sieve selection scheme that ties the activity of the aaRS to the viability of *Escherichia coli* (*3*, *4*). In the positive selection, large libraries of aaRS active site mutants are screened for their ability to incorporate the unnatural amino acid in response to an amber nonsense codon at a permissive site in an essential protein; in the negative selection, aaRSs that incorporate endogenous amino acids into a lethal protein in the absence of the unnatural amino acid are removed.

Because no selective pressure is applied against other unnatural amino acids during the selection, evolved aaRSs may also cross-react with other unnatural amino acids, while maintaining their orthogonality to endogenous amino acids (*5*−*9*). This polyspecificity can be exploited in incorporating additional unnatural amino acids without the evolution of new, orthogonal aaRS/tRNA pairs. Aminoacyl-tRNA synthetase promiscuity is not problematic in this regard as the growth media can be supplemented with only the desired unnatural amino acid. Previous experiments have examined the substrate permissivity of the pyrrolysyl, *p*-benzoylphenylalanine, and naphthylalanine synthetases, and mutations were introduced to broaden their substrate specificity (*7*). However, a thorough analysis of multiple aminoacyl-tRNA synthetases with a broad range of unnatural amino acid substrates has yet to be conducted (*3*). Thus, to systematically analyze and improve our understanding of the permissivity of multiple evolved aaRSs, we have investigated their ability to incorporate a diverse library of unnatural amino acids using a simple fluorescent reporter-based assay.

EXPERIMENTAL PROCEDURES

GFP Expression. GFP expression was conducted as previously described. BL-21(DE3) *E. coli* was cotransformed with the appropriate pEVOL plasmid and pET-GFP_{Y151X} plasmid (*1*), grown in 2×YT medium at 37 °C to saturation with chloramphenicol (40 μg/mL) and ampicillin (100 μg/mL), and diluted in 2×YT to an OD_{600} of 0.2. Diluted cultures were grown at 37 °C to an OD_{600} of 0.7−1.0 and induced with IPTG (1 mM) and arabinose (0.02%). The culture was immediately aliquoted (100 μL/well) into a 96-well plate containing 10 μL of 10 mM unnatural amino acid per well, and background fluorescence was measured on a Spectramax Gemini EM instrument (excitation and emission at 395 and 509 nm, respectively; Molecular Probes). After 14 h at 30 °C, fluorescence was measured again to determine the level of GFP expression. Amino acids that afforded a significantly measurable signal were then employed in similar expression experiments in triplicate. Each sample was normalized to cell density using the OD_{600}, and fluorescence was measured and compared to that of wild-type expression (Table 2), confirming the incorporation of UAA.

Myoglobin Expression. BL-21(DE3) *E. coli* was cotransformed with the appropriate pEVOL plasmid and pET-Myo_{F107X},

[†]This work was funded by Grant DE-FG03-00ER46051 from the Division of Materials Sciences, Department of Energy (P.G.S.). D.D.Y. is grateful for National Institutes of Health Fellowship IF32CA144213, and T.S.Y. is grateful for an Achievement Rewards for College Scientists Scholarship.
[‡]This is manuscript 20691 of the Scripps Research Institute.
*To whom correspondence should be addressed. Phone: (858) 784-9273. Fax: (858) 784-9440. E-mail: schultz@scripps.edu.

 Published on Web 02/01/2011

grown in 2×YT medium at 37 °C to saturation with chloramphenicol (40 μg/mL) and ampicillin (100 μg/mL), and diluted in 2YT to an OD_{600} of 0.2. Diluted cultures were grown at 37 °C to an OD_{600} of 0.7, induced with IPTG (1 mM), arabinose (0.02%), and the appropriate unnatural amino acid (1 mM), and grown at 37 °C for 16 h. The cultures were pelletted and lysed using Bug Buster (Novagen), and the protein was purified on Ni-NTA spin columns (Qiagen) according to the manufacturer's protocol. Myoglobin expression was analyzed on a 4 to 20% Gly-Tris polyacrylamide gel (see the Supporting Information), and the mass of the purified protein was determined by liquid chromatography and mass spectroscopy (LC−MS) on an Agilent 1100 Series LC/MSD instrument. The chromatographic peak corresponding to myoglobin (between 6.1 and 6.5 min) was charge deconvoluted using Agilent LC/MSD ChemStation (revision B.03.02). Deconvolution parameters were set as follows: high M_r = 17000, low M_r = 21000, maximum charge of 50, and 3−8 minimum peaks in set. The error was ±0.02%, as determined from control samples. LC−MS results for Myo_{Y107X} expression experiments are listed in the Supporting Information.

Aminoacyl-tRNA Synthetase Crystallization. Protein expression and crystallization were conducted as previously described (*10*−*12*). The *p*CNF-RS DNA was amplified from the pBK-*p*CNF vector (*13*) (forward primer, 5′ GCAAGCGCATATGGACGAATTTGAAATGA 3′; reverse primer, 5′ GTTCGGCTCGAGTAATCTCTTTCTAATTG 3′), inserted into the pET-22b(+) vector using the XhoI and NdeI restriction sites, and transformed into BL-21(DE3) cells. A 1 L 2×YT (Amp) expression culture was inoculated to an OD_{600} of 0.2, grown at 37 °C to an OD_{600} of 0.6, and induced with IPTG (1 mM). The culture was grown at 37 °C for 16 h, pelleted, and lysed [50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, and 10 mM BME (pH 8.0)] by sonication. The sample was then centrifuged (18000 rpm for 20 min), and the supernatant was incubated with Ni-NTA resin (2 mL; Qiagen) for 1 h at 4 °C. The resin was then washed [100 mL of 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, and 10 mM BME (pH 8.0)] and eluted [5 mL of 50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, and 10 mM BME (pH 8.0)], followed by dialysis [3 × 1 h at room temperature in 25 mM Tris, 50 mM NaCl, 1 mM EDTA, and 10 mM BME (pH 8.0)]. The aaRS was purified by FPLC on a MonoQ column [buffer A consisted of 25 mM Tris, 25 mM NaCl, 1 mM EDTA, and 10 mM BME (pH 8.5); buffer B consisted of 25 mM Tris, 1 M NaCl, 1 mM EDTA, and 10 mM BME (pH 8.5)], concentrated to 2 mL, and dialyzed into crystallization buffer [20 mM Tris, 50 mM NaCl, and 10 mM BME (pH 8.0)]. Crystals were grown by the sitting-drop vapor-diffusion method using a 1:1 mixture of concentrated protein (15 mg/mL, with 2 mM pCNF) and mother liquor [0.2 M (NH_4)tartrate and 20% PEG-3350 (pH 6.6)] at 20 °C.

Data Collection, Structure Determination, and Refinement. Data for the pCNF cocrystals were collected at beamline 5.0.3 of the Advanced Light source (ALS) at a wavelength of 0.9778 Å to a maximum Bragg spacing of 2.3 Å (Table 1). All data were reduced and scaled using the HKL2000 package (*14*). The crystal belonged to the monoclinic crystal system with the β angle approaching 90°. A strong peak in the native Patterson map indicated the presence of pseudotranslational symmetry and indicated that the crystal was pseudo-orthorhombic. This was confirmed by the inability of the unreduced data to scale reasonably in the equivalent orthorhombic space group. The presence of noncrystallographic translational symmetry induces considerable deviation from Wilson statistics. The Wilson ratios ($\langle I^2\rangle/\langle I\rangle^2$) for the data set calculated from phenix xtriage are 2.752 and 2.776 for acentric and centric reflections, respectively, whereas perfect data should have values of 2.0 and 3.0, respectively. The Wilson distribution, which is based on the assumption of a random distribution of atoms in the cell, is therefore considerably skewed by this translational symmetry, and statistics such as the *R* factor and procedures such as the calculation of normalized amplitudes that are based on this assumption can be affected. Despite this, the structure was easily determined using Phaser (*15*) using all data between 50 and 2.3 Å resolution using the wild-type tyrosyl-tRNA synthetase (*16*) [Protein Data Bank (PDB) entry 1J1U] as a probe. The molecular replacement solution confirmed the presence of the pseudotranslational symmetry between the two molecules in the asymmetric unit. Electron density produced from an initial refinement with Buster clearly showed the mutated residues between the probe model and the structure as well as electron density for pCNF in both molecules, confirming the validity of the molecular replacement solution. Model building and refinement were performed by iterative building and refinement with Coot and Buster (*17*, *18*). All other crystallographic manipulations were conducted with the CCP4 program suite (*19*). Convergence of the refinement was checked throughout, with weights between geometric and X-ray terms optimized on the basis of the R_{free} in combination with a criterion for reasonable geometry. The Local Structure Symmetry Restraints (LSSR) procedure was also applied throughout refinement (*20*), also incorportated on the basis of its effect on R_{free}. The validity of the final structure was checked using a combination of the validation tools in Coot (*17*), including inspection of the Ramachandran plot, real space correlation, and rotamer and geometry analysis. Using these refinement and geometrical criteria, the final model converged with R_{cryst} and R_{free} values of 23.0 and 30.0% respectively, excellent geometry, root-mean-square deviations (rmsd) of bonds and angles of 0.013 $Å^2$ and 1.3° respectively, no residues in disallowed regions of the Ramachandran plot (Table 1), and all of the mutations in the structure in a preferred rotamer conformation.

Table 1: Data and Refinement Statistics

space group	$P2_1$
unit cell dimensions	a = 52.52 Å, b = 68.93 Å, c = 82.09 Å, β = 90.59°
wavelength (Å)	0.9778
resolution (Å) (highest-resolution shell)	2.3 (2.34−2.3)
R_{merge} (%) (highest-resolution shell)	0.066 (0.25)
no. of unique reflections	46732
completeness (%) (highest-resolution shell)	96.2 (76.6)
I/σ (highest-resolution shell)	15.0 (2.1)
redundancy (highest-resolution shell)	3.0 (2.1)
Refinement	
R_{factor} (R_{free}[a]) (%)[b]	0.23 (0.30)
no. of protein atoms	4913
no. of water atoms	172
no. of heteroatoms	28
rmsd for bonds (Å)	0.013
rmsd for angles (deg)	1.37
mean B factor ($Å^2$)	47.9

[a] R_{free} equals R_{cryst}, but for 5.0% of the total reflections chosen at random and omitted from the refinement. [b] R factor = $\sum\sum|I_i - \langle I_i\rangle|/\sum|I_i|$, where I_i is the scaled intensity of the ith measurement and $\langle I_i\rangle$ is the mean intensity for that reflection.

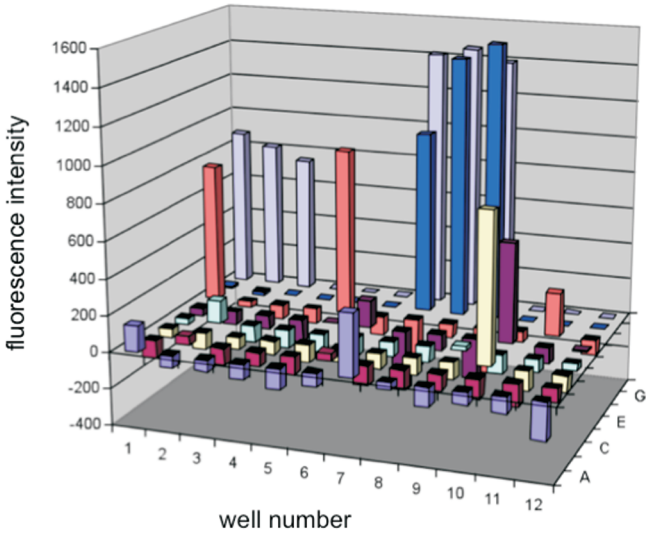

FIGURE 1: GFP fluorescence assay for *p*CNF-RS-dependent incorporation of unnatural amino acids into GFP_{Y151X}. A 96-well plate contained different non-native amino acids [1 mM (see the Supporting Information)] incubated with BL-21(DE3) cells harboring the GFP_{Y151X} reporter and pEVOL-*p*CNF in 2×YT medium. Fluorescence was measured after expression for 14 h at 30 °C and corrected for background by a noninduced culture.

RESULTS AND DISCUSSION

Assay Development. To develop a rapid assay for analyzing the incorporation of unnatural amino acids by an evolved aaRS, we used a fluorescence-based GFP reporter. A GFP gene with an amber mutation at Tyr151 (GFP_{Y151X}) was placed under the control of the T7 promoter in a pET101 vector as previously described (*21*). Tyrosine 151 is a surface residue that can be substituted with a range of structurally diverse amino acids without any effect on protein stability or fluorescence (*21*). This reporter affords fluorescence only when the TAG codon is used as a sense codon for the acylated suppressor $tRNA_{CUA}$ (and not a stop codon). Mutant *Methanocaldococcus jannaschii* tyrosyl-RS/$tRNA_{CUA}$ (*Mj*TyrRS/$tRNA_{CUA}$) pairs encoded in the efficient pEVOL system (*21*) were cotransformed into BL-21(DE3) cells with the pET-GFP_{Y151X} construct to allow rapid screening for substrate permissivity. Aminoacyl-tRNA synthetases specific for *p*AcF (*p*-acetylphenylalanine, **7**) (*22*), *p*AzF (*p*-azidophenylalanine, **8**) (*23*), *p*CNF (*p*-cyanophenylalanine, **1**) (*13*), *p*IF (*p*-iodophenylalanine, **2**) (*24*), *p*BoF (*p*-borophenylalanine) (*25*), *p*PrF (*p*-propargylphenylalanine, **14**) (*26*), *p*CMF (*p*-carboxymethylphenylalanine) (*27*), *p*BpF (*p*-benzoylphenylalanine) (*28*), NapA (naphthylalanine) (*29*), CouA (7-hydroxycoumarin-4-yl ethylglycine) (*30*), PLA (*p*-hydroxy-L-phenyllactic acid) (*31*), BipyA (bipyridylalanine) (*32*), and HQA (8-hydroxyquinolin-3-ylalanine) (*33*), in addition to the wild-type tyrosyl-tRNA synthetase, were analyzed (*1*, *34*). A 96-well format was used to screen 72 unnatural amino acids including *para*- and *meta*-substituted phenylalanine analogues, histidine and alanine derivatives, amine, biaryl, and thiol-containing amino acids (see the Supporting Information). While fluorescent amino acids were included in the plate, other than CouA (well E7), the wavelengths of excitation and emission were nonoverlapping with GFP, and a background subtraction prior to expression was sufficient to allow the assessment of incorporation. The fluorescence of each well supplemented with 1 mM amino acid was measured after expression for 14 h in rich medium (Figure 1). To rapidly assess amino acid incorporation, fluorescence readings were taken on the cultures, as opposed to lysed cells, and potential hits were later confirmed by protein isolation and mass spectrometry.

Many of the aaRSs failed to exhibit a significant degree of substrate promiscuity with respect to the UAA in our screen. Not surprisingly, wild-type *Mj*TyrRS exhibited no detectable incorporation of any unnatural amino acid in the screen. Additionally, the *p*BoF, *p*PrF, *p*CMF, HQA, CouA, PLA, NapA, and BipyA specific aaRSs displayed a very low propensity for incorporation of unnatural amino acids beyond the specific unnatural amino acid for which they evolved. The *p*AcF, *p*IF, and *p*AzF synthetases were able to incorporate several *para*-substituted phenylalanine analogues included in the screen, but only to a limited extent (see the Supporting Information). Most interestingly, *p*CNF-RS (*13*) exhibited a substantial degree of polyspecificity as it afforded increased GFP_{Y151X} fluorescence in the presence of *p*-chlorophenylalanine (**2**), *p*-fluorophenylalanine (**5**), *p*-acetylphenylalanine (**7**), *p*-alkynylphenylalanine (**6**), benzylserine (**15**), and *p*-phenylphenylalanine (**12**) (Figure 1). Screen hits were validated by incorporation of the unnatural amino acid into a His-tagged myoglobin mutant containing an F107TAG amber mutation (Myo_{F107X}) (*21*). The protein was purified using a Ni-NTA resin and analyzed by sodium dodecyl sulfate−polyacrylamide gel electrophoresis. The calculated mass of the purified mutant protein was confirmed by LC−MS. In the absence of any unnatural amino acid, low levels of phenylalanine incorporation in response to the amber codon could be detected (< 20%); however, in the presence of any of these amino acids (≥ 1 mM), only incorporation of the unnatural amino acid was observed within the detection limits of LC−MS (*21*). The behavior of *p*CNF-RS in this assay was unique with respect to both incorporation efficiency and substrate diversity relative to the other 13 aminoacyl-tRNA synthetases.

Determination of Substrate Scope for pCNF-RS Polyspecificity. *p*CNF-RS appears to be highly tolerant of *para*-substituted phenylalanine derivatives. Consequently, the *p*CNF-RS was further screened against additional *para*-substituted aromatic amino acids (Figure 2). The majority of these amino acids (**1**−**18**) were incorporated into GFP_{Y151X} by the *p*CN-RS [relative incorporation was assessed in triplicate and normalized to cell density (Table 2)]; incorporation was further validated by incorporation into Myo_{F107X} with subsequent mass spectral analysis (Table 2). *p*CNF-RS substrates include UAAs that have already been genetically encoded in bacteria, e.g., *p*IF (**2**), *p*BrF (*p*-bromophenylalanine **3**), *p*AcF (**7**), *p*AzF (**8**), *p*PrF (**14**), pNO_2F (*p*-nitrophenylalanine **11**), oNO_2F (*o*-nitrophenylalanine **13**), *O*-methyl-Tyr (**16**), and *O*-allyl-Tyr (**17**) (*2*). *p*CNF-RS substrates also include unnatural amino acids that have not previously been encoded in bacteria, e.g., *p*ClF (**4**), *p*FF (**5**), *p*-ethynylphenylalanine (**6**), *p*-trifluoromethylacetylphenylalanine (**9**), *p*-phenylphenylalanine (**12**), benzylserine (**15**), and *O*-*tert*-butyltyrosine (**18**). On the basis of GFP expression levels, many of the amino acids are good substrates for *p*CNF-RS, affording > 80% protein expression relative to wild-type GFP. Interestingly, several amino acids [*p*IF (**2**), *p*BrF (**3**), *p*ClF (**4**), *p*-alkynylF (**6**), *p*AzF (**8**), *p*-isopropylF (**10**), and OMeY (**16**)] afford higher protein yields than the *p*CNF substrate did (**1**), the UAA for which the *p*CNF-RS was originally selected. Only amino acids **19**−**21** (*p*-carboxyphenylalanine, *p*-methylphosphonic acid phenylalanine, and *p*-aminophenylalanine), which have hydrogen bond donating substituents, were not incorporated. It is interesting to note that despite the apparent polysubstrate specificity of the *p*CNF-RS, **21** (*p*-aminophenylalanine) was

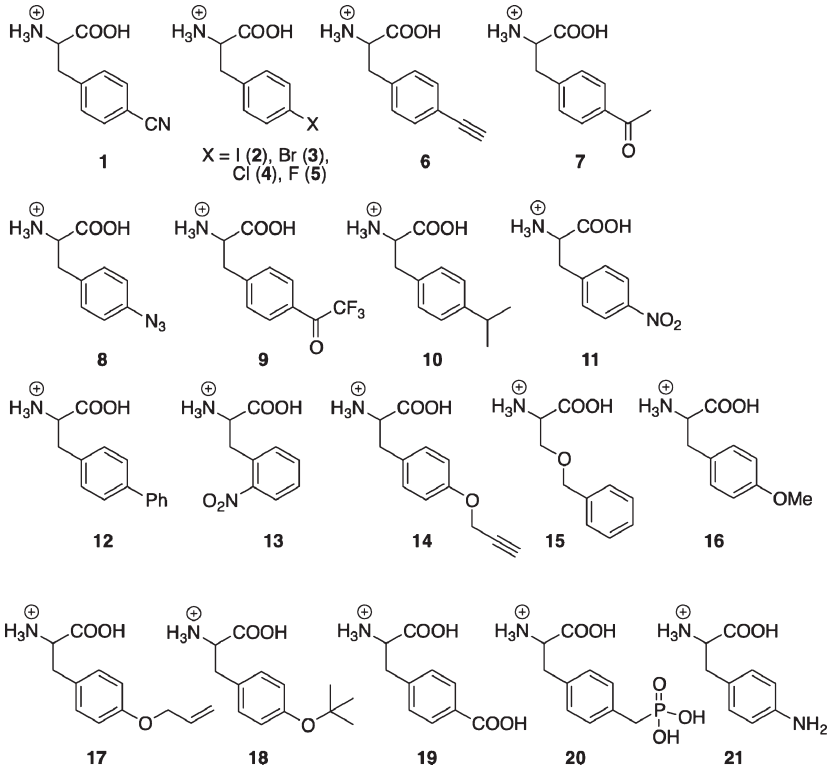

FIGURE 2: Unnatural amino acids screened for incorporation by *p*CNF-RS.

Table 2: Incorporation of Unnatural Amino Acids by *p*CNF-RS

aaRS	UAA	expected[a]	observed	relative level of expression[b]
WT[c]	Phe[c]	18352	18353	100 ± 3.9
*p*CNF	**1**	18377	18376	79 ± 2.9
*p*CNF	**2**	18478	18477	87 ± 2.2
*p*CNF	**3**	18431	18431	96 ± 1.4
*p*CNF	**4**	18386	18386	91 ± 5.2
*p*CNF	**5**	18370	18368	45 ± 4.9
*p*CNF	**6**	18376	18376	84 ± 1.4
*p*CNF	**7**	18394	18396	60 ± 3.0
*p*CNF	**8**	18393	18393	95 ± 2.3
*p*CNF	**9**	18450	18450	43 ± 6.1
*p*CNF	**10**	18394	18393	82 ± 3.9
*p*CNF	**11**	18396	18395	75 ± 3.7
*p*CNF	**12**	18429	18326	81 ± 1.2
*p*CNF	**13**	18396	18394	61 ± 1.8
*p*CNF	**14**	18406	18405	72 ± 3.0
*p*CNF	**15**	18382	18382	45 ± 4.2
*p*CNF	**16**	18383	18383	94 ± 1.4
*p*CNF	**17**	18407	18407	77 ± 3.2
*p*CNF	**18**	18422	18422	61 ± 4.6

[a]LC−MS expected mass for UAA incorporated into F107TAG myoglobin. [b]Based on GFP-Y151TAG expression and normalized to WT GFP expression (16 mg/L). [c]Mass for wild-type Myo with Phe.

not incorporated while other significantly larger substituents were tolerated, indicating the pCNF-RS has some degree of specificity (**21** has been previously shown to be taken up efficiently by *E. coli* and is stable in the cytoplasm) (*35*). In total, the *p*CNF-RS was found to incorporate 18 unnatural amino acids, seven of which had not been previously encoded. A number of the latter UAAs, e.g., *p*-alkynylphenylalanine (**6**) and *p*-trifluoroacetylphenylalanine (**9**), may be useful as orthogonal bioconjugation handles or as warheads for the selective inactivation of enzymes.

X-ray Crystal Structure of pCNF-RS. To improve our understanding of the molecular basis of *p*CNF-RS's polysubstrate specificity (which in contrast to polyspecific enzymes like p450s, is not a solvent-exposed surface cavity), the aaRS was overexpressed and crystallized for X-ray structure analysis (*10*−*12*, *36*). The crystal structure of the *p*CNF-RS−*p*CNF complex (PDB entry 3QE4) was determined to 2.3 Å with an R_{factor} of 0.23 (Table 1 and the Supporting Information). The *p*CNF-RS structure superimposes well with the WT TyrRS-Tyr structure (PDB entry 1J1U) except in the amino acid binding site. The amino acid binding site for *p*CNF-RS is relatively hydrophobic and contains the following mutations relative to the wild-type synthetase: Y32L, L65V, F108W, N109M, D158G, and I159A. A comparison of the WT TyrRS structure with the *p*CNF-bound *p*CNF-RS structure (Figure 4A vs Figure 4B) reveals that mutations create a larger binding pocket that allows the enzyme to accept the nitrile group. In the WT TyrRS structure, residues Y32 and D158 are involved in hydrogen bonding with the tyrosine hydroxyl group. These residues in *p*CNF-RS are mutated to leucine and glycine, respectively, which removes hydrogen bonding potential and likely explains the inability of *p*CNF-RS to accept tyrosine and other phenylalanine derivatives with hydrogen bond donor substituents in the *para* position. Additionally, the L65V mutation creates a larger binding pocket and appears to alter the angular position of the bound amino acid. This mutation is unique to *p*CNF-RS, as no other previously evolved aminoacyl-tRNA synthetases possess the valine residue at this position. Although the binding pocket is larger in *p*CNF-RS, key residues (Leu32, Val65, and Gly158) are still able to stabilize the *p*CNF (**1**) through van Der Waals interactions (Figure 4B).

To determine which mutations might be determinants of polyspecificity, we overlaid the *p*CNF-bound *p*CNF-RS amino acid binding site structure with the substrate-bound *p*AcF-RS [Y32L, D158G, I159C, and L162R (PDB entry 1ZH6)] and

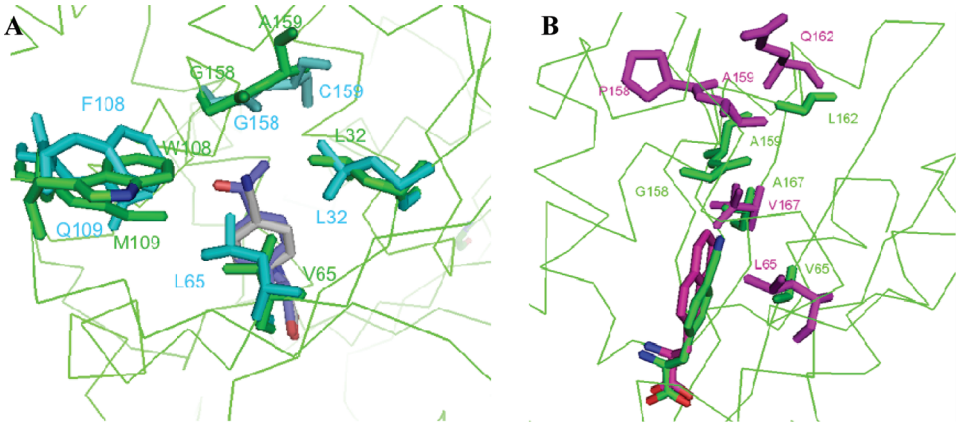

FIGURE 3: (A) Overlay of the key active site residues for substrate-bound *p*AcF-RS·*p*AcF (cyan, PDB entry 1ZH6) and *p*CNF-RS·*p*CNF (green) complexes. The *p*AcF (blue) and *p*CNF (gray) substrates are shown at the center of the structure. (B) Overlay of the substrate-bound NapA-RS·NapA (magenta, PDB entry 1ZH0) and *p*CNF-RS·*p*CNF (green) complexes. The NapA (magenta) and *p*CNF (green) substrates are shown at the bottom of the structure.

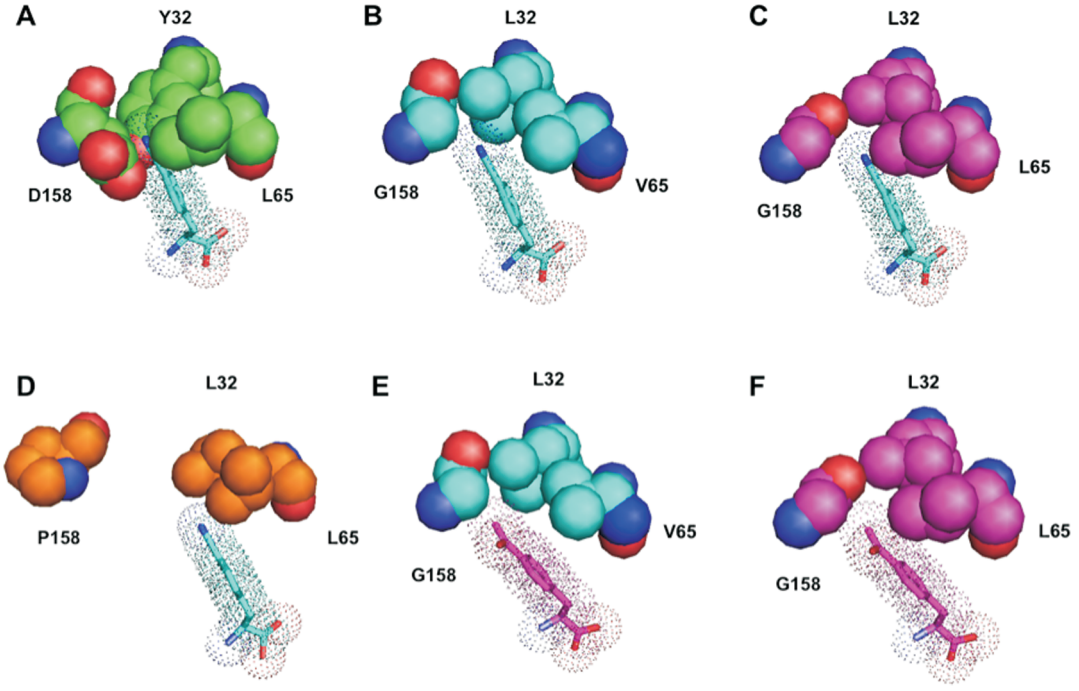

FIGURE 4: Structural analysis of three key residues (32, 65, and 158) for amino acid recognition in various synthetases: (A) wild-type tyrosyl RS (green, PDB entry 1J1U) modeled with *p*CNF, (B) *p*CNF-RS (cyan) modeled with *p*CNF, (C) *p*AcF-RS (magenta, PDB entry 1ZH6) modeled with *p*CNF, (D) NapA-RS (orange, PDB entry 1ZH0) modeled with *p*CNF, (E) *p*CNF-RS (cyan, PDB entry 3QE4) modeled with *p*AcF, and (F) *p*AcF-RS (magenta, PDB entry 1ZH6) modeled with *p*AcF.

NapA-RS [Y32L, D158P, I159A, L162Q, and A167V (PDB 1ZH0)] structures. The former aaRS accepts relatively few other unnatural amino acids, while the latter displays no polyspecificity (Figure 3). As previously noted, the Y32L mutation increases the amino acid binding site size (especially at the *para* position of phenylalanine derivatives); however, this mutation is also found in *p*AcF-RS and NapA-RS (among others), which do not exhibit a degree of polyspecificity comparable to that of *p*CNF-RS. The unique L65V mutation appears to create unoccupied space at the back of the amino acid binding site that is filled by other residues in *p*AcF-RS. This mutation may allow *p*CNF-RS to accept a diverse group of *para*-substituted phenylalanine or tyrosine derivatives (Figure 3). Interestingly, when *p*CNF is modeled into the *p*AcF-RS binding pocket, the leucine at residue 65 interferes with the cyano substituent (Figure 4C; the *p*AcF-RS complex is shown for comparison in Figure 4F). In contrast, *p*AcF (**7**) can readily be docked into the active site of *p*CNF-RS (Figure 4E). Additionally, mutation of leucine 65 to smaller amino acids has been shown to allow the incorporation of larger unnatural amino acids into other aaRSs. For example, the *o*-nitrobenzyltyrosine (ONBY) synthetase contains a L65G mutation, facilitating the incorporation of the sterically bulky bicyclic ONBY amino acid (*37*).

While an increased substrate binding site size may be partially responsible for increased polyspecificity, other factors are involved. This is apparent upon comparison of the *p*CNF-RS and NapA-RS structures. Although the D158P mutation in NapA-RS gives it a significantly larger binding pocket compared to that of *p*CNF-RS, this aaRS exhibits virtually no permissivity. When *p*CNF is modeled into the amino acid binding site of NapA-RS,

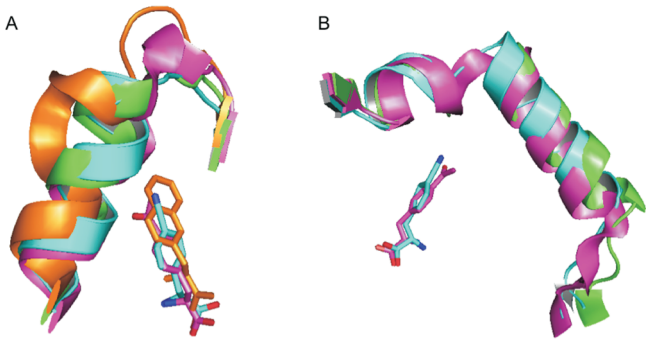

FIGURE 5: Backbone perturbations in various aminoacyl-tRNA synthetases. (A) Overlay of the helices containing mutated residues 158 and 159 in NapA-RS (orange, PDB entry 1ZH0), WT TyrRS (green, PDB entry 1J1U), *p*AcF-RS (magenta, PDB entry 1ZH6), and *p*CNF-RS (cyan). The termination of the helix that accommodates large amino acids is especially apparent in NapA-RS. (B) Overlay of the helices containing mutated residues 108 and 109 in WT TyrRS (green, PDB entry 1J1U), *p*AcF-RS (magenta, PDB entry 1ZH6), and *p*CNF-RS (cyan). *p*CNF is colored cyan, *p*AcF magenta, and NapA orange.

proline 158 prevents side chain van der Waals interactions that might stabilize the bound amino acid (Figure 4D). Moreover, a number of the other aminoacyl-tRNA synthetases examined with larger binding sites (e.g., that accommodate bipyridyl, 8-hydroxyquinolinyl, and benzoylphenyl side chains) do not exhibit significant polyspecificity, suggesting that there is a plasticity unique to the *p*CNF-RS active site.

The uniqueness of the *p*CNF-RS is also evident upon examination of backbone perturbations in the mutant aminoacyl-tRNA synthetases. As previously reported, the D158P and I159A mutations terminate the α_8-helix in NapA-RS; a similar backbone alteration is observed in the *p*AcF-RS (albeit to a lesser extent) (*12*). Interestingly, this perturbation is not observed in the *p*CNF-RS, despite having the same D158G mutation as the *p*AcF-RS and the same I159A mutation as the NapA-RS. In fact, the *p*CNF-RS aligns much more closely with the wild-type tyrosyl-tRNA synthetase than with the other aminoacyl-tRNA synthetases in this region (Figure 5A). Another distinctive difference is apparent in the *p*CNF-RS upon examination of helical residues 108 and 109. Relative to both the WT RS and the *p*AcF-RS, the helix is moderately displaced and is slightly truncated (Figure 5B). This may afford a greater degree of space and plasticity for the accommodation of larger *para*-substituted amino acids.

Importantly, the *p*CNF-RS amino acid binding site still strongly discriminates against Phe, Tyr, and *p*-aminophenylalanine, while it accepts similar amino acids such as *O*-methyltyrosine (**16**), *p*-chlorophenylalanine (**4**), and *p*-fluorophenylalanine (**5**). Clearly, mutations in a relatively small number of active site residues result in the ability of this aaRS to discriminate subtle changes in substrate structure. This result indicates that the negative selection step in the evolution of *p*CNF-RS is quite robust in its ability to remove polyspecific mutants that accept endogenous amino acids. Work is currently underway to cocrystallize *p*CNF-RS and various mutants with additional common and unnatural amino acids in an effort to further decipher this unique enzyme's specificity.

CONCLUSION

The use of previously evolved aaRSs for the incorporation of additional unnatural amino acids will facilitate the expansion of the genetic repertoire to include novel chemical functionalities. Seven amino acids that had previously not been genetically encoded in bacteria were selectively incorporated into proteins using the *p*CNF-RS synthetase. These include *p*-chlorophenylalanine, *p*-fluorophenylalanine, *p*-alkynylphenylalanine (**4**−**6**), *p*-trifluoromethylacetylphenylalanine (**9**), *p*-phenylphenylalanine (**12**), benzylserine (**15**), and *O*-*tert*-butyltyrosine (**18**). Future screens of additional amino acids and aaRSs may yield aaRSs that incorporate classes of unnatural amino acids rather than a single residue.

ACKNOWLEDGMENT

This work is based on experiments conducted at beamline 5.0.3 of the Advanced Light Source (ALS). The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences, Material Sciences Division of the U.S. Department of Energy, under Contract DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory (Berkeley, CA). We thank all of the staff of these beamlines for their continued support.

SUPPORTING INFORMATION AVAILABLE

Unnatural amino acids screened and 96-well plate format, myoglobin mass spectral data, and crystal refinement statistics. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Young, T. S., and Schultz, P. G. (2010) Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *J. Biol. Chem. 285*, 11039–11044.
2. Liu, C. C., and Schultz, P. G. (2010) An Expanding Genetic Code. *Annu. Rev. Biochem.* (in press).
3. Wang, L., and Schultz, P. G. (2001) A general approach for the generation of orthogonal tRNAs. *Chem. Biol. 8*, 883–890.
4. Santoro, S. W., Wang, L., Herberich, B., King, D. S., and Schultz, P. G. (2002) An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol. 20*, 1044–1048.
5. Stokes, A. L., Miyake-Stoner, S. J., Peeler, J. C., Nguyen, D. P., Hammer, R. P., and Mehl, R. A. (2009) Enhancing the utility of unnatural amino acid synthetases by manipulating broad substrate specificity. *Mol. Biosyst. 5*, 1032–1038.
6. Hartman, M. C., Josephson, K., Lin, C. W., and Szostak, J. W. (2007) An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. *PLoS One 2*, e972.
7. Miyake-Stoner, S. J., Refakis, C. A., Hammill, J. T., Lusic, H., Hazen, J. L., Deiters, A., and Mehl, R. A. (2010) Generating permissive site-specific unnatural aminoacyl-tRNA synthetases. *Biochemistry 49*, 1667–1677.
8. Kirshenbaum, K., Carrico, I. S., and Tirrell, D. A. (2002) Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues. *ChemBioChem 3*, 235–237.
9. Munier, R., and Cohen, G. N. (1959) [Incorporation of structural analogues of amino acids into bacterial proteins during their synthesis in vivo]. *Biochim. Biophys. Acta 31*, 378–391.
10. Liu, W., Alfonta, L., Mack, A. V., and Schultz, P. G. (2007) Structural basis for the recognition of para-benzoyl-L-phenylalanine by evolved aminoacyl-tRNA synthetases. *Angew. Chem., Int. Ed. 46*, 6073–6075.
11. Turner, J. M., Graziano, J., Spraggon, G., and Schultz, P. G. (2005) Structural characterization of a p-acetylphenylalanyl aminoacyl-tRNA synthetase. *J. Am. Chem. Soc. 127*, 14976–14977.
12. Turner, J. M., Graziano, J., Spraggon, G., and Schultz, P. G. (2006) Structural plasticity of an aminoacyl-tRNA synthetase active site. *Proc. Natl. Acad. Sci. U.S.A. 103*, 6483–6488.
13. Schultz, K. C., Supekova, L., Ryu, Y., Xie, J., Perera, R., and Schultz, P. G. (2006) A genetically encoded infrared probe. *J. Am. Chem. Soc. 128*, 13984–13985.
14. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307–326.

15. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr. D61*, 458–464.
16. Kobayashi, T., Nureki, O., Ishitani, R., Yaremchuk, A., Tukalo, M., Cusack, S., Sakamoto, K., and Yokoyama, S. (2003) Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Biol. 10*, 425–432.
17. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126–2132.
18. Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, P., Roversi, P., Sharff, A., Smart, O., Vonrhein, C., and Womack, T. (2010) BUSTER, version 2.9, Global Phasing Ltd., Cambridge, U.K.
19. Bailey, S. (1994) The Ccp4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D50*, 760–763.
20. Smart, O., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Wonrhein, C., Womack, T, and Bricogne, G. (2008) Refinement with Local Structure Similarity Restraints (LSSR) Enables Exploitation of Information from Related Structures and Facilitates use of NCS. *Abstracts of the Annual Meeting of the American Crystallographic Association,* 117.
21. Young, T. S., Ahmad, I., Yin, J. A., and Schultz, P. G. (2010) An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *J. Mol. Biol. 395*, 361–374.
22. Wang, L., Zhang, Z. W., Brock, A., and Schultz, P. G. (2003) Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A. 100*, 56–61.
23. Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) Addition of p-azido-L-phenylaianine to the genetic code of *Escherichia coli*. *J. Am. Chem. Soc. 124*, 9026–9027.
24. Xie, J. M., Wang, L., Wu, N., Brock, A., Spraggon, G., and Schultz, P. G. (2004) The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol. 22*, 1297–1301.
25. Brustad, E., Bushey, M. L., Lee, J. W., Groff, D., Liu, W., and Schultz, P. G. (2008) A Genetically Encoded Boronate-Containing Amino Acid. *Angew. Chem., Int. Ed. 47*, 8220–8223.
26. Deiters, A., and Schultz, P. G. (2005) In vivo incorporation of an alkyne into proteins in *Escherichia coli*. *Bioorg. Med. Chem. Lett. 15*, 1521–1524.
27. Xie, J. M., Supekova, L., and Schultz, P. G. (2007) A genetically encoded metabolically stable analogue of phosphotyrosine in *Escherichia coli*. *ACA Chem. Biol. 2*, 474–478.
28. Chin, J. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A. 99*, 11020–11024.
29. Wang, L., Brock, A., and Schultz, P. G. (2002) Adding L-3-(2-naphthyl)alanine to the genetic code of *E. coli*. *J. Am. Chem. Soc. 124*, 1836–1837.
30. Wang, J. Y., Xie, J. M., and Schultz, P. G. (2006) A genetically encoded fluorescent amino acid. *J. Am. Chem. Soc. 128*, 8738–8739.
31. Guo, J. T., Wang, J. Y., Anderson, J. C., and Schultz, P. G. (2008) Addition of an α-hydroxy acid to the genetic code of bacteria. *Angew. Chem., Int. Ed. 47*, 722–725.
32. Xie, J. M., Liu, W. S., and Schultz, P. G. (2007) A genetically encoded bidentate, metal-binding amino acid. *Angew. Chem., Int. Ed. 46*, 9239–9242.
33. Lee, H. S., Spraggon, G., Schultz, P. G., and Wang, F. (2009) Genetic Incorporation of a Metal-Ion Chelating Amino Acid into Proteins as a Biophysical Probe. *J. Am. Chem. Soc. 131*, 2481–2483.
34. Guo, J., Wang, J., Lee, J. S., and Schultz, P. G. (2008) Site-specific incorporation of methyl- and acetyl-lysine analogues into recombinant proteins. *Angew. Chem., Int. Ed. 47*, 6399–6401.
35. Mehl, R. A., Anderson, J. C., Santoro, S. W., Wang, L., Martin, A. B., King, D. S., Horn, D. M., and Schultz, P. G. (2003) Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc. 125*, 935–939.
36. Zhang, Y., Wang, L., Schultz, P. G., and Wilson, I. A. (2005) Crystal structures of apo wild-type *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for O-methyl-L-tyrosine. *Protein Sci. 14*, 1340–1349.
37. Deiters, A., Groff, D., Ryu, Y. H., Xie, J. M., and Schultz, P. G. (2006) A genetically encoded photocaged tyrosine. *Angew. Chem., Int. Ed. 45*, 2728–2731.