Adaptation of an Orthogonal Archaeal Leucyl-tRNA and Synthetase Pair for Four-base, Amber, and Opal Suppression†

J. Christopher Anderson and Peter G. Schultz*

Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received April 8, 2003; Revised Manuscript Received June 2, 2003

ABSTRACT: Recently, it has been shown that an amber suppressor tRNA/aminoacyl-tRNA synthetase pair derived from the tyrosyl-tRNA synthetase of *Methanococcus jannaschii* can be used to genetically encode unnatural amino acids in response to the amber nonsense codon, TAG. However, we have been unable to modify this pair to decode either the opal nonsense codon, TGA, or the four-base codon, AGGA, limiting us to a 21 amino acid code. To overcome this limitation, we have adapted a leucyl-tRNA synthetase from *Methanobacterium thermoautotrophicum* and leucyl tRNA derived from *Halobacterium sp. NRC-1* as an orthogonal tRNA-synthetase pair in *Escherichia coli* to decode amber (TAG), opal (TGA), and four-base (AGGA) codons. To improve the efficiency and selectivity of the suppressor tRNA, extensive mutagenesis was performed on the anticodon loop and acceptor stem. The two most significant criteria required for an efficient amber orthogonal suppressor tRNA are a CU(X)XXXAA anticodon loop and the lack of noncanonical or mismatched base pairs in the stem regions. These changes afford only weak suppression of TGA and AGGA. However, this information together with an analysis of sequence similarity of multiple native archaeal tRNA sequences led to efficient, orthogonal suppressors of opal codons and the four-base codon, AGGA. Ultimately, it should be possible to use these additional orthogonal pairs to genetically incorporate multiple unnatural amino acids into proteins.

A great deal of effort has focused on the cotranslational incorporation of unnatural amino acids into proteins. Early work demonstrated that the translational machinery of *Escherichia coli* would accommodate amino acids similar in structure to the common 20 (*1*). This work was further extended by relaxing the specificity of endogenous *E. coli*[1] synthetases so that they activate unnatural amino acids as well as their cognate natural amino acid. Moreover, it was shown that mutations in editing domains could also be used to extend the substrate scope of the endogenous synthetase (*2*). However, these strategies are limited to recoding the genetic code rather than expanding the genetic code and lead to varying degrees of substitution of one of the common 20 amino acids with an unnatural amino acid.

Later, it was shown that unnatural amino acids could be site-specifically incorporated into proteins *in vitro* by the addition of chemically aminoacylated orthogonal amber suppressor tRNAs to an *in vitro* transcription/translation reaction (*3*−*6*). It is clear from these studies that the ribosome and translation factors are compatible with a large number of unnatural amino acids, even those with unusual structures. Unfortunately, the chemical aminoacylation of tRNAs is difficult, and this method can only produce microgram-scale quantities of protein because of the stoichiometric nature of the process. A catalytic *in vivo* method could overcome these limitations and would also permit the study of proteins containing unnatural amino acids in living cells.

To add additional synthetic amino acids to the genetic code *in vivo*, it is necessary to generate a 21st orthogonal pair of synthetase and tRNA that can function efficiently in the translational machinery. The synthetase must not cross-react with any of the endogenous tRNAs (40 in *E. coli*), and the orthogonal tRNA must not be aminoacylated by any of the endogenous synthetases (21 in *E. coli*). The tRNA must decode only a specific new codon that is not decoded by any endogenous tRNA, and the synthetase must charge its tRNA with only a specific unnatural amino acid. We have successfully generated an orthogonal tRNA-synthetase pair from tyrosyl-tRNA synthetase of *Methanococcus jannaschii* that satisfies these requirements. This system has been used to incorporate a series of unnatural amino acids including keto amino acids (*7*), photocrosslinking amino acids (*8*, *9*), and heavy atom containing amino acids selectively into proteins in response to the TAG codon.

Several other orthogonal pairs have been reported. Glutaminyl (*10*), aspartyl (*11*), and tyrosyl (*12*, *13*) systems derived from *Saccharomyces cerevisiae* tRNAs and synthetases have been described for the potential incorporation of unnatural amino acids in *E. coli*. Systems derived from the *E. coli* glutaminyl (*13*) and tyrosyl (*14*) synthetase have been described for use in *S. cerevisiae*. The *E. coli* tyrosyl system can also function in mammalian cells and has been used for the incorporation of 3-iodo-L-tyrosine *in vivo* (*15*). All of these systems have made exclusive use of the amber stop

† Support was provided by NIH Grant GM62159. J.C.A. is a NSF Pre-doctoral Fellow. This is manuscript 15571-CH of the Scripps Research Institute.

* Corresponding author. E-mail: schultz@scripps.edu.

[1] Abbreviations: Af, Ap, Hh, Mj, Mt, Pf, Ph, and Ec: *Archaeoglobus fulgidus*, *Aeuropyrum pernix*, *Halobacterium sp. NRC-1*, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Pyrococcus furiosus*, *Pyrococcus horikoshi*, and *Escherichia coli*, respectively; LRS, leucyl-tRNA synthetase; *bla*, gene for β-lactamase; *lacZ*, gene for β-galactosidase.

10.1021/bi034550w CCC: $25.00
Published on Web 07/25/2003

codon. To expand the genetic code beyond 21 amino acids, other orthogonal pairs and unique codons must be identified.

A critical requirement of any orthogonal pair is a codon that is unique within the genetic code and that will not cross-react with noncognate tRNAs. In addition to the amber stop codon (TAG), the opal nonsense codon (TGA) is one such candidate. A genetic code in which TAG and TGA encoded unnatural amino acids could encode 22 amino acids while preserving the ochre nonsense codon, UAA, which is the most abundant termination signal. The suppression of opal codons is robust *in vivo* but has not been frequently used for the incorporation of unnatural amino acids *in vitro* because of high background readthrough of TGA codons (*16*). Another possible codon involves unnatural base pairs. Unnatural amino acids have been incorporated in response to novel codons containing the unnatural base (*iso*-dC)AG (*17*) or pyridin-2-one (*18*) using an *in vitro* translation system. Adaptation of unnatural base pairs for the incorporation of unnatural amino acids into proteins *in vivo*, however, would require the faithful replication and transcription of unnatural base pairs in DNA and RNA (*19*). Another codon that might be used to encode additional amino acids are four- and five-base codons. Using a library of tRNAs with randomized anticodon loops coupled with a selection scheme, we recently identified several highly efficient and noncross-reactive four- and five-base codons, including AGGA, UAGA, CCCU, and UAGA (*20*, *21*).

Regardless of the codon chosen, it is necessary to generate additional orthogonal tRNA-synthetase pairs that can translate these codons with high fidelity and good efficiency. Because the tRNA anticodon loop is a major identity element for recognition by most synthetases, one must identify a synthetase that does not recognize these identity elements to generate suppressor tRNAs for these unusual codons. The leucyl-, seryl-, and alanyl-tRNA synthetases of *E. coli* are well-known to tolerate extensive substitutions in the anticodon loop (*22*−*24*). We thought it likely that some homologous archaeal or eukaryotic synthetases might have similar properties. We now report that derivatives of a leucyl-tRNA synthetase from *Methanobacterium thermoautotrophicum* and leucyl tRNAs derived from *Halobacterium sp. NRC-1* act as orthogonal tRNA-synthetase pairs for the amber codon in *E. coli*. Moreover, information gained in these studies, together with multiple sequence alignments of native archaeal tRNA sequences, allowed us to design efficient orthogonal suppressor tRNAs of opal codons and a four-base codon, AGGA.

MATERIALS AND METHODS

Strains, Plasmids, and Materials. All *in vivo* manipulations were carried out in *E. coli* strain TOP10 (Invitrogen) in LB media at 37 °C. *Halobacterium sp. NRC-1* was purchased from the American Type Culture Collection (ATCC). PCR was carried out according to standard protocols with a mixture of *Taq* (Promega) and *Pfu* (Stratagene) polymerases. Oligonucleotides were synthesized by Genosys, Operon, or the UCSF Biomolecular Resource Center. For oligonucleotides containing degenerate bases, the phosphoramidites were premixed to avoid bias. Standard protocols were employed for subcloning with restriction enzymes (NEB) and T4 DNA ligase (NEB). Plasmids were introduced into *E. coli* by electroporation. Sequence analysis was performed using the Genetics Computer Group, Inc. (GCG) software. The sequences of all plasmids were confirmed by restriction mapping and sequencing.

Cloning of tRNA Synthetase Genes. Genomic DNA was either purchased from ATCC or was prepared from a cell pellet purchased from ATCC. Genomic DNA was extracted using the DNeasy kit (Qiagen). Synthetase genes were amplified from genomic DNA by PCR and then subcloned into the *Nco*I and either *Eco*RI, *Kpn*I, or *Pvu*II sites of plasmid pKQ. More details on the cloning of these genes can be found in the Supporting Information. Plasmid pKQ contains the ribosome binding site, multiple cloning site, and *rrnB* terminator from plasmid pBAD-Myc/HisA (Invitrogen) under control of a constitutive glutamine promoter. The plasmid also contains a ColE1 origin of replication and a kanamycin resistance gene for plasmid maintenance.

Constructions of Reporter Plasmids. β-lactamase reporter plasmids were constructed from plasmid pACKO-Bla. This plasmid was constructed with a p15a origin, a chloramphenicol resistance gene, and unique sites for insertion of a gene for β-lactamase and a tRNA under control of the strong, constitutive *lpp* promoter. Site A184 of the β-lactamase gene was changed to TAG, AGGA, or TGA by an overlap PCR strategy, and the genes were subcloned into the *Aat*II and *Xma*I sites of pACKO-Bla to give plasmids pACKO-A184TAG, pACKO-A184AGGA, and pACKO-A184TGA.

Constructions of tRNA Plasmids. Genes for individual tRNAs and for tRNA libraries were constructed by extension reactions and subcloned into the *Eco*RI and *Pst*I sites of pACKO-Bla derivatives. All libraries represented at least 10-fold more members than the theoretical size of the library to ensure high coverage.

Measurement of Suppression Efficiency. A series of LB agar plates were prepared with 25 μg/mL kanamycin, 25 μg/mL chloramphenicol, and concentrations of ampicillin between 5 and 1000 μg/mL. Synthetase and tRNA plasmids were cotransformed and plated at densities below 100 cells per plate. Suppression efficiency was reported as the highest concentration at which cells survived to form colonies among a series of plates for which the next highest and lowest concentrations would be within 20% of the reported value.

Selection of Libraries and Characterization of Selectants. All tRNA libraries were subjected to ampicillin selection, and the surviving colonies were isolated and sequenced by the method described previously (*20*). Briefly, libraries were spread on LB plates containing 25 μg/mL kanamycin and chloramphenicol for plasmid maintenance and varying concentrations of ampicillin for selection. After 24 h of growth, the plates were scraped, and the cells were diluted slightly and then spread again on ampicillin plates. After colonies appeared, plates were again scraped and plated at dilute cell densities on a range of plates with different ampicillin concentrations. Selectants were isolated, sequenced, and then confirmed by retransformation into cells containing synthetase-expressing plasmids.

β-Galactosidase Reporter Assays. The full-length *lacZ* gene of plasmid pBAD-Myc/His/lacZ (Invitrogen) was amplified by PCR and subcloned into plasmid pLASC to obtain plasmid pLASC-lacZ. This pSC101-derived plasmid expresses *lacZ* gene under the control of an *lpp* promoter and has an ampicillin resistance gene for plasmid mainte-

nance. Derivatives of this plasmid were constructed wherein Leu-25 of the peptide VVLQRRDWEN of *lacZ* was replaced by TAG, TGA, or AGGA codons, or sense codons for tyrosine, serine, or leucine. The appropriate pLASC-lacZ-, pACKO-Bla-, and pKQ-derived plasmids were cotransformed and grown to an OD_{600} of 0.5. β-galactosidase assays were performed in quadruplicate using the BetaFluor β-Galactosidase Assay Kit (Novagen). Percent suppression was calculated as the percentage of activity for a sample relative to the value observed from the pLASC-lacZ construct with the corresponding sense codon at position 25. Cells containing pLASC-lacZ plasmids with sense codons at position 25 were also assayed by 2-nitrophenyl-β-D-galactopyranoside assays (*27*), and activity was calculated in Miller units.

Purification of Synthetase Proteins. Synthetase genes were cloned in frame with the C-terminal myc/his tag of pBAD-Myc/HisA (Invitrogen). Protein purification was performed with the Qiaexpressionist kit (Qiagen) by the manufacturer's protocol under native conditions. Protein concentrations were measured by the BCA Protein Assay Kit (Pierce) and analyzed by SDS−PAGE.

In Vitro Aminoacylation Assays. Aminoacylation assays were performed by methods described previously (*28*) in 20 μL reactions containing 50 mM Tris-Cl, pH 7.5, 30 mM KCl, 20 mM $MgCl_2$, 3 mM glutathione, 0.1 mg/mL BSA, 10 mM ATP, 1 μM (79 Ci/mmol) [^{3}H] leucine (Amersham), 750 nM synthetase, and 0, 2, 10, or 40 μM crude total tRNA. Crude total *E. coli* tRNA was purchased from Roche, and halobacterial tRNA was extracted from cultures of *Halobacterium sp. NRC-1* with the RNA/DNA Extraction Kit (Qiagen).

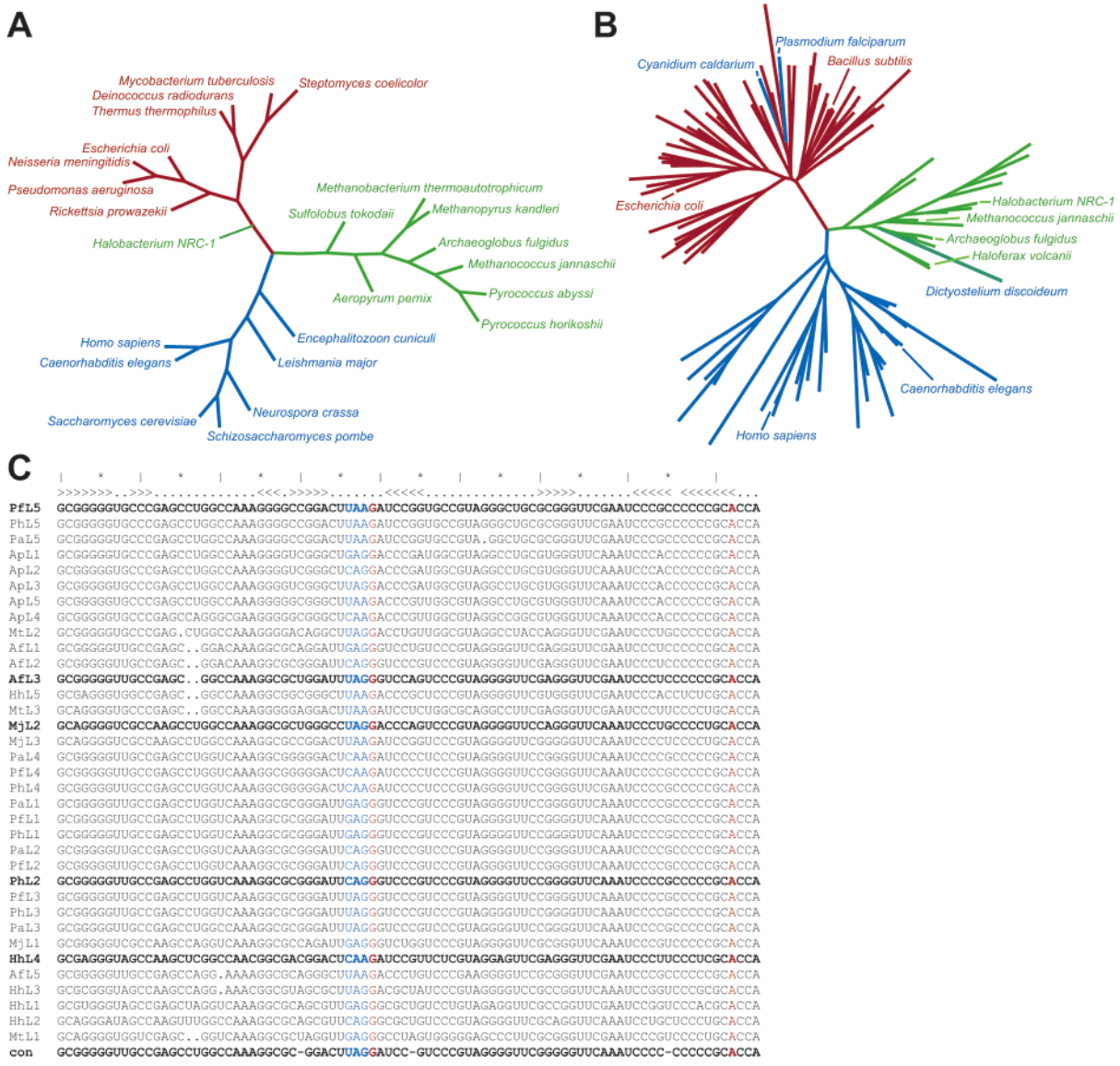
RESULTS

Identification of Orthogonal tRNAs. Previous studies (*29*) have shown that halobacterial tRNAs are inefficiently charged by the *E. coli* leucyl-tRNA synthetase. The similarity between the halobacterial and the other archaeal leucyl-tRNAs (see Figure 1C) led us to believe that tRNAs from other archaeans might also be orthogonal to the *E. coli* synthetases. The sequences were chosen to broadly represent the family of archaeal leucyl-tRNAs and included $tRNA_3^{Leu}$ of *Archaeoglobus fulgidus* (AfL3), $tRNA_4^{Leu}$ of *Halobacterium sp. NRC-1* (HhL4), $tRNA_2^{Leu}$ of *M. jannaschii* (MjL2), $tRNA_5^{Leu}$ of *Pyrococcus furiosus* (PfL5), and $tRNA_2^{Leu}$ of *Pyrococcus horikoshi* (PhL2) (see Figure 1C for sequences). In all cases, the anticodon was changed to CUA, and CCA was added to the 3′ terminus if the sequence was not present in the source gene to obtain an amber suppressor tRNA.

To measure the activity of suppressor tRNAs, a selection system was developed based on the *in vivo* suppression of nonsense or frameshift mutations introduced into the gene for β-lactamase (*bla*). Reporter genes for *bla* variants with TAG, AGGA, and TGA at position A184 (a permissive site (*10*)) were constructed in plasmid pACKO-Bla, a medium copy plasmid derived from pACYC184. Bacteria transformed with these reporter constructs are unable to grow on LB agar plates with ampicillin concentrations greater than 5 μg/mL, only slightly higher than the value (2 μg/mL ampicillin) observed for bacteria transformed with no plasmids. Plasmids derived from pACKO-Bla can also express tRNA genes under the control of a strong *lpp* promoter. When the robust amber suppressor gene *supD*, a tRNA efficiently charged by *E. coli* seryl tRNA synthetase, is expressed from pACKO-A184TAG (which encodes the A184TAG variant of *bla*), host bacteria survive at an ampicillin concentration of 1000 μg/mL. In contrast, in the case of an orthogonal tRNA, which cannot be efficiently charged by endogenous *E. coli* synthetases, ampicillin resistance should be less than 5 ug/mL. Conversely, if the tRNA is not orthogonal, or if a heterologous synthetase capable of charging the tRNA is coexpressed in the system, a higher level of ampicillin resistance should be observed.

The genes for the five potential orthogonal amber suppressor tRNAs were integrated into pACKO-A184TAG. *E. coli* hosts expressing the HhL4-derived suppressor, designated HL(TAG)1, could survive to only 5 μg/mL ampicillin, the MjL2- and PhL2-derived suppressors to 7 μg/mL, and the PfL5- and AfL3-derived suppressors to 20 μg/mL ampicillin. Therefore, all five suppressor tRNAs are either weak suppressor tRNAs or are inefficiently charged by *E. coli* aminoacyl-tRNA synthetases.

Cloning of Archaeal Leucyl-tRNA Synthetases. Because of the high homology of the archaeal leucyl-tRNAs, we anticipated that the archeal leucyl-tRNA synthetases might have similar tRNA recognition properties. Therefore, both species-specific and cross-species combinations of archeal leucyl-tRNAs and synthetases were examined to find an optimal pair for use in *E. coli*. The leucyl-tRNA synthetases from *A. fulgidus* (AfLRS), *Aeropyrum pernix* (ApLRS), *Halobacterium sp. NRC-1* (HhLRS), *M. jannaschii* (MjLRS), *M. thermoautotrophicum* (MtLRS), and *P. horikoshi* (PhLRS) were chosen as initial candidates because of the availability of the genome sequences and commercial availability of the organisms. The genes for these synthetases were cloned under the control of a constitutive glutamine promoter on the high copy plasmid, pKQ, which was constructed from pBR322 and contains a kanamycin resistance gene. The *leuS* gene from *E. coli* (EcLRS) was also cloned as a negative control. Synthetase expression plasmids and reporter constructs were cotransformed and assayed for activity by ampicillin selection (Figure 2). In general, the reporter plasmid containing the HhL4 suppressor tRNA, HL(TAG)1, gave the largest enhancement in suppression efficiency upon cotransformation with synthetase-expressing plasmids, but the PhL2- and AfL3-derived tRNAs also show a suppression enhancement. The MjL2- and PfL5-derived suppressor tRNAs survive to the same concentrations of ampicillin regardless of whether the archaeal synthetase is present and were not pursued further. From all 35 combinations of synthetase and reporter plasmids, the highest levels of ampicillin resistance result when the synthetases, MtLRS or MjLRS, are expressed with the HhL4-derived suppressor tRNA. The AfLRS construct gives slightly lower levels of resistance, and all other synthetases give no increase in suppression efficiency over background levels. With MjLRS or MtLRS, cells expressing HL(TAG)1 survive to 35 μg/mL ampicillin, but only 5 μg/mL with the *E. coli* synthetase or plasmid lacking the synthetase. Cells expressing AfLRS can survive to 25 μg/mL ampicillin when coexpressed with HL(TAG)1. From these *in vivo* suppression screens, three synthetases (MtLRS, MjLRS, and AfLRS) were identified as candidates for an orthogonal pair with the HhL4-derived amber suppressor tRNA, HL(TAG)1.




```
PfL5  GCGGGGGUGCCCGAGCCUGGCCAAAGGGGCCGGACUUAAGAUCCGGUGCCGUAGGGCUGCGCGGGUUCGAAUCCCGCCCCCCGCACCA
PhL5  GCGGGGGUGCCCGAGCCUGGCCAAAGGGGCCGGACUUAAGAUCCGGUGCCGUAGGGCUGCGCGGGUUCGAAUCCCGCCCCCCGCACCA
PaL5  GCGGGGGUGCCCGAGCCUGGCCAAAGGGGCCGGACUUAAGAUCCGGUGCCGUA.GGCUGCGCGGGUUCGAAUCCCGCCCCCCGCACCA
ApL1  GCGGGGGUGCCCGAGCCUGGCCAAAGGGGUCGGGCUGAGGACCCGAUGGCGUAGGCCUGCGUGGGUUCAAAUCCCACCCCCCGCACCA
ApL2  GCGGGGGUGCCCGAGCCUGGCCAAAGGGGUCGGGCUCAGGACCCGAUGGCGUAGGCCUGCGUGGGUUCAAAUCCCACCCCCCGCACCA
ApL3  GCGGGGGUGCCCGAGCCUGGCCAAAGGGGUCGGGCUUAGGACCCGAUGGCGUAGGCCUGCGUGGGUUCAAAUCCCACCCCCCGCACCA
ApL5  GCGGGGGUGCCCGAGCCUGGCCAAAGGGGGCGGGCUUAAGACCCGUUGGCGUAGGCCUGCGUGGGUUCAAAUCCCACCCCCCGCACCA
ApL4  GCGGGGGUGCCCGAGCCAGGGCGAAGGGGGCGGGCUCAAGACCCGUUGGCGUAGGCCGGCGUGGGUUCAAAUCCCACCCCCCGCACCA
MtL2  GCGGGGGUGCCCGAG.CUGGCCAAAGGGGACAGGCUUAGGACCUGUUGGCGUAGGCCUACCAGGGUUCGAAUCCCUGCCCCCGCACCA
AfL1  GCGGGGGUUGCCGAGC..GGACAAAGGCGCAGGAUUGAGGGUCCUGUCCCGUAGGGGUUCGAGGGUUCGAAUCCCUCCCCCCGCACCA
AfL2  GCGGGGGUUGCCGAGC..GGACAAAGGCGCGGGAUUCAGGGUCCCGUCCCGUAGGGGUUCGAGGGUUCGAAUCCCUCCCCCCGCACCA
AfL3  GCGGGGGUUGCCGAGC..GGCCAAAGGCGCUGGAUUUAGGGUCCAGUCCCGUAGGGGUUCGAGGGUUCGAAUCCCUCCCCCCGCACCA
HhL5  GCGAGGGUGGCCGAGC..GGCCAAAGGCGGCGGGCUUAAGACCCGCUCCCGUAGGGGUUCGUGGGUUCGAAUCCCACCUCUCGCACCA
MtL3  GCAGGGGUGCCCGAGC..GGCCAAAGGGGGAGGACUUAAGAUCCUCUGGCGCAGGCCUUCGAGGGUUCGAAUCCCUUCCCCUGCACCA
MjL2  GCAGGGGUCGCCAAGCCUGGCCAAAGGCGCUGGGCCUAGGACCCAGUCCCGUAGGGGUUCCAGGGUUCAAAUCCCUGCCCCUGCACCA
MjL3  GCAGGGGUCGCCAAGCCUGGCCAAAGGCGCCGGACUUAAGAUCCGGUCCCGUAGGGGUUCGGGGGUUCAAAUCCCCUCCCCCUGCACCA
PaL4  GCGGGGGUUGCCGAGCCUGGUCAAAGGCGGGGGACUCAAGAUCCCCUCCCGUAGGGGUUCCGGGGUUCAAAUCCCCGCCCCCGCACCA
PfL4  GCGGGGGUUGCCGAGCCUGGUCAAAGGCGGGGGACUCAAGAUCCCCUCCCGUAGGGGUUCCGGGGUUCAAAUCCCCGCCCCCGCACCA
PhL4  GCGGGGGUUGCCGAGCCUGGUCAAAGGCGGGGGACUCAAGAUCCCCUCCCGUAGGGGUUCCGGGGUUCGAAUCCCCGCCCCCGCACCA
PaL1  GCGGGGGUUGCCGAGCCUGGUCAAAGGCGCGGGAUUGAGGGUCCCGUCCCGUAGGGGUUCCGGGGUUCAAAUCCCCGCCCCCGCACCA
PfL1  GCGGGGGUUGCCGAGCCUGGUCAAAGGCGCGGGAUUGAGGGUCCCGUCCCGUAGGGGUUCCGGGGUUCAAAUCCCCGCCCCCGCACCA
PhL1  GCGGGGGUUGCCGAGCCUGGUCAAAGGCGCGGGAUUGAGGGUCCCGUCCCGUAGGGGUUCCGGGGUUCAAAUCCCCGCCCCCGCACCA
PaL2  GCGGGGGUUGCCGAGCCUGGUCAAAGGCGCGGGAUUCAGGGUCCCGUCCCGUAGGGGUUCCGGGGUUCAAAUCCCCGCCCCCGCACCA
PfL2  GCGGGGGUUGCCGAGCCUGGUCAAAGGCGCGGGAUUCAGGGUCCCGUCCCGUAGGGGUUCCGGGGUUCAAAUCCCCGCCCCCGCACCA
PhL2  GCGGGGGUUGCCGAGCCUGGUCAAAGGCGCGGGAUUCAGGGUCCCGUCCCGUAGGGGUUCCGGGGUUCAAAUCCCCGCCCCCGCACCA
PfL3  GCGGGGGUUGCCGAGCCUGGUCAAAGGCGCGGGAUUUAGGGUCCCGUCCCGUAGGGGUUCCGGGGUUCAAAUCCCCGCCCCCGCACCA
PhL3  GCGGGGGUUGCCGAGCCUGGUCAAAGGCGCGGGAUUUAGGGUCCCGUCCCGUAGGGGUUCCGGGGUUCAAAUCCCCGCCCCCGCACCA
PaL3  GCGGGGGUUGCCGAGCCUGGCCAAAGGCGCGGGAUUUAGGGUCCCGUCCCGUAGGGGUUCCGGGGUUCAAAUCCCCGCCCCCGCACCA
MjL1  GCGGGGGUCGCCAAGCCAGGUCAAAGGCGCCAGAUUGAGGGUCUGGUCCCGUAGGGGUUCGCGGGUUCAAAUCCCGUCCCCCGCACCA
HhL4  GCGAGGGUAGCCAAGCUCGGCCAACGGCGACGGACUCAAGAUCCGUUCUCGUAGGAGUUCGAGGGUUCGAAUCCCUUCCCUCGCACCA
AfL5  GCGGGGGUUGCCGAGCCAGG.AAAAGGCGCAGGGCUUAAGACCCUGUCCCGAAGGGGUCCGCGGGUUCGAAUCCCGCCCCCCGCACCA
HhL3  GCGCGGGUAGCCAAGCCAGG.AAACGGCGUAGCGCUUAGGACGCUAUCCCGUAGGGGUCCGCCGGUUCAAAUCCGGUCCCGCGCACCA
HhL1  GCGUGGGUAGCCGAGCUAGGUCAAAGGCGCAGCGUUGAGGGCGCUGUCCUGUAGAGGUUCGCCGGUUCGAAUCCGGUCCCACGCACCA
HhL2  GCAGGGAUAGCCAAGUUUGGCCAAAGGCGCAGCGUUCAGGGCGCUGUCCCGUAGGGGUUCGCAGGUUCAAAUCCUGCUCCCUGCACCA
MtL1  GCAGGGGUGGUCGAGC..GGUCAAAGGCGCUAGGUUGAGGGCCUAGUGGGGGAGCCCUUCGCGGGUUCGAAUCCCGUCCCCUGCACCA
con   GCGGGGGUUGCCGAGCCUGGCCAAAGGCGC-GGACUUAGGAUCC-GUCCCGUAGGGGUUCGGGGGUUCAAAUCCCC-CCCCCGCACCA
```

FIGURE 1: Leucyl tRNAs and synthetases. Archaeal sequences are labeled in green, prokaryotic in red, and eukaryotic in blue. (A) ClustalW analysis of aminoacyl-tRNA synthetases reveals the halobacterial synthetase to be unusual in its homology to prokaryotic rather than archaeal and eukaryotic synthetases. (B) Halobacterial tRNAs all share high homology to other archaeal tRNAs. Dendrograms were generated using the program PhyloDraw. (C) Multiple sequence alignment of the family of archaeal leucyl tRNAs examined as potential orthogonal suppressors. Sequences examined as potential amber suppressors by changing the anticodon (in blue) to CUA are shown in bold as is the consensus sequence. The highly conserved positions G37 and A73 are highlighted in red.

In Vitro Charging Assays. An *in vivo* suppression screen can distinguish active and inactive aminoacyl-tRNA synthetases, but it cannot distinguish an orthogonal synthetase from one that cross-reacts with *E. coli* tRNA. To determine the permissiveness of AfLRS, MjLRS, and MtLRS for *E. coli* tRNA, the synthetases were overexpressed, purified, and then subjected to *in vitro* aminoacylation assays to measure their ability to charge *E. coli* tRNA. AfLRS, MjLRS, and MtLRS were purified from an arabinose promoter overexpression system by Ni-NTA affinity chromatography in yields of 14, 8, and 3 mg/L, respectively. *In vitro* aminoacylation assays were performed with tritium-labeled leucine and either *E. coli* or *Halobacterium NRC-1* total tRNA (Figure 3). On the basis of the charging of 10 μM crude total tRNA, MtLRS and AfLRS charge halobacterial tRNA 54- and 21-fold more efficiently than *E. coli* tRNA, respectively. The MjLRS enzyme, however, shows only a 6-fold preference for halobacterial tRNA. The *E. coli* enzyme was 100-fold more efficient at charging *E. coli* crude total tRNA than halobacterial tRNA. Therefore, MtLRS and AfLRS are good candidates for orthogonal aminoacyl-tRNA synthetases with respect to *E. coli* tRNA, but MjLRS is not. Since MtLRS showed a higher level of suppression with HL(TAG)1 *in vivo* than did AfLRS, the MtLRS/HL(TAG)1 pair was carried forward as a potential new orthogonal pair for use in *E. coli*.

Optimization of the tRNA Anticodon Loop. The robust endogenous amber suppressor *supD* confers survival to 1000 μg/mL ampicillin when expressed from pACKO-A184TAG. In contrast, cells expressing the MtLRS/HL(TAG)1 pair survive to only 35 μg/mL ampicillin, which corresponds to a 2.9% suppression efficiency as determined from β-galactosidase assays (Table 1). We therefore sought to improve the activity of the system. Previous experiments on frameshift, missense, and nonsense suppression revealed that A37 was a highly conserved feature in robust suppressor tRNAs

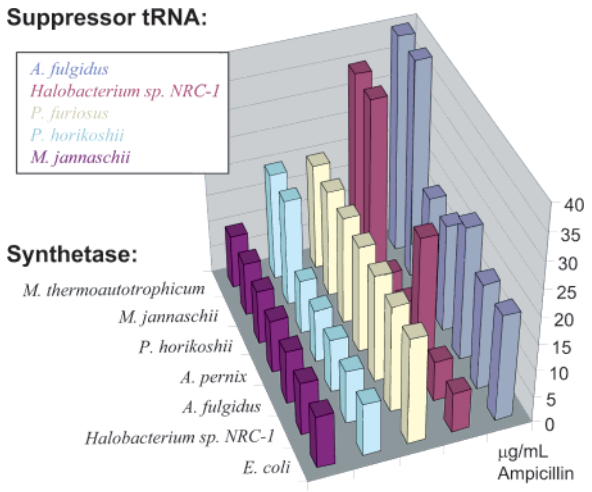


FIGURE 2: Identification of a leucyl orthogonal pair. The suppression efficiency of seven synthetases expressed with five orthogonal amber suppressor reporter constructs as measured using a β-lactamase amber suppression assay.

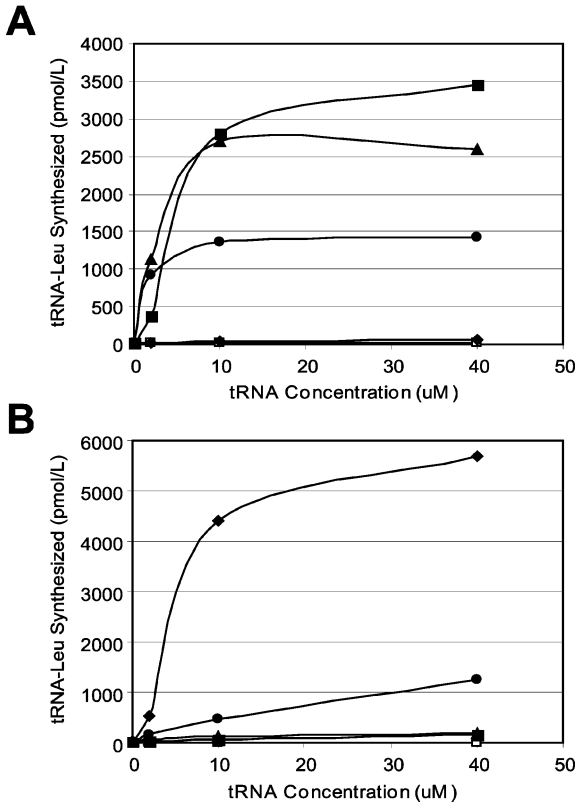


FIGURE 3: Aminoacylation *in vitro* by archaeal leucyl-tRNA synthetases. (A) Charging of crude total halobacterial tRNA determined by aminoacylation assays with [^{3}H] leucine with AfLRS (■), MjLRS (●), MtLRS (▲), EcLRS (♦), and no synthetase (□). (B) Charging of crude total *E. coli* tRNA.

(*20*). HhL4 has a G at position 37; therefore, substitution of G37 to A might be expected to improve suppression efficiency. To examine this and other possible anticodon loop mutants, a library was constructed in which the seven positions of the anticodon loop (positions 32−38, see Figure 4A) in HhL4 were replaced with degenerate bases and subcloned into pACKO-A184TAG. The library of tRNAs was cotransformed with pKQ-MtLRS and subjected to ampicillin selection initially at 35 μg/mL ampicillin for two rounds of selection, then plated on a series of plates with increasing ampicillin concentration in the third round of selection. At the highest concentration of ampicillin for which growth was observed (500 μg/mL), the only clone found had an anticodon loop with the sequence CUCUAAA, corresponding to a simple G37A mutation (Table 1). When cotransformed with pKQ-MtLRS, this clone could survive to 500 μg/mL ampicillin. In the absence of the synthetase, it survived to only 25 μg/mL ampicillin. Under similar conditions, cells containing the wild-type *M. jannaschii* tyrosyl orthogonal amber suppressor tRNA survive to 350 μg/mL ampicillin in the presence of the cognate synthetase and to 60 μg/mL ampicillin without the synthetase.

Randomization of Leucyl Acceptor Stem. Although the activity of the HhL4-derived amber tRNA was significantly improved with the G37A mutation, the suppression level in the absence of the synthetase increased from 5 to 25 μg/mL ampicillin. To overcome the undesired increase in background suppression, a mutant of the HhL4-derived tRNA was sought that would not cross react with *E. coli* aminoacyl-tRNA synthetases. Almost all of *E. coli* synthetases recognize bases within the acceptor stem of their cognate tRNAs. Therefore, we anticipated that mutations within this region of the tRNA might eliminate interactions between the orthogonal tRNA and the cross-reactive synthetase. A library in which the three terminal base pairs of the acceptor stem and the discriminator base were randomized (positions 1−3 and 70−73, the randomized region is outlined in Figure 4A) was constructed from the HL(TAG)2 mutant tRNA and subcloned into pACKO-A184TAG.

To identify members of this tRNA library that retained activity but were even poorer substrates for endogenous synthetases, a selection strategy was adopted from previous work on the *M. jannaschii* system (*30*). To isolate a pool of mutant tRNAs that had comparable activity to the G37A mutant of the HhL4-derived tRNA, the tRNA library in which the acceptor stem was randomized was cotransformed with pKQ-MtLRS and subjected to two rounds of positive selection at 500 μg/mL ampicillin. Six clones surviving the positive selection were sequenced, and all were unique and conserved the discriminator base, A73 (Figure 4). In all cases, the stem positions had standard Watson−Crick base pairs. To identify members of the pool of active clones that would not be charged by endogenous aminoacyl-tRNA synthetases, the surviving tRNA-expressing plasmids were transferred into cells containing a barnase reporter plasmid, pSCB2. This plasmid contains the gene for the RNase, barnase, with two TAG codons at permissive positions 2 and 44, under control of the arabinose promoter, as well as the gene for β-lactamase. Any tRNA that is aminoacylated by an endogenous *E. coli* synthetase will result in suppression of the nonsense codons and cell death. The cells were plated on LB plates containing 25 μg/mL chloramphenicol, 50 μg/mL ampicillin to maintain the plasmids, and 0.2% arabinose to induce expression of the barnase gene. Sixteen survivors were sequenced, and three unique sequences were identified. All three clones had reversed the 3:70 base pair from G:C to C:G. Of these, mutant HL(TAG)3 gave the highest level of suppression in the presence of MtLRS (600 μg/mL ampicillin) and only survived to 7.5 μg/mL ampicillin without the synthetase. These values correspond to 33.2% suppression

Table 1: Suppression Efficiency of Mutant Orthogonal tRNAs

Reporter Plasmid	Miller Units
pLASC-lacZ(Leu)	210 ± 2
pLASC-lacZ(Ser)	200 ± 5
pLASC-lacZ(Tyr)	192 ± 7
pLASC-lacZ(TAG)	1 ± 1
pLASC-lacZ(AGGA)	2 ± 1
pLASC-lacZ(TGA)	1 ± 1

	Percent Suppression[a]		
Suppressor tRNA	with pKQ	with synthetase	Sequence
HL(TAG)1	0.4 ± 0.1%	2.9 ± 0.8%	GCGAGGGTAGCCAAGCTCGGCCAACGGCGACGGACT**CTA**GATCCGTTC TCGTAGGAGTTCGAGGGTTCGAATCCCTTCCCTCGCACCA
HL(TAG)2	0.3 ± 0.1%	9.6 ± 0.4%	GCGAGGGTAGCCAAGCTCGGCCAACGGCGACGGACTCTA**A**ATCCGTTC TCGTAGGAGTTCGAGGGTTCGAATCCCTTCCCTCGCACCA
HL(TAG)3	1.5 ± 1.2%	33.2 ± 4.4%	**CC**AGGGTAGCCAAGCTCGGCCAACGGCGACGGACTCTAAATCCGTTC TCGTAGGAGTTCGAGGGTTCGAATCCCTTCCCT**GG**ACCA
HL(AGGA)1	0.4 ± 0.1%	4.6 ± 2.1%	GCGAGGGTAGCCAAGCTCGGCCAACGGCGACGGACT**TCCTA**ATCCGTT CTCGTAGGAGTTCGAGGGTTCGAATCCCTTCCCTCGCACCA
HL(AGGA)2	0.7 ± 0.3%	14.9 ± 6.1%	GCGAGGGTAGCCAAGCTCGGCCAACGGCGACGGACTTCCTAATCCGTT CTCGTAGGAGTTCGAGGGTTCGAATCCCT**C**CCCTCGCACCA
HL(AGGA)3	7.4 ± 0.4%	35.5 ± 1.4%	GCGGGGGTTGCCGAGCCTGGCCAAAGGCGCCGGACTTCCTAATCC**G**GT CCCGTAGGGGTTC**C**GGGGTTCAAATCCCCGCCCCCGCACCA
HL(TGA)1	4.7 ± 1.5%	60.8 ± 7.0%	GCGGGGGTTGCCGAGCCTGGCCAAAGGCGCCGGACT**TCA**AATCCGGTC CCGTAGGGGTTCCGGGGTTCAAATCCCCGCCCCCGCACCA
J17[b]	0.2 ± 0.1%	18.5 ± 4.8%	CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCATGG CGCTGGTTCAAATCCGGCCCGCCGGACCA
SupD	42.8 ± 7.1%	ND	GGAGAGATGCCGGAGCGGCTGAACGGACCGGTCTCTAAAACCGGAGTA GGGGCAACTCTACCGGGGGTTCAAATCCCCCTCTCTCCGCCA
Ser2AGGA	25.2 ± 0.1%	ND	GGAGAGATGCCGGAGCGGCTGAACGGACCGGTCTTCCTAAACCGGAGT AGGGGCAACTCTACCGGGGGTTCAAATCCCCCTCTCTCCGCCA

[a] β-Galactosidase activity was determined for tRNA reporter plasmids derived from pACKO-Bla cotransformed with the appropriate pLASC-lacZ mutant and either a synthetase-expressing plasmid or a plasmid with no synthetase. Activity is reported as the percentage of activity observed relative to the value observed from the pLASC-lacZ construct with a leucyl (wild-type), seryl, or tyrosyl sense codon at position 25. In each case, the codon at position 25 of *lacZ* is designated in parentheses. [b] J17, the *M. jannaschii* tyrosyl amber suppressor tRNA with improved orthogonality (*30*) was expressed in plasmid pACKO-A184TAG in the presence of pLASC-lacZ(TAG) and either pKQ or pBK-JYRS.

in the presence of MtLRS and 1.5% in the absence of the synthetase as determined by β-galactosidase amber suppression assays (see Table 1). By comparison, the mutant *M. jannaschii* suppressor tRNA, J17, gives values of 18.5 and 0.2% with and without the *M. jannaschii* tyrosine synthetase, respectively.

Identification of AGGA Suppressors. To expand the list of codons that can be used for unnatural amino acid mutagenesis, we attempted to generate a tRNA that could efficiently suppress a four-base codon. Previous studies indicated that the four-base codon AGGA can be efficiently suppressed in *E. coli*, and tRNAs with eight nucleotide anticodon loops were the most efficient suppressors of AGGA codons (*20*). A β-lactamase reporter plasmid analogous to the TAG reporter was constructed but with A184 replaced by AGGA instead of TAG. Normal translation in the absence of a +1 frameshift suppressor tRNA should result in missense errors downstream of position 184 and premature truncation of the protein. A library of tRNAs derived from the HhL4 tRNA was constructed in which the seven nucleotide anticodon loop was replaced with eight random nucleotides. The library was subcloned into pACKO-A184AGGA, cotransformed with pKQ-MtLRS, and then subjected to ampicillin selection. At the highest concentration of ampicillin at which growth was observed, 75 μg/mL, only one clone, HL(AGGA)1, was found. This clone had the anticodon loop sequence CUUCCUAA. As was the case with the *bla* A184TAG reporter plasmid, cells transformed with pACKO-A184AGGA can survive to only 5 μg/mL ampicillin in the absence of a suppressor tRNA. Therefore, the clone identified, HL(AGGA)1, is a weak suppressor of AGGA codons.

During these experiments, serendipitous mutants capable of surviving up to 300 μg/mL ampicillin were identified. These mutants were no longer orthogonal, and all had multiple point mutations relative to the parent sequence. All of the clones contained the substitution T65C. This mutation corrects the G:U mismatch present in the TψC loop stem, suggesting that this G:U base pair might be detrimental to suppressor activity. We therefore decided to randomize this base; a library was made in which the 49:65 base pair was randomized in HL(AGGA)1. The library was subcloned into pACKO-A184AGGA and then cotransformed with pKQ-MtLRS. Of the 16 library members, the most efficient suppressor, HL(AGGA)2, was identified by ampicillin selection. This clone contained a T65C mutation and could survive to 125 μg/mL ampicillin. Nevertheless, this level of activity was far lower than that observed for the corresponding amber suppressors. Consequently, alternative strategies were considered.

Mutations in the D-loop have been previously implicated in frameshift suppression (*31*), and we next hypothesized that such mutations might improve the suppression efficiency of the AGGA suppressor. Libraries wherein the 13 nucleotides of the D-loop (positions 14−21, see Figure 4) were replaced with 11 or 13 random nucleotides were prepared in pACKO-A184AGGA. Although a great deal of sequence diversity was observed among the survivors at the highest

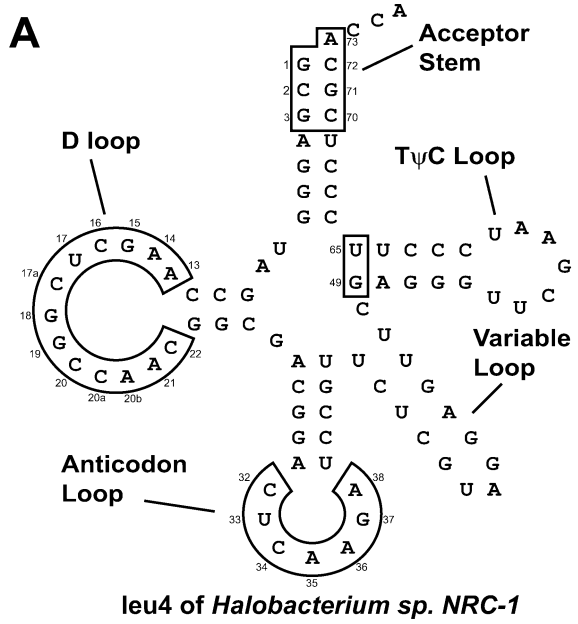


FIGURE 4: Optimization of suppressor tRNAs. (A) Regions of the halobacterial orthogonal tRNA subjected to mutagenesis in an effort to improve the efficiency or selectivity of TAG and AGGA suppressor tRNAs. (B) Active mutant TAG suppressors identified by positive selection conserve A73. Less cross-reactive mutants identified by a double-sieve selection strategy conserve a C3:G70 base pair. The most active and selective suppressor tRNA is shown with double boxes.

concentrations of ampicillin (125 μg/mL), no mutants were observed with increased activity relative to the parent tRNA.

Consensus-Derived AGGA Suppressor tRNA. In examining the sequence of the HhL4-derived tRNA, there was no obvious explanation for the poor activity of this suppressor. Rather than mutate HL(AGGA)2 further, we pursued an alternative approach. The archaeal leucyl-tRNAs are highly similar, varying from each other usually by only a few base substitutions (Figure 1C). The entire family would be well-represented by a library derived from a consensus sequence with many random mutations throughout. The consensus sequence was compiled with the GCG program *pileup*, and those positions considered degenerate by the program were changed to the most frequent base at those positions. The anticodon loop was changed to CUUCCUAA since this sequence was already shown to be the optimal sequence for an AGGA suppressor derived from HhL4. The final sequence used as the consensus sequence is shown in Figure 5. A library was synthesized by overlap extension of oligonucleotides in which each site of the tRNA gene was synthesized as a doped mixture of 90% of the consensus sequence and

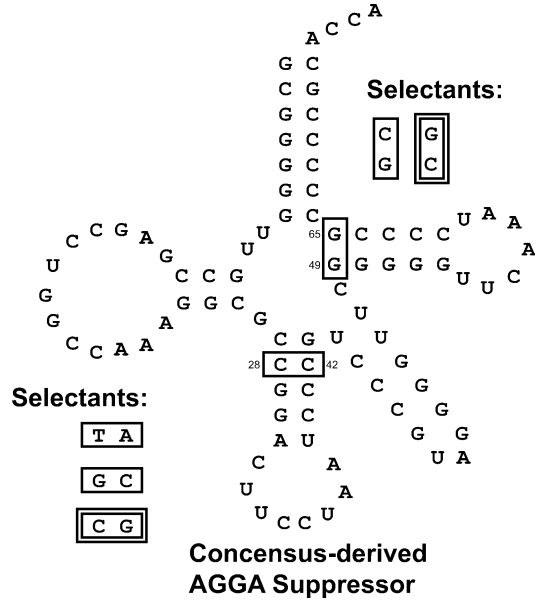


FIGURE 5: Consensus-derived frameshift suppressor. A consensus sequence is obtained by multiple sequence alignment of all known archaeal leucyl tRNAs, and the anticodon loop is changed to UCUCCUAA. The variations observed for tRNAs identified by selection are shown in boxes. The most active mutations are shown with double boxes.

10% of a mixture of the other three bases. The library was subcloned into pACKO-A184AGGA. Sequencing of 24 naïve clones revealed that the average number of mutations per clone was 5.9, and these mutations were randomly distributed throughout the tRNA sequence. After cotransformation with pKQ-MtLRS and selection on ampicillin plates, several clones survived to 300 μg/mL ampicillin and were found to be the original sequence with the 27:42 and 49:65 base pairs changed to the canonical base pairs T27:A42, G27:C42, or C27:G42 and G47:C65 or C47:G65 (Figure 5). The most efficient suppressor, designated HL-(AGGA)3, can survive to 300 μg/mL ampicillin in the presence of pKQ-MtLRS but to only 30 μg/mL in the absence of the synthetase, which correspond to 35.5 and 7.4% suppression, respectively, as determined by β-galactosidase assays (Table 1).

Identification of Opal Suppressor tRNAs. To further expand the list of codons, we sought opal suppressors derived from HhL4. A reporter plasmid, pACKO-A184TGA, was constructed in which the A184 position of β-lactamase was changed to TGA. This *bla* A184TGA reporter plasmid can survive to 10 μg/mL ampicillin without any suppressor tRNA present, whereas the TAG and AGGA reporters could survive to only 5 μg/mL. In the case of opal suppression, there is background read-through that leads to the production of a small amount of protein even in the absence of a suppression system. Nevertheless, this level is quite small. To identify suppressors, a library in which the anticodon loop (positions 31−38) of HhL4 was replaced with seven degenerate nucleotides was prepared in pACKO-A184TGA. When cotransformed with pKQ-MtLRS, no members of this library could survive on ampicillin plates at 50 μg/mL. Instead of HhL4, a library was prepared in which the eight nucleotide anticodon loop was randomized with seven nucleotides in HL(AGGA)3, the most robust AGGA suppressor identified from the consensus sequence. At the highest concentrations

of ampicillin at which growth was observed (300 μg/mL) only one clone, designated HL(TGA)1, with the sequence CUUCAAA was found. The clone can survive to 350 μg/mL ampicillin when coexpressed with pKQ-MtLRS but can survive to only 30 μg/mL without the synthetase plasmid, which corresponds to 60.8% suppression as determined by β-galactosidase assays (Table 1). Apparently, the beneficial effects of using the consensus sequence are not limited to frameshift suppression.

DISCUSSION

Identification of New Orthogonal Pairs. One approach to construct orthogonal tRNA-synthetase pairs is to adapt eukaryotic or archaeal synthetases and tRNAs for use in *E. coli*. Several yeast synthetases, notably glutamine, aspartic acid, arginine, and tyrosine, have been shown not to recognize *E. coli* tRNA and might therefore be useful for the construction of orthogonal tRNA-synthetase pairs. Unfortunately, many eukaryotic synthetases express poorly or have low specific activity in *E. coli*. Eukaryotic synthetases, particularly the mammalian enzymes, are often organized into large complexes (*35*), and the low activity often observed may be related to the inability to form these complexes in *E. coli*.

The success of the *M. jannaschii* tyrosyl orthogonal pair (*36*) suggested that archaebacteria may in general be a good source of orthogonal pairs. Early work on the halophile *Halobacterium cutirebrum* (*29*) indicated that almost all the tRNAs of this archaean (notably leucine, arginine, tyrosine, serine, histidine, and proline) cannot be charged by *E. coli* aminoacyl-tRNA synthetases. Indeed, archaeal tRNA synthetases are more similar to their eukaryotic than prokaryotic counterparts in terms of homology and tRNA recognition elements. Unlike their eukaryotic counterparts, however, there is currently no evidence for their higher order assembly into structured multimers (*37*, *38*). Moreover, since most archaea are thermophiles, active synthetases from archaea can be expressed in good yields in *E. coli* and can be readily purified in active form. Because of extensive sequencing efforts, at least 16 archaeal genome sequences are currently available, which together with the lack of introns in the genome, greatly facilitate the PCR amplification of the archaeal synthetase genes. For all of the above reasons, our attention has focused on the archaea as a source for orthogonal pairs.

Another design issue in the construction of orthogonal tRNA-synthetase pairs is the ability of the aminoacyl-tRNA synthetase to recognize mutants of the cognate tRNA with altered anticodon loops (i.e., nonsense or missense suppressors). Aminoacyl-tRNA synthetases frequently use the anticodon loop as a major positive identity element, and mutations in this region of the tRNA frequently result in impaired synthetase recognition. The leucyl-tRNA synthetases frequently lack strong anticodon recognition elements, and a leucyl orthogonal tRNA-synthetase pair might therefore be able to decode a variety of codons, including amber, opal, and four-base codons. Of the archaeal leucyl-tRNA synthetases, only the enzyme from *Haloferax volcanii* has been thoroughly investigated (*25*). The synthetase does not recognize bases in the anticodon loop; instead, a highly conserved pattern of mismatches within the variable loop is the primary recognition element for the synthetase. Although the cloning of the gene for this enzyme has not been reported, the sequenced genome of a closely related archaean, *Halobacterium sp. NRC-1*, is available. A multiple sequence alignment of leucyl-tRNA synthetases from many phyla including archaeal, prokaryotic, and eukaryotic sequences (Figure 1A) shows that the halophilic enzyme is unusual among the family of archaeal synthetases, having greater homology to the prokaryotic branch than the eukaryotic or archaeal branches. Unlike the synthetases, all archaeal leucyl tRNAs are highly homologous and share absolutely conserved features such as A73, G37, and a 12 nucleotide variable loop with two unpaired bases (Figure 1B). The conservation of these positive recognition elements led us to believe that tRNA recognition by the other archaeal leucyl-tRNAs would be similar to recognition by the halobacterial synthetase. Consequently, these synthetases might be useful in the construction of orthogonal tRNA-synthetase pairs when combined with suppressor tRNAs derived from archaeal leucyl-tRNAs.

Because archaeal leucyl-tRNAs and synthetases are highly homologous to one another, both species-specific and cross-species combinations could potentially function as efficient orthogonal tRNA-synthetase pairs. Therefore, each of the five potential orthogonal tRNAs (AfL3, HhL4, MjL2, PfL5, and PhL2) was examined in the presence of each of the six separate archaeal synthetases (AfLRS, ApLRS, HhLRS, MjLRS, MtLRS, and PhLRS) for the ability to suppress A184TAG in *bla*. All five orthogonal tRNAs afforded a higher level of amber suppression in the absence of an archaeal synthetase than is observed when no amber suppressor tRNA is present in the cell. All five suppressors are, therefore, expressed, processed, and functionally charged to some degree by an endogenous *E. coli* synthetase. Nevertheless, only three of the five tRNAs (PhL2, AfL3, and HhL4) gave a higher level of suppression when a foreign synthetase (either MjLRS, MtLRS, or AfLRS) was coexpressed with the tRNA than was observed with no synthetase. It is unknown why the MjL2 and PfL5 suppressors fail to give an enhancement in suppression when coexpressed with a cognate or noncognate archaeal synthetase. These tRNAs may be expressed as functional suppressor tRNAs in *E. coli* but are unable to be charged because of incompatibility with both cognate and noncognate synthetases. In the case of MjL2, the suppressor is derived from the natural substrate for MjLRS, so it seems unlikely that the tRNA would not be charged, when other tRNAs are efficiently charged by MjLRS. Another explanation might be that the tRNAs are efficiently charged but are incompatible with the *E. coli* translational machinery, but this is not consistent with the fact that some suppression is observed when no archaeal synthetase is present. A third possibility is that MjL2 and PfL5 are efficiently charged with leucine but are deacylated in an editing process by an endogenous *E. coli* synthetase. In any case, it is clear that not all archaeal leucyl isoacceptors are equivalent in their ability to function as orthogonal amber suppressors in *E. coli*.

Only three of the six leucyl-tRNA synthetases (MjLRS, MtLRS, and AfLRS) cloned from archaea gave a higher level of suppression when combined with any of the five orthogonal tRNAs. In the case of HhLRS, the synthetase does not yield protein when overexpressed. Most likely, PhLRS and ApLRS do not express functional protein in *E. coli* either, but it is also possible that the proteins are not active at 37

°C or do not recognize any of the orthogonal tRNAs tested. There was no evidence that some tRNAs are preferred substrates for a specific synthetase. Indeed, although a tRNA from *M. jannaschii* was one of the five orthogonal tRNAs examined, the halobacterium-derived suppressor was the preferred substrate for MjLRS. All three functional tRNAs gave the highest level of suppression when charged by MtLRS or MjLRS, and to a lesser degree with AfLRS.

Although on the whole the archaeal leucyl synthetases have similar tRNA recognition properties, it is clear from *in vitro* charging experiments that there are some differences in their recognition of tRNA. The charging of crude total *E. coli* tRNA by AfLRS and MtLRS is only 5- and 13-fold higher, respectively, than the background reaction observed with no synthetase, whereas MjLRS is able to charge *E. coli* tRNA 50-fold over background. Such differences in tRNA recognition among highly homologous synthetases was unanticipated but not without precedent (*29*). Since aminoacyl-tRNA synthetases have evolved only to be orthogonal to the noncognate tRNAs present in their own host's cytoplasm, it is perhaps not surprising that subtle variations in sequence or chemical modification can lead to mischarging in foreign systems.

Improving the Activity of Orthogonal Suppressor tRNAs. To date, we have successfully identified and characterized three orthogonal tRNA-synthetase pairs: the yeast glutamine (*10*), yeast aspartate (*11*), and archaeal tyrosine pairs (*39*). Of these systems, only the tyrosine system gives levels of amber suppression comparable to the levels observed for strong native amber suppressors such as supD or supF. When expressed with a the high-copy β-lactamase reporter pBLAM (the reporter plasmid for this study was a medium-copy plasmid) in the presence of their cognate synthetase, cells containing the original glutamine, aspartate, and tyrosine orthogonal amber suppressor tRNAs can survive to 140, 60, and 1220 ug/mL ampicillin, respectively (*11*, *39*). The tyrosine system is also the only system that has thus far been successfully modified to specifically incorporate an unnatural amino acid *in vivo* in response to the amber codon. A high level of suppression may be critical to the successful modification of the amino acid specificity of synthetases using a double-sieve selection strategy (*10*). For suppression systems with low activity, it is often difficult to distinguish active and inactive synthetases in selection experiments because of their similarity in phenotype. A high level of suppression is required for the production of protein containing unnatural amino acids. Therefore, a great deal of attention has been paid to those features of orthogonal tRNAs that give rise to robust suppression.

Previous work on frameshift and amber suppression in *E. coli* clearly indicates that positions 31, 32, 37, and 38 of the tRNA anticodon loop have profound effects on suppression efficiency (*23*, *40*−*42*). The presence of G37 in all the archaeal leucyl tRNAs led us to believe that a substitution at this position might lead to a higher suppression efficiency. Indeed, randomization of the anticodon loop showed that the most efficient suppressors have the anticodon loop CUCUAAA. Although the tRNA was toxic, the G37A mutant also emerged through selection with the *M. jannaschii* tyrosine system (*30*) as the most potent suppressor thus far observed for this system. Similar selection experiments with the yeast-derived glutamine and aspartate orthogonal pairs have been performed in which libraries of positions 32−38 of the anticodon loop are replaced with degenerate bases and then subjected to positive ampicillin selection in the presence of the cognate synthetase (J. C. Anderson, P. G. Schultz, and M. Pastrnak, unpublished results). In both cases, the anticodon loop sequence CUCUAAA afforded the highest suppression efficiency corresponding to 6- and 5-fold enhancements in the concentration of ampicillin at which growth is observed for the glutamine and aspartate systems, respectively. In at least three other systems, tRNAs with the anticodon loop sequence CUCUAAA have emerged as the most efficient amber suppressors. The anticodon loop sequence CUUCCUAA was found to be the most efficient sequence for a leucyl AGGA suppressor. Selection experiments on tRNAs with randomized anticodon loops in *E. coli* $tRNA_2^{Ser}$ similarly converged on the anticodon loop sequence CUUCCUAA for AGGA suppression (*20*), and the sequence CUUCAAA also emerged as the most efficient anticodon loop sequence for a leucyl opal suppressor. These results suggest that the preferred anticodon loop sequence is determined by interactions with endogenous translational machinery rather than the particular preferences of the aminoacyl-tRNA synthetases. Indeed, the anticodon loop may require sequence-specific modifications to function optimally (*43*, *44*). Alternatively, Yarus (*45*) has suggested that the entire anticodon stem and loop (positions 27−43 of the tRNA) together function as an extended anticodon that interacts with ribosome as a module. The entire sequence of this region may help to define the identity of the anticodon for proper decoding.

All three codons examined in this study were most efficiently suppressed by tRNAs with the sequence CU(X)XXXAA in the anticodon loop. Although this consensus sequence is preferred for TAG, TGA, and AGGA codons, other sequences may be preferable for other four- and five-base codons. In previous studies (*20*), the most efficient suppressor tRNAs had bases at positions 32, 33, 37, and 38, which differed from the consensus sequence. For example, the most efficient suppressors of the codon CUAG had an anticodon loop with the sequence CGCTAGGA, deviating at both U33 and A37. In addition, some synthetases employ position 37 as a strong positive determinant for recognition, in which case a CU(X)XXXAA anticodon loop sequence may prove to be nonoptimal.

Optimization of the anticodon loop sequence as described above was sufficient to provide an efficient amber suppressor tRNA for the leucine system. Optimization of the anticodon loop of the AGGA frameshift suppressors derived from HhL4 also afforded a viable tRNA. However, the suppression efficiency (4.6%) of this tRNA, HL(AGGA)1, is far lower than that measured for the suppression of amber codons by HL(TAG)2. Indeed, this suppressor permitted survival at only 75 μg/mL ampicillin, significantly less than the seryl AGGA suppressor (Ser2AGGA) identified previously (*20*), which can survive to 275 ug/mL ampicillin when expressed in plasmid pACKO-A184AGGA. In general, the best AGGA suppressors are less active than the best amber suppressors (*21*), but there appears to be something particular to HhL4 that hinders its ability to act as a frameshift suppressor. The only feature obviously different from robust four-base suppressors previously identified (*46*) is the presence of a very large D loop in HhL4. Most suppressors have nine nucleotides in the D loop and four base pairs in the stem. HhL4 has only three base pairs in the stem and 13 bases in

the loop. Moreover, previous studies have shown the D loop to play a role in frameshift suppression (*31*). Not only did we see no increase in activity upon randomization of the D loop, there was also a great deal of sequence variation among the most active suppressors.

The serendipitous appearance of mutations in the G49:U65 base pair of the four-base suppressor tRNAs suggested that noncanonical base pairing in the stem regions of tRNAs has a deleterious effect on suppression efficiency. This hypothesis was further supported by a randomization and selection experiment on the acceptor stem of the HhL4-derived amber suppressor. The three terminal base pairs of the acceptor stem were simultaneously randomized. This library of tRNAs would therefore contain all combinations of mismatched and Watson−Crick base pairs. In fact, 98.4% of the theoretical members of this library should have at least one mismatched base pair. Nevertheless, in the nine active acceptor stem mutants outlined in Figure 4, all positions are occupied by Watson−Crick base pairs. Similarly, the D, TψC, anticodon, and acceptor stems of the yeast glutamine amber suppressor tRNA have been individually randomized and subjected to positive selection (J. C. Anderson and P. G. Schultz, unpublished results). In all surviving clones, every position in these stem regions was occupied by a Watson−Crick pair. In the parent tRNA, the 6:67 base pair is U:G. Mutation of this base pair to U:A results in a doubling of the concentration of ampicillin at which cells can grow. Also, when subjected to positive selection, the only mutations that emerged from random mutagenesis of the leucyl consensus-derived frameshift suppressor appeared at mispaired sites. Others have also noted that mispairing in stem regions adversely affects suppression efficiency (*47*, *48*). It may be that tRNAs with mispaired bases are not readily folded into the correct cloverleaf structure and therefore are not readily processed and modified (*49*). A quantitative analysis of the ratio of charged to uncharged species and of the ratio of fully processed to unprocessed tRNA present in the cell would greatly enhance our understanding of the mechanisms by which these poorly suppressing tRNAs are impaired.

An analysis of multiple sequence alignments of many families of tRNAs reveal multiple examples of conserved non-Watson−Crick pairings. For example, a G3:U70 base pair is a conserved positive determinant for recognition by *E. coli* alanyl-tRNA synthetase (*50*). If the element is a conserved positive determinant for recognition, then it may prove difficult to construct robust suppressor tRNAs for the cognate synthetase. Most often, however, the mispairing present in native sequences is only found in specific isoacceptors. These and other variations from the consensus sequence of the family of tRNAs present in individual isoacceptors may be present as a result of subtle, species-specific adaptations in positive or negative synthetase recognition, optimal processing and modification, or interactions with elongation factors. Alternatively, these variations may simply be the result of neutral evolutionary drift.

When transferred to another species, these variations are unlikely to offer to the new host's translational machinery any advantages they conferred to the source organism. Furthermore, for cross-species pairs, the synthetase is unlikely to recognize any species-specific identity elements present in the tRNA. Only those recognition elements common to the entire family are likely to be useful. Similarly, any processing or modification adaptations particular to a specific tRNA would be lost to the *E. coli* translational apparatus. These variations may even be deleterious to suppression efficiency, particularly when these variations are mismatched bases in stem regions. Suppressor tRNAs derived from the consensus sequence preserve only those features that are broadly shared by the entire family and eliminate potentially deleterious variations. Therefore, suppressor tRNAs derived from the consensus sequence may in general lead to higher suppression efficiencies.

Although optimization of the anticodon loop and elimination of mispairing gave modest to large increases in suppression efficiency, these modifications were not sufficient to provide robust AGGA and opal suppressor tRNAs. Only the consensus-derived suppressors had activities comparable to the $tRNA_2^{Ser}$-derived suppressors described previously (*21*). A comparison of the consensus-derived sequences for HL(AGGA)3 and HL(AGGA)2 reveal that there are 14 base substitutions, but neither sequence has mispairs. It is not obvious which of these 14 substitutions account for the improved activity. Perhaps by using the consensus sequence of the entire family of tRNAs, those bases that are specific to any particular tRNA and may be detrimental to activity are identified and eliminated. With only this one example, it is unclear whether the identification of this sequence is simply serendipitous or represents a generally applicable strategy. We plan to test this strategy in the design of additional orthogonal pairs.

Improving the Selectivity of Orthogonal Suppressor tRNAs. Unfortunately, improvements in the activity of these suppressor tRNAs also brought about an undesirable increase in the level of suppression observed in the absence of synthetase. The original *M. jannaschii* tyrosine orthogonal suppressor tRNA was partially charged by an *E. coli* synthetase, but the reaction was eliminated by mutagenesis (*30*). A double sieve selection was able to identify mutants of the wild-type tRNA with excellent orthogonality, but there was also a significant loss of overall activity. Ideally, mutations could be introduced into the tRNA that would eliminate the cross-reactivity with *E. coli* synthetases but preserve high levels of suppression efficiency. Since aminoacyl-tRNA synthetases frequently recognize positions within the acceptor stem and discriminator base of tRNAs (*50*), it is likely that an *E. coli* synthetase that charges the orthogonal tRNA would have a positive recognition element in this region. If this determinant could be changed without destroying recognition by the foreign synthetase, activity could be preserved while eliminating the background reaction. When this strategy was applied to the HhL4-derived amber suppressor, such mutants were indeed found. Several mutants preserved or even improved suppression efficiency when coexpressed with MtLRS but had nearly background levels of amber suppression (7.5 vs 5 ug/mL ampicillin) in the absence of the synthetase. These mutants had reversed the third base pair from G:C to C:G, and an inspection of the recognition elements known for various *E. coli* synthetases suggests the identity of the *E. coli* synthetase that had cross-reacted with the HhL4-derived suppressor. Both GlnRS and LysRS of *E. coli* conserve G3:C70 and frequently cross-react with amber suppressor tRNAs (*23*). Because LysRS also conserves A73, this is the more likely candidate for the cross-reactive *E. coli* synthetase (*51*, *52*). Perhaps this strategy is a general solution to the problem of improving

the specificity of cross-reactive orthogonal tRNAs since most *E. coli* aminoacyl-tRNA synthetases contain positive determinants within the acceptor stem.

CONCLUSION

We have shown that the leucyl-tRNA synthetase from the archaean *M. thermoautotrophicum* and mutants of a halobacterial tRNA function as an orthogonal pair in *E. coli*. Mutagenesis experiments showed that the two most significant criteria that lead to efficient orthogonal amber suppressor tRNAs are a CU(X)XXXAA anticodon loop and the lack of noncanonical or mismatched base pairs in the stem regions. From these selections we have identified efficient amber, four-base, and opal orthogonal suppressor tRNAs. We have also devised a consensus strategy to rationally design efficient orthogonal tRNAs. We are currently carrying out experiments to change the amino acid specificity of the archaeal synthetase, as well as examining the generality of our consensus suppressor strategy with archaeal glutamate and lysine orthogonal pairs. Ultimately, it should be possible to combine this leucyl-orthogonal pair with the *M. jannaschii* pair to site-specifically incorporate two unique unnatural amino acids simultaneously into proteins *in vivo*.

ACKNOWLEDGMENT

We thank Priscilla Yang and Stephen Santoro for careful review of the manuscript.

SUPPORTING INFORMATION AVAILABLE

Detailed information for the cloning of archaeal leucyl-tRNA synthetases. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Hortin, G., and Boime, I. (1983) *Methods Enzymol. 96*, 777−784.
2. Doring, V., Mootz, H. D., Nangle, L. A., Hendrickson, T. L., de Crecy-Lagard, V., Schimmel, P., and Marliere, P. (2001) *Science 292*, 501−504.
3. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G. (1989) *Science 244*, 182−188.
4. Bain, J. D., Glabe, C. G., Dix, T. A., and Chamberlin, A. R. (1989) *J. Am. Chem. Soc. 111*, 8013−8014.
5. Dougherty, D. A. (2000) *Curr. Opin. Chem. Biol. 4*, 645−652.
6. Cornish, V. W., Mendel, D., and Schultz, P. G. (1995) *Angew. Chem., Int. Ed. 34*, 621−633.
7. Wang, L., Zhang, Z., Brock, A., and Schultz, P. G. (2003) *Proc. Natl. Acad. Sci. U.S.A. 100*, 56−61.
8. Chin, J. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 11020−11024.
9. Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) *J. Am. Chem. Soc. 124*, 9026−9027.
10. Liu, D. R., and Schultz, P. G. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 4780−4785.
11. Pastrnak, M., Magliery, T. J., and Schultz, P. G. (2000) *Helv. Chim. Acta 83*, 2277−2286.
12. Ohno, S., Yokogawa, T., Fujii, I., Asahara, H., Inokuchi, H., and Nishikawa, K. (1998) *J. Biochem. (Tokyo) 124*, 1065−1068.
13. Kowal, A. K., Kohrer, C., and RajBhandary, U. L. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 2268−2273.
14. Edwards, H., and Schimmel, P. (1990) *Mol. Cell. Biol. 10*, 1633−1641.
15. Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., Kobayashi, T., Kitabatake, M., Takio, K., Saito, K., Shirouzu, M., Hirao, I., and Yokoyama, S. (2002) *Nucleic Acids Res. 30*, 4692−4699.
16. Cload, S. T., Liu, D. R., Froland, W. A., and Schultz, P. G. (1996) *Chem. Biol. 3*, 1033−1038.
17. Piccirilli, J. A., Krauch, T., Moroney, S. E., and Benner, S. A. (1990) *Nature 343*, 33−37.
18. Hirao, I., Ohtsuki, T., Fujiwara, T., Mitsui, T., Yokogawa, T., Okuni, T., Nakayama, H., Takio, K., Yabuki, T., Kigawa, T., Kodama, K., Nishikawa, K., and Yokoyama, S. (2002) *Nat. Biotechnol. 20*, 177−182.
19. Wu, Y., Fa, M., Tae, E. L., Schultz, P. G., and Romesberg, F. E. (2002) *J. Am. Chem. Soc. 124*, 14626−14630.
20. Magliery, T. J., Anderson, J. C., and Schultz, P. G. (2001) *J. Mol. Biol. 307*, 755−769.
21. Anderson, J. C., Magliery, T. J., and Schultz, P. G. (2002) *Chem. Biol. 9*, 237−244.
22. Shimizu, M., Asahara, H., Tamura, K., Hasegawa, T., and Himeno, H. (1992) *J. Mol. Evol. 35*, 436−443.
23. Kleina, L. G., Masson, J. M., Normanly, J., Abelson, J., and Miller, J. H. (1990) *J. Mol. Biol. 213*, 705−717.
24. Sampson, J. R., and Saks, M. E. (1993) *Nuclic Acids Res. 21*, 4467−4475.
25. Soma, A., Uchiyama, K., Sakamoto, T., Maeda, M., and Himeno, H. (1999) *J. Mol. Biol. 293*, 1029−1038.
26. Lanyi, J. K. (1974) *Bacteriol. Rev. 38*, 272−290.
27. Miller, J. H. (1972) *Experiments in molecular genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
28. Hoben, P., and Soll, D. (1985) *Methods Enzymol. 113*, 55−59.
29. Kwok, Y., and Wong, J. T. (1980) *Can. J. Biochem. 58*, 213−218.
30. Wang, L., and Schultz, P. G. (2001) *Chem. Biol. 8*, 883−890.
31. Tuohy, T. M., Thompson, S., Gesteland, R. F., and Atkins, J. F. (1992) *J. Mol. Biol. 228*, 1042−1054.
32. Furter, R. (1998) *Protein Sci. 7*, 419−426.
33. Liu, D. R., Magliery, T. J., and Schultz, P. G. (1997) *Chem. Biol. 4*, 685−691.
34. Liu, D. R., Magliery, T. J., Pastrnak, M., and Schultz, P. G. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 10092−10097.
35. Mirande, M., Gache, Y., Le Corre, D., and Waller, J. P. (1982) *EMBO J. 1*, 733−736.
36. Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) *Science 292*, 498−500.
37. Tumbula, D., Vothknecht, U. C., Kim, H. S., Ibba, M., Min, B., Li, T., Pelaschier, J., Stathopoulos, C., Becker, H., and Soll, D. (1999) *Genetics 152*, 1269−1276.
38. Woese, C. R., Olsen, G. J., Ibba, M., and Soll, D. (2000) *Microbiol. Mol. Biol. Rev. 54*, 202−236.
39. Wang, L., Magliery, T. J., Liu, D. R., and Schultz, P. G. (2000) *J. Am. Chem. Soc. 122*, 5010−5011.
40. Yarus, M., Cline, S. W., Wier, P., Breeden, L., and Thompson, R. C. (1986) *J. Mol. Biol. 192*, 235−255.
41. Smith, D., Breeden, L., Farrell, E., and Yarus, M. (1987) *Nucleic Acids Res. 15*, 4669−4686.
42. Raftery, L. A., and Yarus, M. (1987) *EMBO J. 6*, 1499−1506.
43. Soderberg, T., and Poulter, C. D. (2000) *Biochemistry 39*, 6546−6553.
44. Sussman, J. L., and Kim, S. (1976) *Science 192*, 853−858.
45. Yarus, M. (1982) *Science 218*, 646−652.
46. Atkins, J. F., Weiss, R. B., Thompson, S., and Gesteland, R. F. (1991) *Annu. Rev. Genet. 25*, 201−228.
47. Buttcher, V., Senger, B., Schumacher, S., Reinbolt, J., and Fasiolo, F. (1994) *Biochem. Biophys. Res. Commun. 200*, 370−377.
48. Hou, Y. M., Schimmel, P., and Houweling, P. L. (1992) *Biochemistry 31*, 4157−4160.
49. Furdon, P. J., Guerrier-Takada, C., and Altman, S. (1983) *Nucleic Acids Res. 11*, 1491−1505.
50. Martinis, S. A., and Schimmel, P. (1995) in *tRNA: Structure, Biosynthesis, and Function* (Soll, D., and RajBhandary, U., Eds.) pp 349−370, ASM Press, Washington, DC.
51. Freist, W., and Gauss, D. H. (1995) *Biol. Chem. Hoppe-Seyler 376*, 451−472.
52. McClain, W. H., Chen, Y. M., Foss, K., and Schneider, J. (1988) *Science 242*, 1681−1684.

BI034550W