ACS chemical biology

Letters

pubs.acs.org/acschemicalbiology

Recoding Aminoacyl-tRNA Synthetases for Synthetic Biology by Rational Protein-RNA Engineering

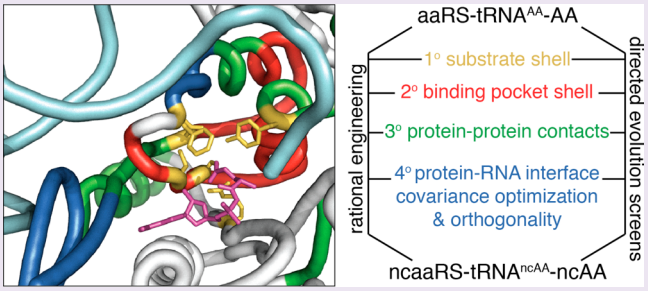
Andrew Hadd[†] and John J. Perona*[,†,‡]

[†]Department of Biochemistry & Molecular Biology, Oregon Health & Sciences University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97239, United States

[‡]Department of Chemistry, Portland State University, PO Box 751, Portland, Oregon 97207, United States

Supporting Information

ABSTRACT: We have taken a rational approach to redesigning the amino acid binding and aminoacyl−tRNA pairing specificities of bacterial glutaminyl−tRNA synthetase. The four-stage engineering incorporates generalizable design principles and improves the pairing efficiency of noncognate glutamate with $tRNA^{Gln}$ by over 10^5-fold compared to the wild-type enzyme. Better optimized designs of the protein−RNA complex include substantial reengineering of the globular core region of the tRNA, demonstrating a role for specific tRNA nucleotides in specifying the identity of the genetically encoded amino acid. Principles emerging from this engineering effort open new prospects for combining rational and genetic selection approaches to design novel aminoacyl−tRNA synthetases that ligate noncanonical amino acids onto tRNAs. This will facilitate reconstruction of the cellular translation apparatus for applications in synthetic biology.



The expansion of genetic codes through incorporation of noncanonical amino acids (ncAAs) offers substantial potential to develop designer proteins capable of yielding new insights into cellular processes.[1,2] Engineering of the translational apparatus is also a key foundational technology in synthetic biology, with possible applications to metabolic pathway reconstruction, design of antimicrobial resistance, cancer therapeutics, and other areas.[3] These aspirations depend on the creation of novel aminoacyl−tRNA synthetases (aaRS) capable of ligating one or more ncAAs to a new tRNA that reads an unused codon triplet or quadruplet—most commonly the UAG amber stop codon.[4−7] A directed evolution approach involving both positive and negative selections, applied to libraries targeting an aaRS active site, has allowed incorporation of over 100 ncAAs in bacterial or eukaryotic cells.[5,8,9] The choice of which aaRS scaffold to employ in these selections has been largely dictated by the requirement for orthogonality: neither the new aaRS nor the new tRNA may substantially cross-react with endogenous parallel aaRS−tRNA systems. In some cases, orthogonality may also require that similar positive and negative selections be carried out to optimize the tRNA sequence.[10,11]

Although this directed evolution approach has been successful in producing ncAA-containing proteins *in vivo*, better integration of the new aaRS−tRNA pairs into the translational apparatus requires further efforts. Very high concentrations of ncAAs and overexpression of tRNAs and engineered aaRS are presently needed for efficient protein synthesis, but these features diminish cellular fitness by increasing competitions with endogenous aaRS for both amino acid and tRNA pools.[6,12,13] Further, limits to the number of amino acid positions that can be sampled in libraries have generally confined explorations of aaRS structures to the immediate amino acid binding site environments. The development of multiplex automated genome engineering (MAGE) greatly streamlines library constructions and allows for introduction of mutations at many loci but does not solve the problem of where to target.[14] Limitation of the selections to a small number of aaRS scaffolds, primarily methanogen tyrosyl− and pyrrolysyl−tRNA synthetases (TyrRS; PylRS), likely further limits the range of ncAAs that can ultimately be incorporated.[8,15,16]

Weak activities of genetically selected aaRS in amino acid activation and tRNA transfer,[6,8,17] and susceptibilities of enzyme-bound noncanonical aminoacyl adenylates to competitive attack by water,[18,19] underlie the requirements for high concentrations of ncAAs and orthogonal tRNAs in selections. We suggest that these deficiencies in the catalytic performance of aaRS enzymes emerging from directed evolution may be remediated through methodologies that incorporate rational design.[6] This approach is facilitated by reliable crystal structures bound to cognate amino acid for nearly all of the 24 phylogenetically independent families of aaRS.[20] The aaRS families are partitioned into two structural classes (class I and

Received: August 19, 2014
Accepted: October 13, 2014
Published: October 13, 2014

   dx.doi.org/10.1021/cb5006596 | ACS Chem. Biol. 2014, 9, 2761−2766

class II), possessing class-specific homologous catalytic domains. Within each class, the exhaustive sequence information for aaRS (and tRNAs) then allows construction of structure-based alignments and predictions of selectivity determinants for coded amino acids. An early application of this rational design approach to a different enzyme family was the successful identification of chymotrypsin specificity determinants by transplantation of amino acids into the homologous trypsin scaffold.[21] We posited that rational design could be similarly effective in elucidating how homologous aaRS architectures have differentiated to select among the large number of amino acids present in the cell. This approach is analogous to the transplantation of proposed tRNA identity nucleotides into noncognate tRNA frameworks, which is long established in the aaRS field.[22] Insights derived from this rational engineering should then assist in formulating new approaches to obtain novel aaRS capable of encoding ncAAs for synthetic biology.

We have chosen *Escherichia coli* glutaminyl−tRNA synthetase (GlnRS) as a model class I enzyme for rational design based on extensive structural and mechanistic studies, and on its function as a relatively small 553 amino acid monomer (Figure 1A).[23−25] GlnRS requires $tRNA^{Gln}$ as a cofactor for synthesis of the activated glutaminyl adenylate intermediate and is thus properly considered as a ribonucleoprotein (RNP) enzyme.[20] Previously, we reported early steps in a multistage design process in which amino acids derived from human glutamyl−tRNA synthetase (GluRS) were introduced into the GlnRS scaffold.[26,27] GlnRS originated in early eukaryotes from the duplication and subsequent differentiation of a gene encoding a nondiscriminating archaeal GluRS ($GluRS^{ND}$) capable of aminoacylating both $tRNA^{Glu}$ and $tRNA^{Gln}$ with glutamate (Glu).[28] The presence of GlnRS in some contemporary bacterial taxa is solely a consequence of horizontal gene transfer from eukaryotes, and the enzyme is more closely related to eukaryotic than bacterial GluRS.

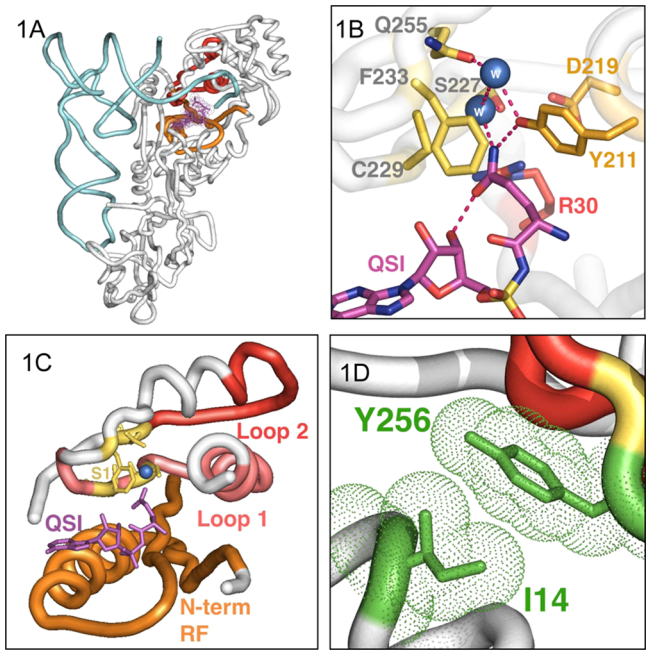


Figure 1. (A) Ribbon representation of the *E. coli* GlnRS-$tRNA^{Gln}$ complex. The tRNA is in light blue. The N-terminal and C-terminal portions of the catalytic Rossmann fold (RF) are depicted in orange and red, respectively. An analog of glutaminyl adenylate is shown bound in the active site. (B) Depiction of the hydrogen bond interactions (dotted lines) made with the glutamine substrate within the immediate binding posket (S1). Residues Q255, F233, S227, and C229 were replaced to generate the S1 mutant (R30, Y211, and D219 are conserved in eukaryotic GluRS). Two interstitial water molecules are shown as blue spheres. (C) Depiction of the RF bound to the glutaminyl−AMP analog QSI. The structurally conserved N-terminal portion of the RF (orange) forms most of the ATP binding interactions. The two exchanged loops in the L1L2 variants are shown in pink and red, respectively. The full list of mutations is provided in Table 1. (D) van der Waals contacts between Tyr256 in L2 and the distal Ile14 residue, preserved in the GlnRS L1L2 W256Y hybrid enzyme.

All GluRS possess two arginine (Arg) residues that bind the Glu substrate carboxylate group, while all GlnRS retain only one Arg (Figure 1B; Supporting Information Figure 1). We previously showed that introduction of the second Arg into GlnRS (GlnRS C229R) lowers K_m(Glu) from >750 mM to 240 mM, but without improving k_{cat}/K_m for Glu−$tRNA^{Gln}$ synthesis (which is diminished by 10^7-fold compared to Gln−$tRNA^{Gln}$ synthesis by wt GlnRS (Table 1)).[26] Most GlnRS and GluRS also possess three additional distinguishing residues in their homologous primary amino acid binding sites (S1 sites). Further introduction of these three GluRS residues into GlnRS generated the S1 GlnRS enzyme (Table 1; Figure 1B), which exhibited 20-fold improved k_{cat}/K_m for Glu−$tRNA^{Gln}$ synthesis compared to WT GlnRS. However, S1 GlnRS remains highly inefficient, with k_{cat}/K_m still reduced by nearly 10^6-fold compared to Gln−$tRNA^{Gln}$ synthesis by wt GlnRS.[26] This full replacement of all first-shell protein residues represents the first stage of the rational design effort. It demonstrated that determinants of amino acid selectivity in GlnRS must be primarily located outside of the immediate amino acid binding pocket. Neither S1 GlnRS nor the enzymes produced in the later design stages retain any detectable activity for cognate Gln−$tRNA^{Gln}$ synthesis.

In the second stage of design, we examined more distal elements of protein structure within the second half of the catalytic Rossmann fold (RF), which is primarily responsible for amino acid binding in class I aaRS (Figure 1C). Because protein sequence alignments did not identify significant conserved differences between GlnRS and GluRS enzymes outside of the S1 site (Supporting Information Figure 1), we swapped distal human GluRS peptides into S1 GlnRS based on structural considerations. Replacement of two surface loops bridging RF secondary structure elements (loops L1 and L2; Figure 1C) improved k_{cat}/K_m for Glu−$tRNA^{Gln}$ synthesis by 10^3-fold compared to S1 GlnRS and reduced K_m(Glu) by 40-fold (L1L2 GlnRS; Table 1).[26] Replacement of distal surface loops in serine proteases similarly effected conversion of amino acid selectivity in the context of peptide bond hydrolysis, providing inspiration for these experiments.[21]

In the third design stage, we examined all interactions made by the 23 replaced amino acids in L1L2 GlnRS with surrounding regions of the protein structure.[27] This allowed assessment of the role of selected third-shell residues in amino acid substrate discrimination and of the potential for non-complementarity between the introduced mutations and surrounding regions of the protein. By this process, we found that the W256Y mutation, representing a reversion back to the GlnRS sequence, improved K_m(tRNA) and K_m(Glu) each by 2-fold compared to L1L2 GlnRS—probably by alleviating a steric clash with nearby Ile14 (Table 1; Figure 1D).[27] K_m(Glu) in

dx.doi.org/10.1021/cb5006596 | *ACS Chem. Biol.* 2014, 9, 2761−2766

Table 1. Steady State Aminoacylation Parameters

		k_{cat} (s^{-1})	K_M [tRNA] (μM)	k_{cat}/K_M [tRNA] (s^{-1} M^{-1})	K_M [GLU] (mM)	k_{cat}/K_M [GLU] (s^{-1} M^{-1})
E. coli GlnRS[a]	$tRNA^{GLN}$	0.046 ± 0.013	19 ± 3	2.4×10^3	>750	9.5×10^{-4}
C229R GlnRS[b]	$tRNA^{GLN}$	3.2 ± 0.1 x10^{-4}			240 ± 10	1.3×10^{-3}
S1 GlnRS[b]	$tRNA^{GLN}$	5.0 ± 0.6 x10^{-3}			230 ± 17	2.2×10^{-2}
L1L2[b]	$tRNA^{GLN}$	0.09 ± 0.02	7.6 ± 3.0	1.2×10^4	5.8 ± 0.5	15.5
L1L2 W256Y[c]	$tRNA^{GLN}$	0.10 ± 0.04	4.8 ± 0.2	2.1×10^4	2.6 ± 1.0	38.5
L1L2 W256Y[d]	Q/E tRNA1	0.014 ± 0.01	1.2 ± 0.4	1.2×10^4	14 ± 7	1.0
L1L2 W256Y	Q/E tRNA2	0.05 ± 0.01	1.4 ± 0.7	3.6×10^4	53 ± 22	0.9
L1L2 W256Y	Q/E tRNA3	0.32 ± 0.05	2.4 ± 0.8	1.3×10^5	1.9 ± 0.2	168.4
L1L2 W256Y	Q/E tRNA4	0.069 ± 0.01	0.74 ± 0.07	9.3×10^4	3.3 ± 0.6	20.9
L1L2 R237D/R238E/W256Y	Q/E tRNA1	0.092 ± 0.01	3.4 ± 0.7	2.7×10^4	12 ± 5	7.7
L1L2 R237D/R238E/W256Y	Q/E tRNA4	0.29 ± 0.05	3.4 ± 0.5	8.5×10^4	3.0 ± 0.7	97.7
benchmark enzymes						
E. coli GlnRS + GLN[a]	$tRNA^{GLN}$	3.2 ± 0.5	0.31 ± 0.09	1.0×10^7	0.26 ± 0.04[a]	1.2×10^4
M. thermautotrophicus $GluRS^{ND(c)}$	$tRNA^{GLN}$	0.12 ± 0.01	0.038 ± 0.010	3.2×10^6	6.2 ± 0.6	19.4
S. cerevisiae $GluRS^{D(e)}$	$tRNA^{GLU}$	1.6 ± 0.6	0.16 ± 0.03	1.0×10^7	39 ± 14	41.0

[a]Reported in ref 25. All values in this table reflect glutamylation except for the benchmark represented by *E. coli* GlnRS for its cognate glutaminylation reaction (*E. coli* GlnRS + GLN). The value 0.26 ± 0.04 represents K_M[GLN]. No activity with GLN as a substrate is detectable for S1 GlnRS or for any of the L1L2 hybrids. [b]Reported in ref 26. S1 GlnRS includes the following mutations: C229R/Q255I/S227A/F233Y. S1L1L2 GlnRS adds the following mutations: T214A/H215C/C216P/S218V/A220S/L221I/I224V/L231T/V243I/L244I/D245E/N246A/I247L/T248G/P250R/ΔV251/H252K/R254Y/Y256W [c]Reported in ref 27 [d]Sequences of hybrid Q/E tRNAs are provided in Figure 2 and the Supporting Information. [e]Reported in ref 32.

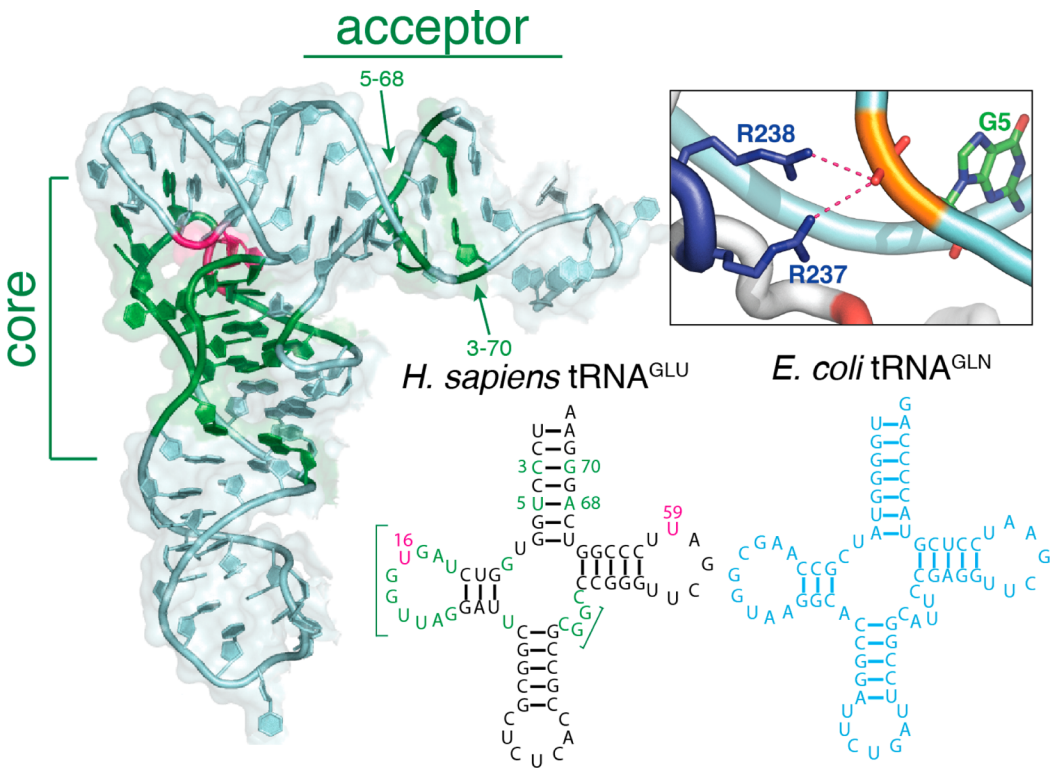


Figure 2. tRNA structure and protein interactions. Left: tertiary structure of $tRNA^{Gln}$ in the conformation bound to GlnRS. The 3−70 and 5−68 base pairs replaced in the acceptor stem are shown in green. The C16 and A59 core region nucleotides, replaced with U16 and U59 in some of the hybrid RNPs, are shown in red, while the remainder of the replaced tRNA core region is shown in green and consists of the C9G substitution together with replacements of the full D and variable loops. Upper right: Structural details of the protein−tRNA interface showing the backbone contacts made at the G5 phosphate of $tRNA^{Gln}$ by Arg237 and A238 (dark blue). Bottom right: secondary structure depictions of human $tRNA^{Glu}$ and *E. coli* $tRNA^{Gln}$.

L1L2W256Y GlnRS is well below the intracellular Glu concentration in *E. coli* and is lower than measured for yeast GluRS and archaeal $GluRS^{ND}$ (Table 1).[27,29] The k_{cat} in L1L2W256Y GlnRS is identical to $GluRS^{ND}$, although it remains 15- to 30-fold lower than that of *E. coli* GlnRS or yeast GluRS. However, K_m(tRNA) is 15- to 125-fold higher than measured in cognate aminoacylation by any of the three benchmark enzymes annotated in Table 1, despite the fact that

dx.doi.org/10.1021/cb5006596 | ACS Chem. Biol. 2014, 9, 2761−2766

no mutations at the $tRNA^{Gln}$ interface were introduced. Therefore, while alterations to protein structure alone fully recreate a functional Glu binding site in GlnRS, the resulting hybrid enzyme remains deficient in pairing amino acid with tRNA. These observations underscore the interdependence of amino acid and tRNA binding in GlnRS,[24,30] as previously illustrated by the sharply elevated K_m(tRNA) for Glu−$tRNA^{Gln}$ synthesis by the wt enzyme (Table 1).[31] Although comparable protein engineering experiments have not been performed with other aaRS, it is likely that such substrate interdependence is embedded in the architectures of all aaRS from both structural classes.[20]

These observations led us to consider a fourth rational design stage to explore the possibility that RNA structure is required for amino acid coding. Sequence comparisons among bacterial $tRNA^{Gln}$ and eukaryotic $tRNA^{Glu}$ species reveal conserved differences in the acceptor arm at positions 3−70 and 5−68, and in the globular core at positions 16 and 59 in the T and D loops, respectively (Figure 2A,B).[32] We began by pairing GlnRS L1L2W256Y with a mutant $tRNA^{Gln}$ in which the G3−C70 and G5−C68 pairs were altered to C3−G70 and U5−A68, respectively (Q/E_tRNA1; Figure 2). Remarkably, K_m(tRNA) is reduced 4-fold by incorporation of these $tRNA^{Glu}$ acceptor-stem base pairs, suggesting improved complementarity at the protein−RNA interface in a region close to the introduced eukaryotic GluRS peptides (Table 1; Supporting Information Figure 2). However, k_{cat} for Glu−$tRNA^{Gln}$ synthesis is diminished 7-fold while K_m(Glu) is elevated by 5-fold in this hybrid RNP, thus diminishing its overall catalytic performance.

These findings suggest an allosteric mechanism connecting protein−RNA contacts at base pairs 3−70 and 5−68 in the acceptor stem, with both the Glu binding site and the catalytic center where the two-step aminoacylation reaction occurs. To test this, we examined distinguishing bacterial GlnRS and eukaryotic GluRS interactions at these acceptor-stem positions. Arg237 and Arg238 in *E. coli* GlnRS, which interact with the sugar−phosphate backbone at nucleotide G5 (Figure 2), are replaced by Asp237 and Glu238 in GluRS. We therefore constructed the L1L2W256Y/R237D/R238E GlnRS enzyme and examined its capacity to glutamylate Q/E_tRNA1. k_{cat} for Glu−$tRNA^{Gln}$ synthesis is fully reconstituted in this RNP, and K_m(Glu) is unchanged, but K_m(tRNA) is elevated 3-fold (Table 1). The reconstitution of k_{cat} indeed suggests intramolecular signaling of acceptor-stem backbone interactions to the catalytic center, but the other measurements indicate that this subregion of the protein−RNA interface does not operate as an independent module but instead depends on other portions of the RNP for its function.

Next, we examined the role of the tRNA core domain. Introducing the C16U and A59U mutations into $tRNA^{Gln}$, to generate the Q/E_tRNA2 species for pairing with GlnRS L1L2W256Y, again improves K_m(tRNA) but diminishes k_{cat} while sharply elevating K_m(Glu) above 50 mM. This behavior is qualitatively similar to that observed with Q/E_tRNA1 (Table 1). However, the sizes of the D and variable loops differ between bacterial $tRNA^{Gln}$ and eukaryotic $tRNA^{Glu}$, suggesting that C16U/A59U has a disrupted tRNA structure. We therefore reconstituted the full human $tRNA^{Glu}$ core region by introducing nine additional alterations in the D and variable loops to generate Q/E_tRNA3 (Figure 2). Kinetic analysis shows that L1L2W256Y GlnRS paired with Q/E_tRNA3 is significantly improved for Glu coding compared to all prior RNP designs. Compared to L1L2W256Y paired with wt $tRNA^{Gln}$, k_{cat} for glutamylation is increased 3-fold, K_m(tRNA) is improved by 2-fold, and K_m(Glu) is maintained at about 2 mM (Table 1). These experiments demonstrate that the $tRNA^{Glu}$ core region is an RNA determinant that specifies Glu for coding within the general architecture of class I aaRS RNPs.

Combining the acceptor stem and full tRNA core region replacements into one RNA (Q/E_tRNA4) has distinct effects on tRNA versus Glu-related kinetic parameters. K_m(tRNA) shows substantial additivity between the two "single" substitutions in the acceptor stem and core (Q/E_tRNA1 and Q/E_tRNA3, respectively) and the "double" substitution in Q/E_tRNA4 (see Supporting Information Figure 3). Therefore, enzyme−tRNA interactions at the inner elbow and acceptor stem regions appear to be largely independent of each other with respect to a Michaelis parameter that approximates tRNA binding. Because of the additive effects, K_m(tRNA) for the RNP composed of GlnRS L1L2W256Y paired with Q/E_tRNA4 falls below 1 μM and is elevated by just 2-fold compared to the native GlnRS−$tRNA^{Gln}$ interaction (Table 1). This demonstrates that rational design can be successfully applied to adjust this key steady-state kinetics parameter into a physiologically relevant range commensurate with intracellular tRNA concentrations.

In contrast, the GlnRS L1L2W256Y:Q/E_tRNA4 RNP exhibits little additivity with respect to K_m(Glu); instead, the "double" substitution maintains a relatively low value for this parameter that primarily reflects the influence of the core region (Table 1; Supporting Information Figure 3). It is remarkable that the more distal $tRNA^{Glu}$-like core dominates the formation of the Glu binding site, since the introduction of positive RNA determinants in this region blocks the disruptive effect of the proximal acceptor stem base-pair swaps in Q/E_tRNA1. These data again emphasize that long-range allosteric signaling from the tRNA core region interface to the Glu binding site is a key aspect of Glu coding in this bacterial system.

Unlike all prior experiments based on subtractive mutagenesis, this rational engineering approach definitively demonstrates that RNA determinants are required for efficient amino acid−RNA pairing, at both k_{cat} and K_m(tRNA) levels. Such RNA elements may represent vestiges of an early RNA world in which amino acid coding for protein synthesis was facilitated by direct amino acid−RNA interactions.[33] The exclusive role of protein in forming the amino acid binding pocket of contemporary aaRS RNPs obscures this function of the RNA, and indeed we have shown that $tRNA^{Glu}$ nucleotides are not required to optimize K_m(Glu) in the engineered enzymes. However, the protein binding site for Glu does not function optimally in RNA pairing unless RNA nucleotides that encode Glu are also included in the RNP. We suggest that optimization of allosteric linkages operating between the mutated tRNA core region and acceptor stem base pairs, by rational design of the intervening protein elements (Figure 3), should further improve function. k_{cat}/K_m(tRNA) for the engineered RNPs remains 30−100-fold below that of the naturally occurring benchmark enzymes (Table 1), and improving this parameter likely requires repair of protein structure in peptides adjacent to the R237D/R238E substitutions (Figure 2).

Our findings provide insights that should be useful in creating new aaRS that encode ncAAs for synthetic biology applications (Figure 3). Randomizing the local amino acid binding sites of archaeal TyrRS and PylRS generates enzymes capable of ncAA incorporation, an outcome that likely rests on

dx.doi.org/10.1021/cb5006596 | ACS Chem. Biol. 2014, 9, 2761−2766

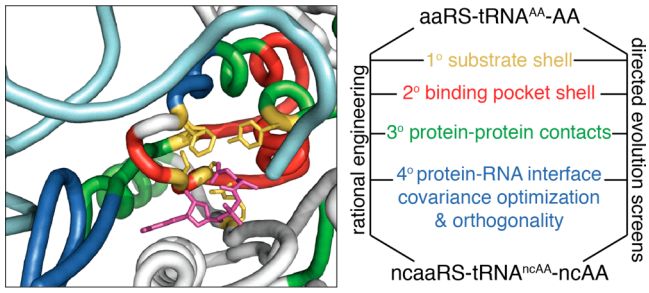


Figure 3. General strategy for rational design of amino acid specificity in the aaRS, illustrated in *E. coli* GlnRS: (i) the S1 site is depicted in yellow; (ii) the distal L1 and L2 loops are depicted in red; (iii) surrounding regions of the protein structure that contact S1, L1, and L2 are depicted in green; (iv) peptides that directly contact tRNA in regions that assist amino acid−RNA pairing specificity are depicted in dark blue. Binding of the enzyme at the tRNA inner elbow is shown at the bottom left; the separate contacts at the tRNA acceptor stem are shown at the top center. The tRNA backbone is shown in light blue, and the glutaminyl adenylate analog is shown at bottom center in magenta. All four elements of the rational design approach should be considered in the context of directed evolution experiments that make use of both wild-type and rationally engineered enzyme and tRNA scaffolds. Details of experimental strategies will depend on the particular characteristics of each orthogonal system.

the fact that the aaRS scaffolds were not under selective pressure to resist these ncAAs. However, our experiments suggest that directed evolution will not succeed in generating highly efficient enzymes unless the second/third shell protein residues and the allostery between amino acid and tRNA substrates are also considered. All aaRS are large multidomain proteins that bind tRNA in a mutual induced fit process that likely proceeds by formation of an initial encounter complex, followed by first-order rearrangements that may also be affected by the binding of ATP and amino acid.[20] These common features observed in X-ray structures predict that allosteric functional linkages between tRNA and amino acid binding, and the broad dispersal of amino acid specificity determinants, will also be prevalent if not universal among the aaRS.

Although the design of ncAA-coding aaRS cannot benefit from structure/sequence databases, one approach that may allow generation of more catalytically proficient enzymes could be to rationally engineer sets of hybrid methanogen TyrRS and PylRS that incorporate distal structural elements from other homologous class 1c and class 2c aaRS, respectively. This would substantially expand the number of subtly distinct scaffolds for use in genetic selections (Figure 3). For example, structural elements from the closely homologous class Ic TrpRS could be swapped into the TyrRS scaffold to generate hybrids capable of Trp-tRNA^{Tyr} synthesis; these hybrids would then provide a starting point for directed evolution. The exquisite sensitivities of naturally occurring aaRS to mutational perturbation suggests that even relatively small differences in active site structures and dynamics, within families of such rationally engineered hybrid enzymes, could generate significant differences in capacities for ncAA incorporation when subjected to directed evolution. Systematic rational engineering of methanogen TyrRS and PylRS along the lines described here would allow assessment of the nature and spatial distribution of amino acid specificity determinants, and the extent to which they may reside in the RNA. Precise choices of residues to be randomized, within the context of rationally engineered scaffolds, will clearly best be made when a quantitative database of structure−function information is available.

A second insight arising from these experiments is that directed evolution strategies that include selections for optimizing orthogonality in the tRNA should account for the possibility that the derived tRNAs may acquire nucleotides that render them less efficient for incorporation of a desired ncAA. This notion emerges from our finding that tRNA^{Glu} core elements are positive determinants for encoding Glu. Careful measurements of kinetic parameters for aminoacylation of ncAAs by selected RNPs are required to evaluate this and other detailed properties of ncAA-encoding aaRS.[34] RNA and protein sequences in ncAA-encoding RNPs should be optimized for both orthogonality and efficiency of aminoacylation.

Finally, it is worth noting that none of the efficiently functioning hybrid RNPs constructed in our studies alter any part of the tRNA anticodon arm, despite the important role for tRNA anticodon loop nucleotides in specifying tRNA^{Glu} and tRNA^{Gln} identities.[24,30] Although further experiments are required to fully evaluate this notion, it appears that anticodon recognition by GlnRS may be less fully integrated into the protein−RNA allosteric network than we have previously suggested.[24] Perhaps either GlnRS-like or GluRS-like anticodon recognition may be substantially compatible with coding of either amino acid. The relative independence of anticodon recognition is beneficial to the further development of orthogonal translation systems.

METHODS

Construction and expression of mutant enzymes was performed as described previously.[26,27] tRNAs were generated by *in vitro* transcription from templates assembled from overlapping nucleotides and were refolded and further purified as described.[35] Modified nucleotides do not have kinetic effects in this system.[24] All enzyme assays utilized tRNAs that are ^{32}P-labeled at the 3′-internucleotide linkage via the exchange activity of tRNA nucleotidyltransferase.[36] Plateau aminoacylation values were 65% or better for all measurements. Details of experimental approaches and representative primary data, protein sequence alignments, discussions of the rationale for the kinetics methodology chosen, and calculations of free energy additivities may be found in the associated Supporting Information.

ASSOCIATED CONTENT

Supporting Information

Details of experimental approaches and representative primary data, protein sequence alignments, discussions of the rationale for the kinetics methodology chosen, and calculations of free energy additivities. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: perona@pdx.edu. Tel.: 503-725-2426.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (GM63713).

DEDICATION

This paper is dedicated to the memory of Martin John Rogers, outstanding scientist and public servant, who made major contributions to our understanding of bacterial GlnRS.

REFERENCES

(1) Davis, L., and Chin, J. W. (2012) Designer proteins: applications of genetic code expansion in cell biology. *Nat. Rev. Mol. Cell Biol. 13*, 168−182.

(2) Neumann, H. (2012) Rewiring translation - Genetic code expansion and its applications. *FEBS Lett. 586*, 2057−2064.

(3) Cheng, A. A., and Lu, T. K. (2012) Synthetic biology: an emerging engineering discipline. *Annu. Rev. Biomed Eng. 14*, 155−178.

(4) Chin, J. W. (2014) Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem. 83*, 379−408.

(5) Liu, C. C., and Schultz, P. G. (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem. 79*, 413−444.

(6) O'Donoghue, P., Ling, J., Wang, Y. S., and Soll, D. (2013) Upgrading protein synthesis for synthetic biology. *Nat. Chem. Biol. 9*, 594−598.

(7) Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., and Chin, J. W. (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature 464*, 441−444.

(8) Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of Escherichia coli. *Science 292*, 498−500.

(9) Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z., and Schultz, P. G. (2003) An expanded eukaryotic genetic code. *Science 301*, 964−967.

(10) Wang, L., and Schultz, P. G. (2001) A general approach for the generation of orthogonal tRNAs. *Chem. Biol. 8*, 883−890.

(11) Liu, D. R., Magliery, T. J., Pastrnak, M., and Schultz, P. G. (1997) Engineering a tRNA and aminoacyl−tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc. Natl. Acad. Sci. U. S. A. 94*, 10092−10097.

(12) Fan, C., Ho, J. M., Chirathivat, N., Soll, D., and Wang, Y. S. (2014) Exploring the substrate range of wild-type aminoacyl−tRNA synthetases. *ChemBioChem 15*, 1805−1809.

(13) Bianco, A., Townsley, F. M., Greiss, S., Lang, K., and Chin, J. W. (2012) Expanding the genetic code of Drosophila melanogaster. *Nat. Chem. Biol. 8*, 748−750.

(14) Gallagher, R. R., Li, Z., Lewis, A. O., and Isaacs, F. J. (2014) Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA. *Nat. Protoc. 9*, 2301−2316.

(15) Neumann, H., Peak-Chew, S. Y., and Chin, J. W. (2008) Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat. Chem. Biol. 4*, 232−234.

(16) Mukai, T., Kobayashi, T., Hino, N., Yanagisawa, T., Sakamoto, K., and Yokoyama, S. (2008) Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun. 371*, 818−822.

(17) Nehring, S., Budisa, N., and Wiltschi, B. (2012) Performance analysis of orthogonal pairs designed for an expanded eukaryotic genetic code. *PLoS One 7*, e31992.

(18) Guo, L. T., Helgadottir, S., Soll, D., and Ling, J. (2012) Rational design and directed evolution of a bacterial-type glutaminyl−tRNA synthetase precursor. *Nucleic Acids Res. 40*, 7967−7974.

(19) Dulic, M., Cvetesic, N., Perona, J. J., and Gruic-Sovulj, I. (2010) Partitioning of tRNA-dependent editing between pre- and post-transfer pathways in class I aminoacyl−tRNA synthetases. *J. Biol. Chem. 285*, 23799−23809.

(20) Perona, J. J., and Hadd, A. (2012) Structural diversity and protein engineering of the aminoacyl−tRNA synthetases. *Biochemistry 51*, 8705−8729.

(21) Hedstrom, L., Szilagyi, L., and Rutter, W. J. (1992) Converting trypsin to chymotrypsin: the role of surface loops. *Science 255*, 1249−1253.

(22) Giege, R., Sissler, M., and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res. 26*, 5017−5035.

(23) Rould, M. A., Perona, J. J., Soll, D., and Steitz, T. A. (1989) Structure of E. coli glutaminyl−tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 A resolution. *Science 246*, 1135−1142.

(24) Rodriguez-Hernandez, A., and Perona, J. J. (2011) Heat maps for intramolecular communication in an RNP enzyme encoding glutamine. *Structure 19*, 386−396.

(25) Bullock, T. L., Uter, N., Nissan, T. A., and Perona, J. J. (2003) Amino acid discrimination by a class I aminoacyl−tRNA synthetase specified by negative determinants. *J. Mol. Biol. 328*, 395−408.

(26) Bullock, T. L., Rodriguez-Hernandez, A., Corigliano, E. M., and Perona, J. J. (2008) A rationally engineered misacylating aminoacyl−tRNA synthetase. *Proc. Natl. Acad. Sci. U. S. A. 105*, 7428−7433.

(27) Rodriguez-Hernandez, A., Bhaskaran, H., Hadd, A., and Perona, J. J. (2010) Synthesis of Glu−tRNA(Gln) by engineered and natural aminoacyl−tRNA synthetases. *Biochemistry 49*, 6727−6736.

(28) Woese, C. R., Olsen, G. J., Ibba, M., and Soll, D. (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol. Biol. Rev. 64*, 202−236.

(29) Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat. Chem. Biol. 5*, 593−599.

(30) Ibba, M., Hong, K. W., Sherman, J. M., Sever, S., and Soll, D. (1996) Interactions between tRNA identity nucleotides and their recognition sites in glutaminyl−tRNA synthetase determine the cognate amino acid affinity of the enzyme. *Proc. Natl. Acad. Sci. U. S. A. 93*, 6953−6958.

(31) Uter, N. T., Gruic-Sovulj, I., and Perona, J. J. (2005) Amino acid-dependent transfer RNA affinity in a class I aminoacyl−tRNA synthetase. *J. Biol. Chem. 280*, 23966−23977.

(32) Hadd, A., Perona, J. J. (2014) Coevolution of specificity determinants in eukaryotic glutamyl- and glutaminyl−tRNA synthetases, *J. Mol. Biol.* Aug 19. doi: 10.1016/j.jmb.2014.08.006. PMID: 25149203 [Epub ahead of print].

(33) de Duve, C. (1988) Transfer RNAs: the second genetic code. *Nature 333*, 117−118.

(34) Francklyn, C. S., First, E. A., Perona, J. J., and Hou, Y. M. (2008) Methods for kinetic and thermodynamic analysis of aminoacyl−tRNA synthetases. *Methods 44*, 100−118.

(35) Sherlin, L. D., Bullock, T. L., Nissan, T. A., Perona, J. J., Lariviere, F. J., Uhlenbeck, O. C., and Scaringe, S. A. (2001) Chemical and enzymatic synthesis of tRNAs for high-throughput crystallization. *RNA 7*, 1671−1678.

(36) Wolfson, A. D., and Uhlenbeck, O. C. (2002) Modulation of tRNAAla identity by inorganic pyrophosphatase. *Proc. Natl. Acad. Sci. U. S. A. 99*, 5965−5970.