Rational design of an evolutionary precursor of glutaminyl-tRNA synthetase

Patrick O'Donoghue[a], Kelly Sheppard[b], Osamu Nureki[c], and Dieter Söll[a,d,1]

[a]Departments of Molecular Biophysics and Biochemistry, and [d]Chemistry, Yale University, New Haven, CT 06520; [b]Department of Chemistry, Skidmore College, Saratoga Springs, NY 12866; and [c]Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Bunkyo-ku, Tokyo 113-0032, Japan

Contributed by Dieter Söll, October 20, 2011 (sent for review September 9, 2011)

The specificity of most aminoacyl-tRNA synthetases for an amino acid and cognate tRNA pair evolved before the divergence of the three domains of life. Glutaminyl-tRNA synthetase (GlnRS) evolved later and is derived from the archaeal-type nondiscriminating glutamyl-tRNA synthetase (GluRS), an enzyme with relaxed tRNA specificity capable of forming both Glu-tRNA^{Glu} and Glu-tRNA^{Gln}. The archaea lack GlnRS and use a specialized amidotransferase to convert Glu-tRNA^{Gln} to Gln-tRNA^{Gln} needed for protein synthesis. We show that the *Methanothermobacter thermautotrophicus* GluRS is active toward tRNA^{Glu} and the two tRNA^{Gln} isoacceptors the organism encodes, but with a significant catalytic preference for tRNA$^{Gln2}_{CUG}$. The less active tRNA$^{Gln1}_{UUG}$ responds to the less common CAA codon for Gln. From a biochemical characterization of *M. thermautotrophicus* GluRS variants, we found that the evolution of tRNA specificity in GlnRS could be recapitulated by converting the *M. thermautotrophicus* GluRS to a tRNA^{Gln} specific enzyme, solely through the addition of an acceptor stem loop present in bacterial GlnRS. One designed GluRS variant is also highly specific for the tRNA$^{Gln2}_{CUG}$ isoacceptor, which responds to the CAG codon, and shows no activity toward tRNA$^{Gln1}_{UUG}$. Because it is now possible to eliminate particular codons from the genome of *Escherichia coli*, additional codons will become available for genetic code engineering. Isoacceptor-specific aminoacyl-tRNA synthetases will enable the reassignment of more open codons while preserving accurate encoding of the 20 canonical amino acids.

The genetic code depends on the catalytic action of the aminoacyl-tRNA synthetases (aaRSs) that are responsible for accurately ligating amino acids to their cognate tRNAs. The high fidelity of the genetic code is derived principally from the exclusive interaction between an aaRS and its cognate tRNA, and this property of "orthogonality" has been exploited to expand the genetic codes of bacterial and eukaryotic cells to encode noncanonical amino acids (reviewed in refs. 1 and 2).

In nature, some aaRSs evolved and are selectively maintained to promiscuously recognize more than one tRNA species. Selenocysteine (Sec) is biosynthesized on its tRNA from a Ser-tRNA^{Sec} precursor, which is generated by a regular seryl-tRNA synthetase (SerRS) that ligates serine to tRNA^{Sec} and to tRNA^{Ser}. Glutamine and asparagine are also biosynthesized on their tRNAs in many organisms (3). In these organisms, Gln-tRNA^{Gln} (4) and Asn-tRNA^{Asn} (5) are synthesized by the action of specialized amidotransferase enzymes (6) from the respective precursors Glu-tRNA^{Gln} and Asp-tRNA^{Asn}. In both cases, the seemingly misacylated precursor aminoacyl-tRNA is formed by a nondiscriminating aaRS—i.e., a glutamyl-tRNA synthetase (GluRS) (7) that glutamylates both tRNA^{Glu} and tRNA^{Gln} or an aspartyl-tRNA synthetase (AspRS) (8) that forms both Asp-tRNA^{Asp} and Asp-tRNA^{Asn}.

There are two types of nondiscriminating GluRS (ND-GluRS). All GluRSs share the class I Rossman fold catalytic domain, but the bacterial GluRS has an α-helical bundle anticodon binding domain that is unrelated to the dual ß-barrel anticodon binding domain found in all glutaminyl-tRNA synthetases (GlnRSs) and archaeal and eukaryotic GluRSs. Although the bacterial ND-GluRS is well characterized (9, 10), the distinct architecture of the archaeal ND-GluRS indicates that its mechanism of dual tRNA recognition is also distinct and yet uncharacterized.

We recently described the crystal structure of an archaeal ND-GluRS (11). Structural comparison of the *Methanothermobacter thermautotrophicus* ND-GluRS with the *Escherichia coli* GlnRS tRNA^{Gln} complex (12) revealed two significant loops in the protein that could contribute to the high specificity of GlnRS for tRNA^{Gln}, whereas their absence in the archaeal ND-GluRS may allow the enzyme to recognize both tRNA^{Glu} and tRNA^{Gln} (11). To investigate tRNA recognition by the archaeal ND-GluRS, we determined the enzyme kinetics of Glu-tRNA^{Glu} and Glu-tRNA^{Gln} formation for the wild-type *M. thermautotrophicus* ND-GluRS (WT-GluRS) and rationally designed GluRS variants that were engineered to selectively aminoacylate tRNA^{Gln}. GlnRS evolved from an ancestor closely similar to the archaeal ND-GluRS, so the designed GluRSs represent plausible intermediate forms preceding GlnRS evolution.

Results

Biochemical Characterization of the *M. thermautotrophicus* Nondiscriminating GluRS. Because the catalytic preference of the archaeal-type nondiscriminating GluRS has not been documented, we measured the kinetic constants of the ND-GluRS toward its two tRNA^{Gln} isoacceptors (tRNA$^{Gln1}_{UUG}$, tRNA$^{Gln2}_{CUG}$) and tRNA^{Glu} (Fig. 1).

The WT-GluRS showed greatest activity toward the tRNA^{Gln2} isoacceptor with a $k_{cat} = 0.41 \pm 0.04$ s^{-1} and $K_M = 1.33 \pm 0.10$ μM (Table 1). All tRNA substrates used in this study were in vitro produced transcripts (see *Methods*). Although in some cases lacking base modifications can lead to inactive tRNAs (10), the kinetic values measured for WT-GluRS are within the typical range for aaRSs and indicate the role of modified bases is not critical for WT-GluRS activity. The kinetic constants of WT-GluRS for in vitro transcribed tRNA^{Gln2} were recently measured (13), showing the enzyme to be marginally (8.5 ± 3.6-fold) less efficient in our experiments.

Although still a nondiscriminating enzyme, WT-GluRS shows a 24-fold catalytic preference for tRNA^{Gln2} over tRNA^{Glu} (Table 1). This preference did not manifest in the plateau charging reaction (Fig. 2*A*). Compared to tRNA^{Gln2}, WT-GluRS has a 2.7-fold greater K_M for tRNA^{Glu} and an even greater reduction (10-fold) in enzyme turnover rate. Surprisingly, tRNA^{Gln1} is a far less catalytically competent substrate for WT-GluRS (Fig. 2*A* and Table 1). Whereas the K_M is only 1.5-fold greater for tRNA^{Gln1} than tRNA^{Glu}, the larger effect is again on k_{cat} (62-fold lower for tRNA^{Gln1} than tRNA^{Gln2}). WT-GluRS displays a marked catalytic preference (257-fold) for tRNA^{Gln2} over tRNA^{Gln1}.

Author contributions: P.O., K.S., O.N., and D.S. designed research; P.O. performed research; P.O. contributed new reagents/analytic tools; P.O. analyzed data; and P.O. and K.S. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: dieter.soll@yale.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1117294108/-/DCSupplemental.

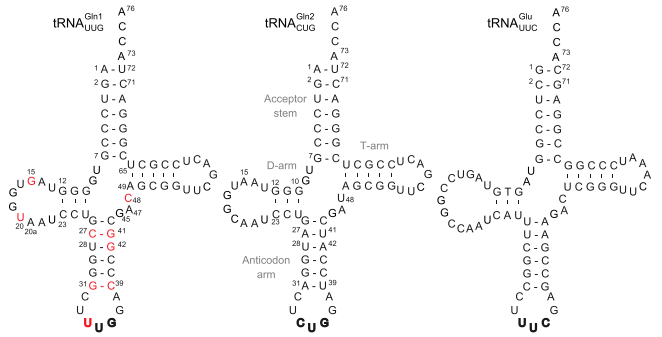

Fig. 1. Cloverleaf structures of the *M. thermautotrophicus* tRNA^{Gln} and tRNA^{Glu} molecules. Sequence differences between the tRNA^{Gln} isoacceptors (tRNA^{Gln1} has the UUG anticodon and tRNA^{Gln2} has the CUG anticodon) are highlighted in red on tRNA^{Gln1}. The structures are annotated with canonical tRNA numbering. The anticodon sequences are shown in bold.

Rational Design of a WT-GluRS Variant Specific for tRNA^{Gln}. GlnRS evolved from an ancestor of the eukaryotic/archaeal-type non-discriminating GluRS (11, 14–17). A series of GluRS mutant enzymes were designed in order to define the molecular basis by which WT-GluRS achieves relaxed tRNA specificity and to demonstrate how a tRNA-specific enzyme can evolve from a non-discriminating ancestor.

A comparison of the WT-GluRS structure (11) with the *E. coli* GlnRS tRNA^{Gln} complex (12) revealed two loops in GlnRS that contact tRNA^{Gln} and are absent from the ND-GluRS structure (Fig. 3). One loop (*E. coli* GlnRS 134–140) found in the acceptor binding domain (CP1-domain in other aaRSs) apparently acts to disrupt the first base pair of the acceptor stem (Figs. 3*C* and 4*A*), allowing the protein to recognize nucleotide identity elements in the second and third base pairs of the acceptor stem (12). The second loop is found at the opposite end of the protein (*E. coli* GlnRS 392–408) and includes Arg402 that makes direct contact with base 36 of the tRNA, a distinguishing feature between tRNA^{Gln} (^{34}YUG^{36} anticodon) and tRNA^{Glu}(^{34}YUC^{36} anticodon) (Figs. 3*D* and 4*B*). Arg402 is almost completely conserved among GlnRS sequences (Fig. S1*B*).

We attempted to design a variant GluRS specific for tRNA^{Gln} by inserting one or both of the anticodon (AL) or acceptor stem (ASL) loops into the WT-GluRS (Fig. 3). We constructed a number of variants including a larger or smaller insertion of the relevant GlnRS peptide (Fig. 4). Although some (AL1, AL2, AL4) displayed low levels of protein production, a number of variants were highly produced and active in Glu-tRNA formation.

Standard aminoacylation plateau charging curves were used to screen the WT-GluRS variants for enhanced activity toward tRNA^{Gln2} and/or decreased activity toward tRNA^{Glu} (Fig. 2 *B–D*). The AL3 (Fig. 2*B*) and AL5 variants showed 80% charging level for Glu-tRNA^{Gln2} production, but only AL3 showed potential discrimination against tRNA^{Glu} with a reduction to 70% plateau level for tRNA^{Glu} charging. The acceptor stem loop GluRS variants (ASL1 and ASL2) displayed significant preference for tRNA^{Gln2}, and ASL2-GluRS showed no detectable activity for tRNA^{Glu}

Table 1. Aminoacylation kinetics of WT-GluRS

WT-GluRS	k_{cat} , s^{-1}	K_M, μM	k_{cat}/K_M, s^{-1} μM^{-1}	Loss of efficiency*
tRNA^{Gln2}	0.41 ± 0.04	1.33 ± 0.40	0.31 ± 0.10	1.0
tRNA^{Glu}	0.04 ± < 0.01	3.61 ± 0.88	0.01 ± < 0.01	24
tRNA^{Gln1}	0.006 ± < 0.001	5.39 ± 1.44	0.001 ± < 0.001	257

*Loss of catalytic efficiency (*x* fold) is the relative fold decrease in k_{cat}/K_M that is calculated as the ratio of k_{cat}/K_M for tRNA^{Gln2} over the k_{cat}/K_M for the tRNA species listed in the far left column. Standard deviations are reported. Reaction conditions are given in *SI Methods*.

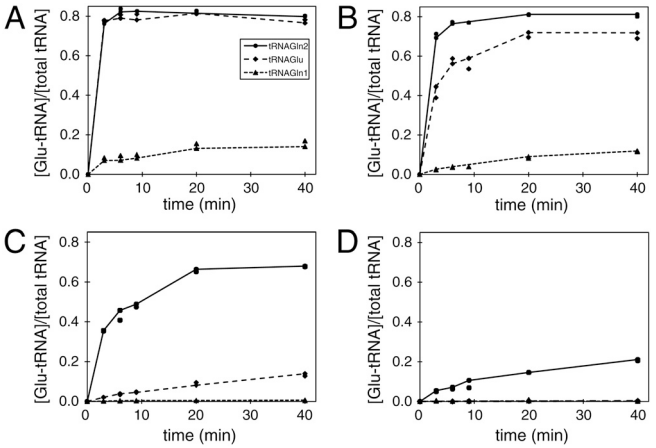

Fig. 2. Plateau tRNA charging curve. The plot shows the fraction of Glu-tRNA formation over the total amount of tRNA during the time course of the reaction. Plateau charging levels were measured for the enzymes WT-GluRS (*A*), AL3-GluRS (*B*), ASL1-GluRS (*C*), and ASL2-GluRS (*D*) with the tRNA substrates (see Fig. 1): tRNA^{Gln2} (●), tRNA^{Glu} (◆), and tRNA^{Gln1} (▲). In the reactions, 1 μM enzyme and 10.4 μM tRNA were used.

(Fig. 2 *C* and *D*). Both ASL1 and ASL2-GluRS showed no activity toward tRNA^{Gln1} under the assay conditions (Fig. 2 *C* and *D*).

Because the AL3 mutant enzyme showed a modest increase in tRNA^{Gln} over tRNA^{Glu} discrimination, we constructed additional mutants that contained the AL3 loop with either the ASL1 (Fig. S2*A*) or ASL2 (Fig. S2*B*) loops. Neither mutant appeared to increase specificity for tRNA^{Gln2}, so they were not further characterized.

Kinetic Characterization of GluRS Variants. A kinetic characterization was conducted to establish the magnitude of tRNA discrimination in the AL3, ASL1, and ASL2-GluRS variants (Table 2). The AL3-GluRS is of similar catalytic efficiency (within error) toward the tRNA^{Glu} compared to the WT-GluRS and shows slightly less tRNA discrimination than WT-GluRS.

Insertion of the acceptor stem loop from *E. coli* GlnRS into WT-GluRS does result in GluRS variants with enhanced specificity for tRNA^{Gln2} over tRNA^{Glu}. The ASL1 mutant replaces residues 203–205 of WT-GluRS with 134–140 from *E. coli* GlnRS, whereas the insertion in the ASL2-GluRS is a larger segment of the GlnRS sequence that replaces residues 191–209 of WT-GluRS with 122–144 of *E. coli* GlnRS (Fig. 4*A*). These insertions lead to less active enzymes than WT-GluRS, but they show a greater degree of tRNA^{Gln} specificity. With only sixfold less efficiency for tRNA^{Gln2} compared to WT-GluRS, the ASL1-GluRS catalytically favors tRNA^{Gln2} by 61-fold, which is a 2.5-fold enhancement in specificity for tRNA^{Gln2} as compared to the WT-GluRS. The specificity was achieved by kinetic discrimination against tRNA^{Glu} by ASL1-GluRS resulting from an approximately fourfold reduced k_{cat} and a fivefold increase in K_M for tRNA^{Glu} compared to WT-GluRS (Table 2).

The ASL2-GluRS, although far less catalytically efficient (by about 500-fold) than WT-GluRS toward tRNA^{Gln2}, was highly specific for tRNA^{Gln2} and showed no activity toward tRNA^{Glu}. The ASL2-GluRS exhibited a 100-fold reduction in k_{cat} as compared to wild-type GluRS, but the K_M for tRNA^{Gln2} was only fivefold greater than for WT-GluRS. The data indicate that the large ASL2 insertion may have perturbed the active site structure and this could explain the larger reduction on k_{cat} versus K_M.

Behavior of GluRS Variants Toward Acceptor Stem tRNA Mutants. Because the acceptor stem loop appears to endow the ASL2-GluRS with tRNA^{Gln} specificity, we created tRNA mutants to define which nucleotides in tRNA^{Gln} help ensure the engi-

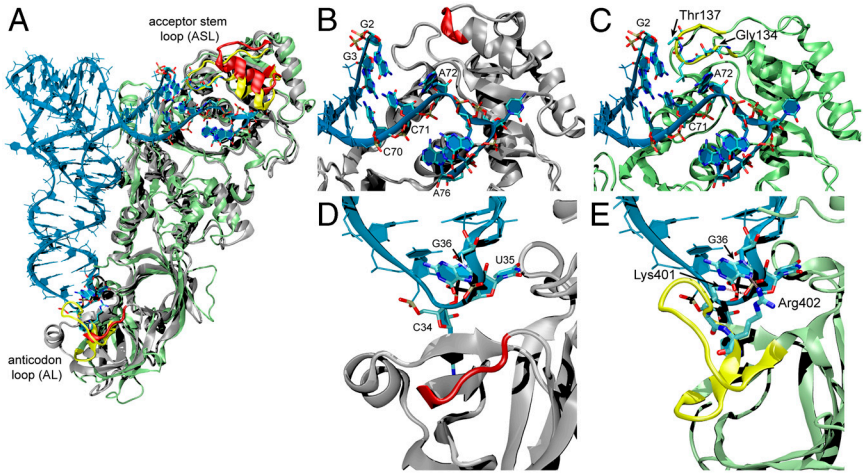

Fig. 3. A structural comparison (*A*) of the *M. thermautotrophicus* GluRS (silver) (11) and the *E. coli* GlnRS (green) $tRNA^{Gln}$ (blue) complex (12). Enlarged views of the ASL region are shown for (*B*) GluRS and (*C*) GlnRS. Larger view of the AL region are also shown for (*D*) GluRS and (*E*) GlnRS. Regions of GlnRS (yellow) swapped for regions in GluRS (red) in the ASL and AL-GluRS variants are highlighted. Sequences for each mutant are in Fig. 4. In panels showing the GluRS (*B*, *D*), the *E. coli* $tRNA^{Gln}$ is included for reference.

neered specify. The acceptor stem loop in *E. coli* GlnRS interacts (nonspecifically via hydrophobic contacts, see Fig. 3*C*) with the second base pair of the tRNA acceptor stem. In the complex structure, the first base pair (U_1:A_{72}) is open and the U_1 residue is disordered (12).

Given that the first base pair could be a critical element in $tRNA^{Gln}$ versus $tRNA^{Glu}$ discrimination as it is with the archaeal-specific amidotransferase Glu-$tRNA^{Gln}$ amidotransferase (GatDE) (18), we constructed $tRNA^{Gln2}$ mutants in which the wild-type A_1:U_{72} base pair is mutated to G_1:C_{72} (as in *M. thermautotrophicus* $tRNA^{Glu}$) and to U_1:A_{72} as in *E. coli* $tRNA^{Gln}$. Additional details are given in *SI Results* (Table S1 and Fig. S3), but the data show that WT and ASL GluRSs both recognize the first base pair, but respond differently to its mutation. The ASL GluRSs display significantly decreased affinity for the G_1:C_{72} mutant, whereas the same mutation affects both tRNA binding and catalysis for the WT-GluRS.

tRNA Phylogeny. The GluRS and GlnRS phylogenetic tree is well characterized (11, 14–17), but a detailed phylogeny of their cognate tRNAs has not been presented. We calculated a large scale phylogenetic tree of $tRNA^{Gln}$ and $tRNA^{Glu}$ sequences to determine if the evolution of GlnRS from the archaeal/eukaryotic GluRS is imprinted in the evolutionary history of the tRNAs.

A phylogenetic tree including all the $tRNA^{Glu}$ and $tRNA^{Gln}$ sequences available from the tRNA database (19) is presented in Fig. 5*B* and in complete detail in Dataset S1. Although some branches are not statistically supported and collapsed in the tree, most of the major clades in the tree (i.e., the archaeal $tRNA^{Gln}$, eukaryotic $tRNA^{Gln}$, bacterial $tRNA^{Gln}$, archaeal $tRNA^{Glu}$, eukaryotic $tRNA^{Glu}$, and bacterial $tRNA^{Glu}$) form well-separated and statistically significant groups, and the archaeal and eukaryotic tRNAs are more similar to each other than their bacterial counterparts as in the ribosomal RNA tree. These observations indicate that the evolutionary history of organisms is to some extent retained in the tRNA molecules. Furthermore, the $tRNA^{Glu}$ and $tRNA^{Gln}$ sequences are separated (supported by 100% bootstrap confidence). Although GlnRS evolved from the archaeal/eukaryotic GluRS (Fig. 5*A*), a similar event (i.e., a distinct $tRNA^{Gln}$ evolving from an archaeal/eukaryotic $tRNA^{Glu}$) is not present in the tRNA phylogeny (Fig. 5*B*). Further details regarding the tRNA phylogeny are given in *SI Results*.

In order to understand how the two $tRNA^{Gln}$ isoacceptors evolved in *M. thermautotrophicus*, we constructed a detailed phylogeny of the archaeal $tRNA^{Gln}$ sequences (Fig. S4). The tree is well resolved with a distinct separation between crenarchaeal and euryarchaeal tRNAs and other standard taxonomic divisions are also visible. The pattern of evolutionary divergence experienced by the archaeal species is evident and, therefore, retained in the tRNA sequences.

Despite being catalytically distinct substrates, $tRNA^{Gln1}$ and $tRNA^{Gln2}$ from *M. thermautotrophicus* are evolutionarily more closely related to each other than to any other $tRNA^{Gln}$ isoacceptor from a different archaeal genus. The $tRNA^{Gln2}$ occupies a long branch in its clade in the phylogeny (Fig. S4), so $tRNA^{Gln2}$ underwent a greater rate of evolutionary change than $tRNA^{Gln1}$. The data support a scenario in which the increase in aminoacylation efficiency of WT-GluRS for $tRNA^{Gln2}$ compared to $tRNA^{Gln1}$ is the principally the result of positive selection on $tRNA^{Gln2}$.

Discussion

Biochemical Properties of ND-aaRSs. Although the existence of tRNA-dependent amino acid biosynthesis was established more than 40 y ago with the demonstration that in *Bacillus* species

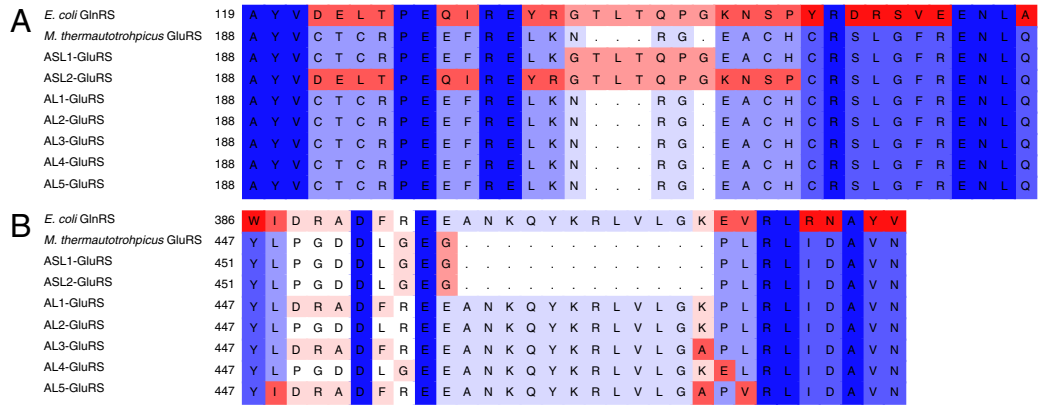

Fig. 4. Alignment of *E. coli* GluRS, *M. thermautotrophicus* GluRS (WT-GluRS), and GluRS variants. Regions surrounding the (*A*) acceptor stem loop and the (*B*) anticodon loop are shown. GluRS variants only differ from WT-GluRS at positions indicated. The enzymes are otherwise sequence identical. Sequences are color coded according to amino acid identity (descending from blue to red).

Table 2. Aminoacylation kinetics of AL3, ASL1, and ASL2-GluRS variants

	k_{cat}, s^{-1}	K_M, μM	k_{cat}/K_M, s^{-1} μM^{-1}	$tRNA^{Gln}$ specificity*	Loss of efficiency†
AL3-GluRS					
$tRNA^{Gln2}$	0.38 ± 0.09	2.22 ± 1.30	0.17 ± 0.11	1.0	1.8
$tRNA^{Glu}$	0.04 ± < 0.01	2.89 ± 0.92	0.01 ± < 0.01	14	26
ASL1-GluRS					
$tRNA^{Gln2}$	0.12 ± < 0.01	2.36 ± 0.28	0.05 ± 0.01	1.0	6.1
$tRNA^{Glu}$	0.01 ± < 0.01	16.4 ± 3.6	$(8.3 \pm 2.0) \times 10^{-4}$	61	374
ASL2-GluRS					
$tRNA^{Gln2}$	0.004 ± < 0.001	6.49 ± 0.90	$(5.8 \pm 0.8) \times 10^{-4}$	1.0	535
$tRNA^{Glu}$	no activity	no activity	no activity	ND	ND

*$tRNA^{Gln}$ specificity is calculated as the ratio of k_{cat}/K_M of the GluRS variant listed at left for $tRNA^{Gln2}$ over the k_{cat}/K_M of the same GluRS variant for $tRNA^{Glu}$.

†Loss of catalytic efficiency (*x* fold) calculated as the ratio of k_{cat}/K_M of WT-GluRS toward $tRNA^{Gln2}$ (data in Table 1) over the k_{cat}/K_M for the GluRS variant and tRNA indicated in the first column. No activity, no aminoacylation detectable with conditions in *SI Methods*; ND, not determinable.

Gln-$tRNA^{Gln}$ was synthesized from Glu-$tRNA^{Gln}$ and not from free Gln (4), only in the last decade has attention been devoted to understanding how relaxed tRNA specificity is achieved by the aaRS enzymes in these pathways.

The *Thermus thermophilus* nondiscriminating AspRS (ND-AspRS2) has 11-fold higher catalytic efficiency toward $tRNA^{Asp}$ compared to $tRNA^{Asn}$ (20). A mutant of the discriminating AspRS (D-AspRS) from *Pyrococcus kodakaraensis* could be converted to a nondiscriminating enzyme (still showing 10-fold catalytic preference for $tRNA^{Asp}$) by swapping the larger L1 loop in the anticodon binding domain of the D-AspRS with that found in the ND-AspRS (20, 21). A closely related *Deinococcus radiodurans* ND-AspRS2 mutant in the L1 loop (P77K) increased enzymatic discrimination threefold (22). The work proved that a loop in the anticodon binding domain of the aaRS affects $tRNA^{Asp}$ discrimination by differentiating the base at position 36 in the Asp versus the Asn anticodon.

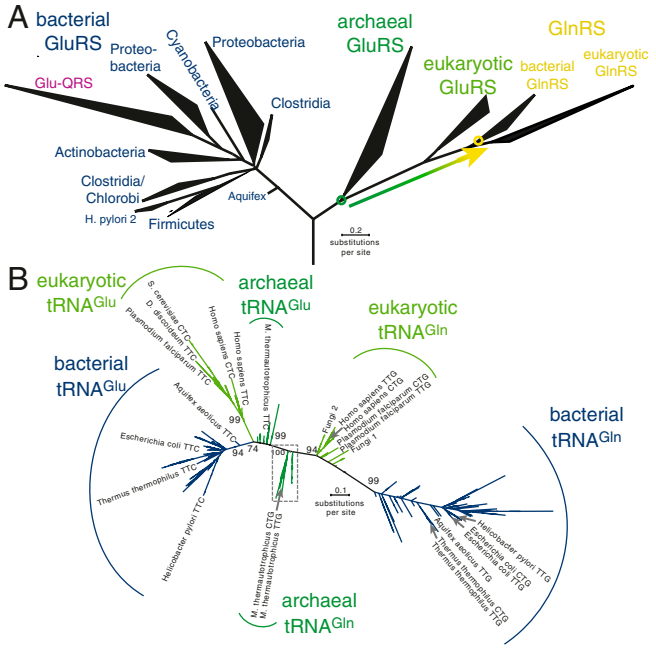

Fig. 5. Phylogenetic trees showing the evolution of (*A*) GluRS and GlnRS, and (*B*) $tRNA^{Glu}$ and $tRNA^{Gln}$. The schematic GluRS and GlnRS phylogeny (*A*) shows only the major phylogenetic groups from a previously calculated tree (11). The arrow indicates evolution of GlnRS from an archaeal-like nondiscriminating GluRS ancestor. A maximum likelihood phylogeny is shown (*B*) that was calculated from an alignment of 753 $tRNA^{Glu}$ and $tRNA^{Gln}$ sequences. Major clades and a few representative taxa are labeled. Bootstrap values are given for the major clades only (for a tree showing all taxa and bootstrap values see Dataset S1). Branches completely lacking support are collapsed. The boxed region is shown in complete detail in Fig. S4.

The bacterial-type nondiscriminating GluRS is reminiscent of the ND-AspRSs. A crystal structure of the *T. thermophilus* discriminating GluRS (D-GluRS) indicated that Arg358 is the critical element that distinguishes the Glu (^{34}YUC^{36}) and Gln (^{34}YUG^{36}) anticodons (9) based upon the identity of the nucleotide at position 36. The Arg358Gln mutation led to an enzyme with relaxed anticodon specificity (9). A structural and biochemical characterization of the *Thermosynechococcus elongatus* ND-GluRS found a Gly in place of Arg358 from *T. thermophilus* and other bacterial-type D-GluRSs (10). Enzyme kinetics revealed that this ND-GluRS is 13-fold more catalytically efficient toward $tRNA^{Glu}$ compared to $tRNA^{Gln}$.

Certain bacterial D-GluRS enzymes are able to discriminate $tRNA^{Glu}$ from $tRNA^{Gln}$ at the acceptor stem. The catalytic domain of the *E. coli* D-GluRS alone is able to discriminate $tRNA^{Glu}$ from $tRNA^{Gln}$ (21). Discrimination of $tRNA^{Gln}$ from $tRNA^{Glu}$ isoacceptors by the *Helicobacter pylori* GluRS2 is also achieved by recognizing the acceptor stem, in particular the $U_1{:}A_{72}$ base pair in $tRNA^{Gln}$ (23).

As the example system for archaeal-type ND-GluRSs, we found the *M. thermautotrophicus* GluRS to be biochemically distinct from the bacterial ND-GluRS and other ND-aaRSs characterized previously. In sharp contrast to the bacterial-type *T. elongatus* ND-GluRS, the *M. thermautotrophicus* GluRS was shown to prefer the major $tRNA^{Gln}$ isoacceptor ($tRNA^{Gln2}$) by 24-fold over $tRNA^{Glu}$, a twofold greater tRNA specificity than typically observed for ND-aaRSs. In addition, the D-AspRSs and bacterial type D-GluRSs typically recognize base 36 in the anticodon as a major element of discrimination between the tRNA substrates, whereas nondiscriminating relatives of the enzymes fail to recognize base 36 and are therefore able to aminoacylate both tRNA species. Our engineered GluRS variants indicate that the evolution of $tRNA^{Gln}$ specificity in GlnRS resulted from differentiating the $tRNA^{Glu}$ and $tRNA^{Gln}$ at the acceptor stem.

$tRNA^{Gln}$ Isoacceptors: Aminoacylation Efficiency and Codon Usage. The *M. thermautotrophicus* GluRS displays a striking catalytic preference (250-fold difference in k_{cat}/K_M) for $tRNA^{Gln2}$ compared to $tRNA^{Gln1}$. Interestingly, $tRNA^{Gln1}_{UUG}$ responds to the CAA codon, which appears 700 times in the 1,869 protein coding ORFs in the *M. thermautotrophicus* genome (24, 25). The other isoacceptor, $tRNA^{Gln2}_{CUG}$ responds to the far more common glutamine codon CAG (9,286 occurrences). The catalytic competence of the tRNA as a substrate for aminoacylation is, therefore, correlated with the tRNA's cognate codon usage. It is well known that tRNA expression levels are correlated

with codon usage bias (26, 27), but there are few examples of a correlation with aminoacylation efficiency. A detailed kinetic analysis of ArgRSs and their specificity for different $tRNA^{Arg}$ isoacceptors did not show such a correlation (28).

There appear to be two potential evolutionary forces responsible for the great difference in catalytic competence for $tRNA^{Gln1}$ versus $tRNA^{Gln2}$. Perhaps, as suggested by the phylogeny (Fig. S4), the enhanced aminoacylation efficiency of $tRNA^{Gln2}$ resulted from positive selection. The $tRNA^{Gln1}$ responds to fewer codons, so it is likely under a lower selective pressure than $tRNA^{Gln2}$ and could have acquired deleterious mutations from genetic drift. There is evidence supporting both scenarios in the nine nucleotide positions that differ between the $tRNA^{Gln}$ isoacceptors (Fig. 1). The energetically unfavorable *U*28:*G*42 base pair found in $tRNA^{Gln1}$ is the likely the most disruptive element for its catalytic performance. Because this base pair is found in no other archaeal tRNA, genetic drift may apply. At two of these positions, $tRNA^{Gln1}$ encodes bases (G15, C48) conserved in all other archaea, whereas $tRNA^{Gln2}$ has a unique mutation (A15, T48). Further experimentation could show if these changes enhanced the activity of $tRNA^{Gln2}$.

The Emergence of $tRNA^{Gln}$ Specificity. We found that insertion of the acceptor stem loop from GlnRS was sufficient to convert the archaeal-type nondiscriminating GluRS into a $tRNA^{Gln}$ specific enzyme. Because GlnRS evolved from an ancestor similar to the archaeal-type ND-GluRS, our transplantation of $tRNA^{Gln}$ specificity from GlnRS into the ND-GluRS recapitulates part of this evolutionary pathway.

Although Gln-$tRNA^{Gln}$ formation predates the emergence of GlnRS (11, 14–17), GlnRS was the first (known) enzyme that evolved to specifically recognize $tRNA^{Gln}$ and discriminate against $tRNA^{Glu}$. Besides noting an increased evolutionary rate among bacterial and eukaryotic tRNAs (*SI Results*), we were unable to find a compelling imprint of GlnRS evolution in the $tRNA^{Glu}$ and $tRNA^{Gln}$ phylogeny (Fig. 5). The finding indicates that GlnRS evolved to specifically recognize the extant $tRNA^{Gln}$ and did not require coevolution of an entirely new type of tRNA.

A second part of the evolutionary transition from ND-GluRS ancestor to GlnRS involved active site mutations that converted amino acid specificity from Glu to Gln. In a recent study, a GlnRS mutant capable of forming Glu-$tRNA^{Gln}$ was engineered from an impressive total of 22 amino acid replacements and one deletion (13, 29). It is, therefore, more likely that $tRNA^{Gln}$ specificity (requiring fewer mutations) evolved first, and direct Gln-$tRNA^{Gln}$ formation activity evolved subsequently. An analogous evolutionary intermediate exists in one of the two bacterial type GluRSs from *H. pylori*. One is specific for $tRNA^{Gln}$ and the other for $tRNA^{Glu}$ (30, 31). If amino acid specificity had evolved first, an undesirable evolutionary intermediate enzyme would result, which could form Gln-$tRNA^{Glu}$, potentially disrupting translation fidelity, whereas any Glu-$tRNA^{Gln}$ formed by a GluRS specific for $tRNA^{Gln}$ could be converted to Gln by the action of the GatDE amidotransferase.

Because the acceptor stem loop is critical for converting WT-GluRS to a $tRNA^{Gln}$ specific enzyme, sequence alignment of this region sheds light on the evolution of $tRNA^{Gln}$ specificity in the archaeal/eukaryotic type GluRS and GlnRS. The acceptor stem loop of *E. coli* GlnRS is found in nearly all GlnRS sequences (GTLTXXG consensus, see Fig. S1*A*). The loop is notably deleted in the *T. thermophilus* GlnRS and indicates that this enzyme may show some activity toward both $tRNA^{Gln}$ and $tRNA^{Glu}$. Eukaryotic GlnRSs likely evolved a distinct mechanism for discriminating $tRNA^{Gln}$ from $tRNA^{Glu}$ because this loop is usually absent in their GlnRS sequences. Exceptions include the GlnRS from yeast and other fungi that have a similar but smaller loop (Fig. S1*A*).

In keeping with the prediction that all archaeal GluRSs are nondiscriminating (32), the acceptor stem loop is indeed absent from all archaea, with only one exception. The *Sulfolobus solfataricus* GluRS has a similar loop (Fig. S1*A*) that is longer by one residue and of distinct sequence from the bacterial GlnRS loop. Because the organism lacks GlnRS, the *S. solfataricus* GluRS must be a nondiscriminating enzyme, yet we predict the enzyme will show a significantly higher catalytic preference for $tRNA^{Gln}$.

Conclusion

Biochemical measurements of the *M. thermautotrophicus* GluRS enzyme with its three homologous tRNA substrates ($tRNA^{Gln1}$, $tRNA^{Gln2}$, $tRNA^{Glu}$) served as the basis for comparison to our rationally designed enzymes. Although we expected $tRNA^{Gln}$ specificity could be controlled by the anticodon loop in GlnRS, activity of the engineered GluRS variants indicates that the acceptor stem loop is the principle discrimination element because insertion of this loop alone enhanced the specificity of archaeal GluRS toward $tRNA^{Gln2}$, significantly in the case of ASL2-GluRS.

There is now an increasing need for designed aminoacyl-tRNA synthetases. Recent efforts (reviewed in refs. 1 and 2) relied on unusual and engineered aaRSs as the principal vehicle for expanding the genetic code. Although over 100 different noncanonical amino acids have been genetically encoded already, for sufficient incorporation to support efficient recombinant protein synthesis only one amino acid at a time can be added to the genetic code. Genetic code expansion strategies are limited to tRNAs that in many cases inefficiently read amber, opal, or even four-base codons. Because tRNAs that read amber and opal codons must compete with the release factor (RF), an engineered *E. coli* strain lacking RF1 improved the read through efficiency of amber codons (33). Whereas four-base codons might allow 200 new open codons, experiments using orthogonal ribosomes selected to enhance read-through of four-base codons still lead to mostly truncated protein with inefficient synthesis of full-length product (34). These techniques were further manipulated to include two new amino acids simultaneously (35), but unless additional open codons are created the genetic code will be limited to about 22 amino acids for efficient production of protein containing multiple noncanonical amino acids.

Although in the past an unthinkable task, the ability to "write" (36) or recode (37) an entire genome may become routine before long. By, for example, reassigning all the glutamine codons to CAA, the CAG codon would then become open. The *Acidithiobacillus ferrooxidans* GluRS2, which is specific for $tRNA^{Gln1}_{UUG}$ that only reads the CAA codon in *E. coli* (38), could be employed with the amidotransferase GatDE to ensure that CAA remains a Gln codon. In such an organism, our designed GluRS and cognate $tRNA^{Gln2}_{CUG}$ would become a vehicle for genetic code expansion. The suggestion is supported by earlier work that showed *E. coli* tRNA isoacceptors are poor substrates for archaeal GluRS and that established an archaeal GluRS and an engineered $tRNA^{Gln}_{CUA}$ as an orthogonal amber decoding pair in *E. coli* (39).

Additional codons could be "opened," providing a platform for a vastly expanded genetic code. Suitable model organisms with such a dramatically expanded code would lead to breakthroughs in various fields including protein engineering, the encoded synthesis of biomaterials, and would provide a means to experimentally probe the optimality and structure of the genetic code itself. In order for an organism to accurately translate a genetic code containing many more amino acids, aminoacyl-tRNA synthetases that are specific for one (and only one) tRNA isoacceptor (itself specific for only one codon) will be needed to accurately synthesize all the necessary aminoacyl-tRNAs required to translate a maximally expanded code. Our data show that it is possible to develop an aaRS that is specific for a particular tRNA isoacceptor, suggesting that the engineered GluRS variant would be functional as an orthogonal pair in the context of a genome lacking the CAG codon. In such a background, further engineering of the GluRS active site could allow for incorporation of a

selected noncanonical amino acid, and so represents an approach toward further expansion of the code.

Methods

Plasmids and Bacterial Strains. The WT-GluRS was previously cloned (40) into the pTYB1 vector (New England Biolabs) and then transformed into an *E. coli* BL21/DE3 strain. Additional details are in *SI Methods*.

Protein and tRNA Purification and Preparation. Pure WT-GluRS, GluRS variants, and tRNA transcripts were produced as before (40), but with slight modification (see *SI Methods*).

Aminoacylation Assay. Formation of Glu-tRNA was monitored by measuring the amount of aminoacylated [^{32}P] labeled tRNA during the reaction time course. The reaction products were separated by thin layer chromatograph (see *SI Methods*), and during development radioactive spots for AMP and Glu-AMP (representing free tRNA and Glu-tRNA, respectively) were separated and then visualized and quantified by phosphorimaging.

Determination of Enzyme Kinetics. Enzyme kinetics were determined from experiments performed in duplicate, conducted independently at least twice. As the number of active molecules in an enzyme preparation of the WT-GluRS cannot be determined (13), we also assume the enzyme preparations to be fully active. Precise experimental details for all aminoacylation reactions reported are given in *SI Methods*.

Phylogeny and Bioinformatics. The tRNA gene sequences and alignments were downloaded from the transfer RNA database (19). Additional details and phylogenetic calculation parameters are given in *SI Methods*.

ACKNOWLEDGMENTS. We are grateful to Ilka Heinemann for scientific discussions and a critical reading of the manuscript, and to Dr. Michael J. O'Donoghue for encouragement. This work was supported by grants to D.S. from the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the US Department of Energy (DE-FG02-98ER20311), and the National Institute of General Medical Sciences (GM22854).

1. Ambrogelly A, Palioura S, Söll D (2007) Natural expansion of the genetic code. *Nat Chem Biol* 3:29–35.
2. Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. *Annu Rev Biochem* 79:413–444.
3. Sheppard K, et al. (2008) From one amino acid to another: tRNA-dependent amino acid biosynthesis. *Nucleic Acids Res* 36:1813–1825.
4. Wilcox M, Nirenberg M (1968) Transfer RNA as a cofactor coupling amino acid synthesis with that of protein. *Proc Natl Acad Sci USA* 61:229–236.
5. Curnow AW, Ibba M, Söll D (1996) tRNA-dependent asparagine formation. *Nature* 382:589–590.
6. Curnow AW, et al. (1997) Glu-tRNA^{Gln} amidotransferase: A novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc Natl Acad Sci USA* 94:11819–11826.
7. Lapointe J, Duplain L, Proulx M (1986) A single glutamyl-tRNA synthetase aminoacylates tRNA^{Glu} and tRNA^{Gln} in *Bacillus subtilis* and efficiently misacylates *Escherichia coli* tRNA^{Gln1} in vitro. *J Bacteriol* 165:88–93.
8. Becker HD, Reinbolt J, Kreutzer R, Giege R, Kern D (1997) Existence of two distinct aspartyl-tRNA synthetases in *Thermus thermophilus*. Structural and biochemical properties of the two enzymes. *Biochemistry* 36:8785–8797.
9. Sekine S, Nureki O, Shimada A, Vassylyev DG, Yokoyama S (2001) Structural basis for anticodon recognition by discriminating glutamyl-tRNA synthetase. *Nat Struct Biol* 8:203–206.
10. Schulze JO, et al. (2006) Crystal structure of a non-discriminating glutamyl-tRNA synthetase. *J Mol Biol* 361:888–897.
11. Nureki O, et al. (2010) Structure of an archaeal non-discriminating glutamyl-tRNA synthetase: A missing link in the evolution of Gln-tRNA^{Gln} formation. *Nucleic Acids Res* 38:7286–7297.
12. Rould MA, Perona JJ, Söll D, Steitz TA (1989) Structure of *E.coli* glutaminyl-tRNA synthetase complexed with tRNA^{Gln} and ATP at 2.8 Å resolution. *Science* 246:1135–1142.
13. Rodríguez-Hernández A, Bhaskaran H, Hadd A, Perona JJ (2010) Synthesis of Glu-tRNA^{Gln} by engineered and natural aminoacyl-tRNA synthetases. *Biochemistry* 49:6727–6736.
14. Brown JR, Doolittle WF (1999) Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. *J Mol Evol* 49:485–495.
15. Lamour V, et al. (1994) Evolution of the Glx-tRNA synthetase family: The glutaminyl enzyme as a case of horizontal gene transfer. *Proc Natl Acad Sci USA* 91:8670–8674.
16. Siatecka M, Rozek M, Barciszewski J, Mirande M (1998) Modular evolution of the Glx-tRNA synthetase family—rooting of the evolutionary tree between the bacteria and archaea/eukarya branches. *Eur J Biochem* 256:80–87.
17. Woese CR, Olsen GJ, Ibba M, Söll D (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol Biol Rev* 64:202–236.
18. Oshikane H, et al. (2006) Structural basis of RNA-dependent recruitment of glutamine to the genetic code. *Science* 312:1950–1954.
19. Jühling F, et al. (2009) tRNAdb 2009: Compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* 37:D159–162.
20. Charron C, Roy H, Blaise M, Giege R, Kern D (2003) Non-discriminating and discriminating aspartyl-tRNA synthetases differ in the anticodon-binding domain. *EMBO J* 22:1632–1643.
21. Dasgupta S, et al. (2009) The role of the catalytic domain of E. coli GluRS in tRNA^{Gln} discrimination. *FEBS Lett* 583:2114–2120.
22. Feng L, Yuan J, Toogood H, Tumbula-Hansen D, Söll D (2005) Aspartyl-tRNA synthetase requires a conserved proline in the anticodon-binding loop for tRNA^{Asn} recognition in vivo. *J Biol Chem* 280:20638–20641.
23. Chang KM, Hendrickson TL (2009) Recognition of tRNA^{Gln} by Helicobacter pylori GluRS2—a tRNA^{Gln}-specific glutamyl-tRNA synthetase. *Nucleic Acids Res* 37:6942–6949.
24. Smith DR, et al. (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: Functional analysis and comparative genomics. *J Bacteriol* 179:7135–7155.
25. Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: Status for the year 2000. *Nucleic Acids Res* 28:292.
26. Dong H, Nilsson L, Kurland CG (1996) Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J Mol Biol* 260:649–663.
27. Ikemura T (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol* 146:1–21.
28. Fender A, Sissler M, Florentz C, Giegé R (2004) Functional idiosyncrasies of tRNA isoacceptors in cognate and noncognate aminoacylation systems. *Biochimie* 86:21–29.
29. Bullock TL, Rodriguez-Hernandez A, Corigliano EM, Perona JJ (2008) A rationally engineered misacylating aminoacyl-tRNA synthetase. *Proc Natl Acad Sci USA* 105:7428–7433.
30. Salazar JC, et al. (2003) Coevolution of an aminoacyl-tRNA synthetase with its tRNA substrates. *Proc Natl Acad Sci USA* 100:13863–13868.
31. Skouloubris S, Ribas de Pouplana L, De Reuse H, Hendrickson TL (2003) A noncognate aminoacyl-tRNA synthetase that may resolve a missing link in protein evolution. *Proc Natl Acad Sci USA* 100:11297–11302.
32. Sheppard K, Söll D (2008) On the evolution of the tRNA-dependent amidotransferases, GatCAB and GatDE. *J Mol Biol* 377:831–844.
33. Mukai T, et al. (2010) Codon reassignment in the *Escherichia coli* genetic code. *Nucleic Acids Res* 38:8188–8195.
34. Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* 464:441–444.
35. Neumann H, Slusarczyk AL, Chin JW (2010) De novo generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs. *J Am Chem Soc* 132:2142–2144.
36. Gibson DG, et al. (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 319:1215–1220.
37. Isaacs FJ, et al. (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* 333:348–353.
38. Núñez H, Lefimil C, Min B, Söll D, Orellana O (2004) In vivo formation of glutamyl-tRNA^{Gln} in *Escherichia coli* by heterologous glutamyl-tRNA synthetases. *FEBS Lett* 557:133–135.
39. Santoro SW, Anderson JC, Lakshman V, Schultz PG (2003) An archaebacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in *Escherichia coli*. *Nucleic Acids Res* 31:6700–6709.
40. Feng L, Sheppard K, Tumbula-Hansen D, Söll D (2005) Gln-tRNA^{Gln} formation from Glu-tRNA^{Gln} requires cooperation of an asparaginase and a Glu-tRNA^{Gln} kinase. *J Biol Chem* 280:8150–8155.