

# Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA

Randall A. Hughes<sup>1,2</sup> and Andrew D. Ellington<sup>1,3,\*</sup>

<sup>1</sup>Department of Chemistry and Biochemistry, The University of Texas at Austin, 1 University Station A5300, Austin, TX 78712, <sup>2</sup>Applied Research Laboratories, The University of Texas at Austin, 10000 Burnet Road, Austin, TX 78758 and <sup>3</sup>Institute for Cellular and Molecular Biology, The University of Texas at Austin, 1 University Station A4800, Austin, TX 78712, USA

Received February 18, 2010; Revised May 20, 2010; Accepted May 24, 2010

## ABSTRACT

While a number of aminoacyl tRNA synthetase (aaRS):tRNA pairs have been engineered to alter or expand the genetic code, only the *Methanococcus jannaschii* tyrosyl tRNA synthetase and tRNA have been used extensively in bacteria, limiting the types and numbers of unnatural amino acids that can be utilized at any one time to expand the genetic code. In order to expand the number and type of aaRS/tRNA pairs available for engineering bacterial genetic codes, we have developed an orthogonal tryptophanyl tRNA synthetase and tRNA pair, derived from *Saccharomyces cerevisiae*. In the process of developing an amber suppressor tRNA, we discovered that the *Escherichia coli* lysyl tRNA synthetase was responsible for misacylating the initial amber suppressor version of the yeast tryptophanyl tRNA. It was discovered that modification of the G:C content of the anticodon stem and therefore reducing the structural flexibility of this stem eliminated misacylation by the *E. coli* lysyl tRNA synthetase, and led to the development of a functional, orthogonal suppressor pair that should prove useful for the incorporation of bulky, unnatural amino acids into the genetic code. Our results provide insight into the role of tRNA flexibility in molecular recognition and the engineering and evolution of tRNA specificity.

## INTRODUCTION

Engineering the genetic code has recently emerged as a method to potentially create novel ‘allo-proteins’ which could have a myriad of applications in the pharmaceutical and biotechnology industries as well as serve as novel components for synthetic biology. To facilitate the site-specific incorporation of novel amino acids into

proteins, additional engineered orthogonal aminoacyl tRNA synthetase (aaRS):tRNA pairs can be expressed in cells [reviewed in (1,2)]. Orthogonal aaRS-tRNA pairs frequently take advantage of the cross-species differences in the recognition elements of aaRSs and their cognate tRNAs (3), and have utilized nonsense codons (3) or frameshift codons (4) to expand the canonical genetic code.

To date, there have been only eight completely unique orthogonal pairs reported for use in the prokaryotic (*Escherichia coli*) translational system. These orthogonal pairs include the aspartic acid (5), glutamine (6), and phenylalanine (7) aaRS/tRNA pairs from *Saccharomyces cerevisiae*; the glutamic acid pair from *Methanosarcina mazei* (8); the leucine pair from *Methanobacterium thermoautotrophicum* (9); the lysine pair from *Pyrococcus horikoshii* (4); the pyrrolysine pair from *Methanosarcinia barkeri* (10); and the tyrosyl pair from *Methanococcus jannaschii* (3). However, only the TyrRS-tRNA<sup>Tyr</sup> pair from *M. jannaschii* has been used extensively to expand the genetic code of bacteria (2).

The tryptophanyl tRNA synthetase (WRS) is part of the class Ic subclass of aaRSs that also includes the evolutionarily related tyrosyl tRNA synthetase (11). Based on the structural and phylogenetic studies of the tryptophanyl and the tyrosyl tRNA synthetases, it is generally believed that the WRS was the last synthetase to evolve (12–14). Both the tyrosyl tRNA and WRSs lack any sort of an editing domain. Instead, they rely on the interaction of conserved active site residues to specifically recognize and position their substrates into their binding pockets (15).

The lack of an editing domain in WRS and its capacious binding pocket for the largest natural amino acid have led to its use for the incorporation of amino acid analogs into proteins (16,17). The expanded substrate flexibility of WRS has even supported adaptive evolution of whole bacteria and bacteria phage proteomes to tryptophan analogs (18–20). However, the engineering of the WRS for the site-specific insertion of large, unnatural amino

\*To whom correspondence should be addressed. Tel: +1 512 471 6445; Fax: +1 512 471 7014; Email: andy.ellington@mail.utexas.edu

acids has largely been ignored, especially when compared with tyrosyl tRNA synthetase (2). Zhang *et al.* (21) have used the *Bacillus subtilis* WRS and an opal suppressor (anticodon UCA) variant of the *B. subtilis* tRNA<sup>Trp</sup> for the site-specific incorporation of 5-hydroxytryptophan into proteins in mammalian cells, but this orthogonal pair cannot be used in *E. coli* since they have overlapping recognition elements.

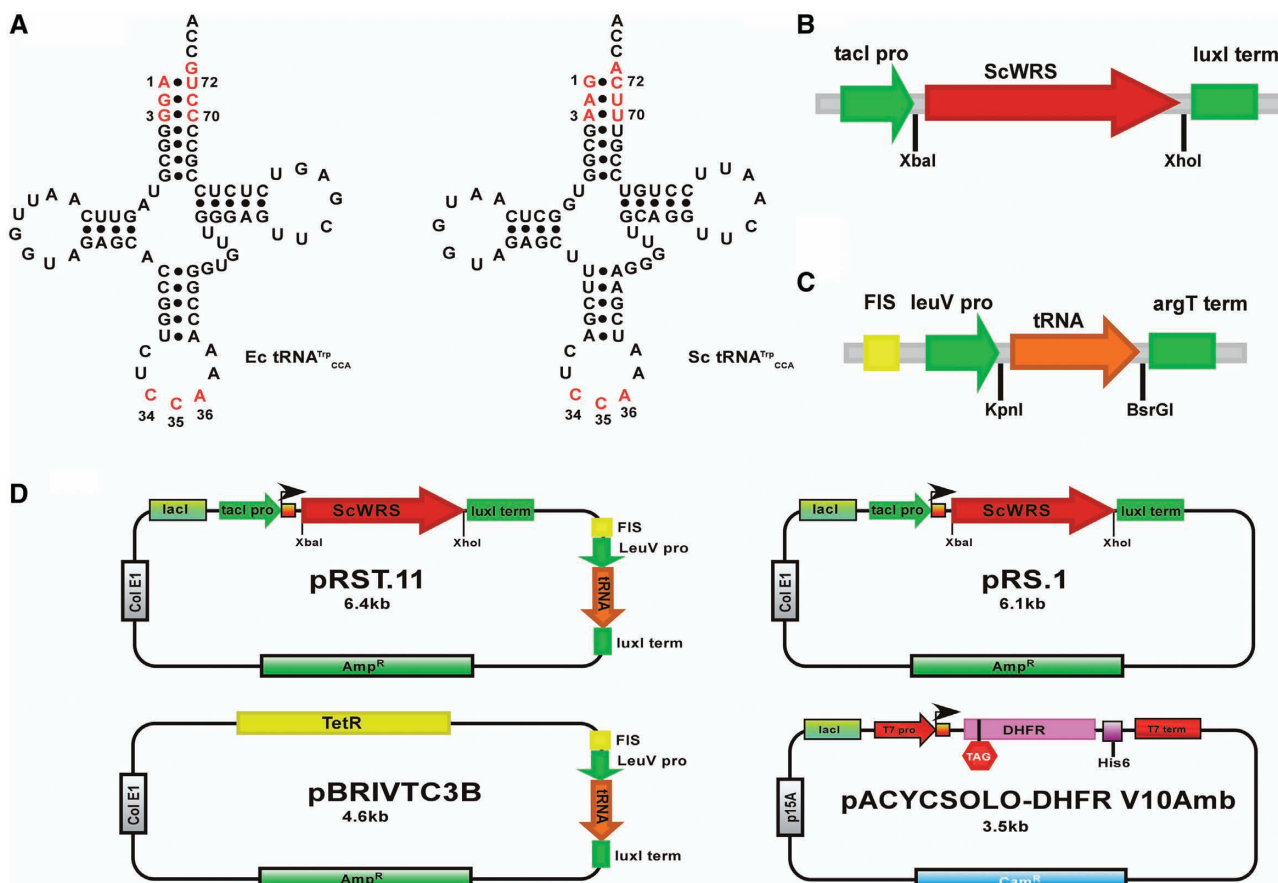
Therefore, to further explore the types of unnatural amino acids that can be site-specifically encoded in bacteria and to expand the potential utility of synthetic orthogonal pairs both *in vivo* and *in vitro*, we have sought to develop a new tryptophanyl orthogonal pair. The *E. coli* tRNA<sup>Trp</sup> has a major identity element at the G73 discriminator base and the anticodon CCA sequence, and additional weak identity elements in the first 3 bp of the acceptor stem (22). In contrast, yeast tRNA<sup>Trp</sup> has an adenosine residue at position 73, and the identity elements in the first 3 bp of the acceptor stem also differ from those found in the *E. coli* tRNA<sup>Trp</sup> [(23); Figure 1A]. Himeno *et al.* have shown that mutations at A73 and the G1-C72 base pair of the *E. coli* tRNA<sup>Trp</sup> that make it more similar to the yeast tRNA<sup>Trp</sup> also lead

to inactivity with the *E. coli* WRS. Moreover, amber suppressor versions of *E. coli* tRNA<sup>Trp</sup> (anticodon CUA) lose their ability to be recognized by the *E. coli* TrpRS and are instead aminoacylated by the *E. coli* GlnRS (24). However, this is not the case for the yeast tRNA<sup>Trp</sup>, which not only maintains its tryptophan identity in yeast cells but also is an efficient amber suppressor (25,26). Taken together, these results suggest that it should be possible to develop a new orthogonal pair for use in *E. coli* based on the WRS and the amber suppressor variant of its cognate tRNA from *S. cerevisiae*.

## MATERIALS AND METHODS

### Materials

*E. coli* strain, CSH108 (Genotype: F' [F128: LacZ8(Am), LacI373] Δ(gpt-lac)5, λ-, ara(FG), gyrA-(NAIR), argE-(Am), rpoB-(rif<sup>R</sup>), thi-1) was obtained from the *E. coli* Genetic Stock Center (Strain # 8081) <http://cgsc.biology.yale.edu/>. Top10 (for routine cloning) and BL21 (DE3)Star chemically competent cells were from



**Figure 1.** Tryptophan tRNAs and DNA expression constructs. (A) Cloverleaf representations of *E. coli* tRNA<sup>Trp</sup> and *S. cerevisiae* tRNA<sup>Trp</sup>. The known identity elements for each tRNA are shown in red. (B) Expression construct for ScWRS in plasmids pRS.1 and pRST.11B. (C) Expression construct for the Sc-tRNA<sup>Trp</sup><sub>Amb</sub> and its variants in plasmids pBRIVTC3B and pRST.11B. (D) Plasmid constructs. pRST.11B is the dual tRNA/aaRS expression vector containing both of the expression constructs shown in (B) and (C). Plasmid pRS.1 expresses only the ScWRS and is similar to pRST.11B except that it lacks the tRNA expression construct shown in (B). Plasmid pBRIVTC3B expresses only the tRNA. Plasmid pACYCSOLO-DHFR V10Amb is the target protein (DHFR) expression vector. The encoded DHFR gene contains an in-frame amber (TAG) mutation at the 10th amino acid position and a C-terminal HisTag.

Invitrogen (Carlsbad, CA, USA). Terrific Broth Powder was from VWR Scientific (West Chester, PA, USA).

All restriction enzymes, T4 DNA ligase and vector pBR322 were from New England Biolabs (Ipswich, MA, USA). Vector pACYCDuet-1 was obtained from Novagen (San Diego, CA, USA). Vector pET28b and KOD polymerase were from Novagen (Gibbstown, NJ, USA). Wizard SV Gel and PCR cleanup kit were from Promega (Madison, WI, USA). Oligonucleotides were obtained from Integrated DNA Technologies (IDT; Coralville, IA, USA). The Klenow fragment polymerase, SimplyBlue Safestain and NuPAGE 4–12% Bis-Tris precast gels were from Invitrogen (Carlsbad, CA, USA). Quickchange site-directed mutagenesis kits and Pfu polymerase were from Stratagene (La Jolla, CA, USA).

Protease inhibitor cocktail was from Roche (Indianapolis, IN, USA). Ni-NTA resin was obtained from Qiagen (Germantown, MD, USA). The anti-polyhistidine alkaline phosphatase primary conjugate antibody was from Sigma-Aldrich (St. Louis, MO, USA). All buffers and chemicals were obtained from Sigma-Aldrich (St. Louis, MO, USA) or Fisher Scientific (Waltham, MA, USA).

### Cloning of the *S. cerevisiae* WRS gene

The coding sequence for the WRS from *S. cerevisiae* (ScWRS) (27) was amplified using the PCR and the primers ScWRS.f-3 5'-TCGAAAAGCTTCCATGAGCAACGACGAAACTGTAGAG-3' (HindIII), ScWRS.r-3 5'-GCAGCCTCGAGTTACTTCTTTTCTTGCTTAGT TTTTGGC-3' (XhoI) from a glass-bead lysed yeast cell extract. The restriction sites introduced are underlined and specified in parentheses. The resulting approximately 1.3 kb PCR product was digested with XhoI and HindIII and ligated into a similarly digested pET-28b vector using T4 DNA ligase following the conditions recommended by the enzyme supplier. The ligation mixture was transformed into Top10 cells and the resulting kanamycin resistant colonies were screened via colony PCR for the ScWRS insert. Positive clones were verified by DNA sequencing from purified clonal plasmids using the primers T7 Terminator Primer 5'-GCTAGTTATTGCTCAGCGG-3' and T7 Promoter Primer 5'-TAATACGACTCACTATAGGG-3'. The resulting sequence-confirmed plasmid was termed pET28b-ScWRS. This expression vector expresses ScWRS from the T7 promoter with an N-terminal His-tag.

### Design and construction of the pRS.1 tRNA synthetase expression vector

In order to express the ScWRS gene in *E. coli*, the expression construct (module) shown in Figure 1B was constructed. This expression construct was designed to be modular in that different components of the construct could be replaced using unique restriction sites flanking each of the three portions of the construct (i.e. promoter-RBS, aaRS, transcription terminator). For example, the gene for the ScWRS shown in green in

Figure 1B could be replaced by restriction digestion with XbaI and XhoI. Since successful nonsense suppression is the result of an acylated suppressor tRNA successfully competing with endogenous release factors for termination codons, maintenance of high levels of acylated suppressor tRNAs within the cell is important, therefore we expressed ScWRS from the strong, synthetic tacI promoter (28) and the strong Shine Dalgarno sequence (AAGGAG) from the expression vector pET14b. Transcription termination was mediated by the luxICDABEG terminator sequence (Biobrick part: BBa\_B0021) obtained from the Biobrick parts registry at MIT ([http://partsregistry.org/Main\\_Page](http://partsregistry.org/Main_Page)). The overall construct was assembled by overlap extension PCR.

The plasmid construct for expressing the ScWRS is shown in Figure 1D. The aforementioned synthetase expression construct was cloned into the HindIII and EagI sites of pBR322 to yield the vector pRS-ScWRS. This cloning step removes the majority of the tetracycline resistance element found in pBR322, but leaves the ampicillin resistance element. This vector contains the ColE1 origin of replication which maintains the plasmid at 15–30 copies per cell (29). To help regulate the expression of the ScWRS from the strong tacI promoter, the expression sequence for LacIq was cloned into the ZraI and HindIII sites of this vector to yield vector pRS.1. The IPTG-induced expression of the ScWRS gene from this plasmid was verified by a western blot assay by probing for the C-terminal His-Tag in a His tagged version of ScWRS (in vector pRS.1ht) using an alkaline phosphatase conjugated anti-polyhistidine antibody (Sigma-Aldrich; St. Louis, MO, USA). As can be seen from Supplementary Figure S1, the ScWRS is efficiently expressed from this construct, though some aggregation in the pellet occurs due to overexpression. A more detailed description of the construction of pRS.1 and related vectors can be found in the Supplementary Data.

### Design and construction of the IVTC3 tRNA expression plasmids

To express the yeast tRNA<sup>Trp</sup> amber suppressor, the synthetic construct shown in Figure 1C was designed and built from overlapping oligonucleotides. This synthetic construct expresses the tRNA under the control of the tRNA leuV 5'-UTR (region from –112 to +33), which contains the stable RNA promoter –35 and –10 regions and an upstream FIS element (30,31). The FIS element was included in the 5'-UTR as it is known to enhance transcription initiation from this promoter (32). In addition, the +1 to +33 sequences from the leuV 5'-UTR was maintained as it contains the natural tRNA processing sequence. The tRNA sequence is flanked by unique KpnI and BsrGI sites to facilitate cloning of different tRNA sequences into the expression construct. Transcription termination is carried out by the rho independent terminator contained in the argT 3'-UTR (33). This construct was cloned into vector pBR322 or vector pRS.1 to yield the tRNA expression vectors pBRIVTC3 and pRST.11, respectively (Figure 1D). The expression and maturation of the yeast tRNA<sup>Trp</sup> amber suppressor

from these plasmid constructs *in vivo* was verified via northern blotting with a Sc-tRNA<sup>Trp</sup>-specific DNA oligonucleotide probe (Supplementary Figure S2). While the recombinant Sc-tRNA<sup>Trp</sup><sub>Amb</sub> was cleaved from the expressed pre-tRNA transcript at sufficient levels (~20-30%) to observe nonsense suppression *in vivo*, future improvements to the processing efficiency of tRNAs expressed from this construct could lead to increased nonsense suppression efficiency *in vivo*. A thorough description of the design, construction and assay of these constructs and vectors can be found in the Supplementary Data.

### Construction of vector pACYCSolo

Vector pACYCDuet-1 was digested with EcoNI and BsrGI to remove the MCS1 and redundant T7promoter. The double-digested plasmid band was gel-purified from a low-melt agarose gel using the Promega (Madison, WI, USA) PCR and gel purification kit. The linearized vector was 'blunted' by reaction with the Klenow fragment polymerase using the following conditions: linearized plasmid, 0.5mM each dNTP, 1× React2 Buffer (Klenow reaction buffer, supplied with enzyme) and 1.5U Klenow fragment. The elongation reaction was carried out at room temperature for 30min, followed by agarose gel purification. The blunted vector was closed with T4 DNA ligase under the following conditions: linearized vector, 1× T4 DNA ligase buffer (50mM Tris pH 7.5, 10mM MgCl<sub>2</sub>, 10mM DTT, 1mM ATP) and 20 U T4 DNA ligase. The ligation mixture was transformed into Top10 cells and colonies were screened via colony PCR using the primers pACYCDuet-f and pACYCDuet-r. Plasmids that appeared to contain the correct insert were submitted for sequencing at the ICMB DNA Core Facility. The sequence verified vector was named pACYCSolo. This vector contained a multiple cloning site sequence flanked by the T7lac promoter and a transcription terminator. It also contained the coding sequences for the Lac repressor and chloramphenicol acetyl transferase. This vector is a pACYC184 derivative and contains the P15A origin which is compatible with other vectors containing ColE1 origins.

### Construction of vectors pACYCSolo-DHFR and pACYCSolo-DHFR\_V10Amb

The coding sequence for *E. coli* dihydrofolate reductase (DHFR) and DHFR\_V10Amb was amplified from the plasmids pIVEX1.4WG-DHFR or pIVEX1.4WG-DHFR\_V10Amb (R. A. Hughes and A. D. Ellington, unpublished results), respectively, with primers DHFR-pivex1.4-f-1 5'-GTTACTTTACATATGATCAGTCTG ATTGCGGCGTTAGC (NdeI), DHFR-HIS6-r 5'-GCT GCTCGAGTTAATGATGATGATGATGATGATGGCTGC CCCGCCGCTCCAGAATCTC-3'(XhoI). The restriction sites used for cloning are underlined. The flanking primer, DHFR-His6-r, encoded a C-terminal His-tag (amino acid sequence GSGHHHHHH). The amplified genes were digested with NdeI and XhoI and cloned into similarly digested pACYCSolo. Cloned sequences were screened by colony PCR for insertion of the gene sequence and

verified by sequencing with primers pACYCDuet-f 5'-TT GCGCCATTCGATGGTGTGTC and pACYCDuet-r 5'-AA AACCCCTCAAGACCCCGTT. Sequence verified clones were designated pACYCSolo-DHFR and pACYCSolo-DHFR\_V10 (Figure 1D) for the expression constructs containing C-terminal His-tagged DHFR and DHFR\_V10Amb, respectively.

### DHFR nonsense suppression assay

Plasmids pBRIVTC3B and pRST.11B containing Sc-tRNA<sup>Trp</sup><sub>Amb</sub> or both ScWRS and Sc-tRNA<sup>Trp</sup><sub>Amb</sub>, respectively, were co-expressed with pACYCSolo-DHFR\_V10Amb in BL21(DE3) or BL21(DE3)Star cells. A control strain containing pACYCSolo-DHFR\_V10Amb and pBR322, which lacks both ScWRS and Sc-tRNA<sup>Trp</sup><sub>Amb</sub>, was used to determine background suppression rates. The strains containing these plasmids were grown in Terrific Broth at 37°C to OD<sub>600</sub> ~0.7-0.8 at which point the expression of DHFR and ScWRS was induced by the addition of IPTG to 1mM. Cells were grown overnight at 37 or 30°C. Following induction and co-expression, the cells were collected by centrifugation and resuspended in binding buffer (50mM Tris, pH 8.0, 0.5M NaCl, 5mM imidazole) containing 1× protease inhibitor cocktail and 1mg/ml lysozyme. The resuspended cells were lysed by sonication on ice using 30% probe amplitude and five pulse cycles (30s ON, 15s OFF). Following sonication, the cell debris was pelleted via centrifugation at 10000g for 20min. The cleared lysate was transferred to a new centrifuge tube and centrifuged again at 11000g for 15min to remove any remaining insoluble material.

The His-tagged DHFR was purified via immobilized metal affinity chromatography (IMAC) (34). The cleared lysate was applied to a 1ml (bead volume) Ni-NTA gravity column pre-incubated with 4× column volumes of binding buffer and allowed to enter the column by gravity flow. The column was washed once with 10× column volumes of Binding Buffer, once with 3× column volumes of Wash Buffer 1 (50mM Tris, pH 8.0, 0.5M NaCl, 20mM imidazole) and once with 3× column volumes of Wash Buffer 2 (50mM Tris, pH 8.0, 0.5M NaCl, 30mM imidazole). DHFR-His6 was eluted from the Ni-NTA column by the addition of 4× column volumes of Elution Buffer (50mM Tris, pH 8.0, 0.5M NaCl, 250mM imidazole). Protein fractions were analyzed by SDS-PAGE on a 4-12% Bis-Tris NuPAGE developed in MES SDS buffer (50mM MES pH 7.3, 50mM Tris, 3.5mM SDS, 1mM EDTA) under reduced, denaturing conditions and stained using SimplyBlue Safestain. Each fraction was normalized to the wet cell pellet weight prior to loading on the gel. Purified proteins were quantitated using a modified Bradford Assay (Bio-Rad; Hercules, CA, USA).

### Edman N-terminal protein sequencing

Proteins (usually ~300pmol) to be sequenced were separated on a 4-12% NuPAGE Bis-Tris gradient gels in MES SDS buffer and transferred to a Polyvinylidene fluoride (PVDF) membrane using the Bio-Rad semidry

transfer apparatus (Biorad; Hercules, CA, USA) in NuPAGE transfer buffer (25 mM Bis-Tris, pH 7.2, 25 mM Bicine, 1 mM EDTA, 2 mM DTT, 20% methanol). The transfer was performed at 25 V for 1 h. Following transfer, the membrane was washed with HPLC grade water containing 2.5 mM DTT for 2 min. The membrane was then stained in 0.1% Coomassie Blue + 2 mM DTT until the protein bands were visible. The membrane was destained via multiple washes in 50% methanol, 10% acetic acid, and 1 mM DTT. The destained membrane was washed in HPLC grade water containing 1 mM DTT for 5–10 min. Destained blots were submitted to the Protein Core Facility (Institute for Cellular and Molecular Biology, The University of Texas at Austin) where the first 15–20 amino acids of the membrane-bound 18 kDa DHFR band were excised and sequenced according to Edman degradation protocols on an Applied Biosystems model 477 protein sequencer (ABI; Carlsbad, CA, USA). Sequenced amino acids were identified and quantitated using reference standards. Identification of the amino acid substituted at amber codon 10 of the modified DHFR protein was done by subtractive comparison of the chromatographs produced for position 9 from those produced at position 10. The identity of the amino acid incorporated at position 10 within DHFR was defined as the most abundant (in pmol) amino acid in the subtracted spectrum.

### Construction and cloning of Sc-tRNA<sup>Trp</sup> variants

All of the tRNA gene sequences were designed to be cloned between the KpnI and BsrGI sites of the tRNA expression cassette in vectors pBRIVTC3B and pRST.11B. The constructs contained tRNA processing sequences (following the KpnI site and preceding the BsrGI site).

The construct sequences were assembled from four overlapping oligonucleotides according to the gene assembly PCR procedure developed by Stemmer (35). The component ~60 nt oligonucleotides were designed using DNAWorks software (available at: <http://helixweb.nih.gov/dnaworks/>) (36). The oligonucleotides necessary to assemble the tRNA sequences are shown in Table 1. Each tRNA variant was assembled from four overlapping oligonucleotides (numbered XXX.1–XXX.4). The oligonucleotides were resuspended in sterilized water to a final concentration of 250 μM, and were mixed together to give a 2.5 μM oligonucleotide mixture of each oligonucleotide. An initial assembly PCR was carried out under the following conditions: 1 μl 2.5 μM oligo mix, 5 μl 10× PfuTurbo buffer, 2.5 μl 4 mM dNTP mix, 1 μl Pfu polymerase (added after an initial denaturation step) and water to 50 μl total volume. The assembly reactions were thermally cycled under the following regime: (1) 95°C-5min, (2) 95°C-30sec, (3) 50°C-30sec, (4) 72°C-60sec, (5) Go to step (2) 25 times and (6) 72°C-10min. Full-length genes were subsequently amplified by PCR: 1 μM assembly PCR, 5 μl 10× PfuTurbo buffer, 2.5 μl 4 mM dNTP mix, 1 μl of each flanking primer (20 μM; Table 1), 1 μl Pfu polymerase (added after an initial denaturation step) and water to 50 μl total volume. The amplification reactions were

thermally cycled: (1) 95°C-5min, (2) 95°C-30sec, (3) 62°C-30sec, (4) 72°C-60sec, (5) Go to step (2) 25 times and (6) 72°C-10min. The final PCR was separated on agarose gels to verify the presence of full-length product.

The assembled tRNA constructs were digested with KpnI and BsrGI and ligated into similarly digested pBRIVTC3B and pRST.11B with T4 DNA ligase. Clones were screened via colony PCR and sequences were confirmed by sequencing with primers pRS1871-94 5'-AACCCCTTGGCAGAACATATCCATC-3' or pRS2022-47r 5'-CTCGCGTATCGGTGATTCATTCTGCT-3'. Plasmids containing the tRNA variants are designated by the following nomenclature: plasmid name-tRNA variant (i.e. pBRIVTC3B-AS3.4).

### β-galactosidase colony spot assay

Plasmids pBRIVTC3B (tRNA only) and pRST.11B (ScWRS and tRNA) containing Sc-tRNA<sup>Trp</sup><sub>Amb</sub> variants were transformed into strain CSH108 via electroporation. Transformed colonies were picked and suspended in 10 μl of LB broth. Two microliter aliquots of this mixture were spotted in triplicate onto LB agar plates containing 50 μg/mL ampicillin, 1 mM IPTG and 0.2 mg/mL X-gal. The spotted plates were incubated at 37°C overnight.

LB broth cultures (3 ml) containing 50 μg/ml ampicillin and 1 mM IPTG were also inoculated (10 μl) and grown overnight at 37°C. Some 2 ml of each of the cultures was spun down and the cells lysed in 300 μl of B-PER reagent containing 1 × protease inhibitor and 200 μg/ml lysozyme. The cell debris was removed by centrifugation at 15000g for 15 min. The protein concentration in the lysate was determined using a modified Bradford Assay. The β-galactosidase activity in the lysate samples was assayed using the high sensitivity β-galactosidase assay kit from Stratagene (Cedar Creek, TX, USA). Triplicate samples from each culture were measured by optical absorbance (at 570 nm) of the chromogenic product, chlorophenol red.

## RESULTS

### Assaying the functionality of the ScWRS/Sc-tRNA<sup>Trp</sup><sub>Amb</sub> pair

In order to test the functionality of the ScWRS/Sc-tRNA<sup>Trp</sup><sub>Amb</sub> pair, the plasmids pBRIVTC3B, pRS.1 and pRST.11B which express only the Sc-tRNA<sup>Trp</sup><sub>Amb</sub>, only the ScWRS or both the ScWRS and Sc-tRNA<sup>Trp</sup><sub>Amb</sub> together, respectively, were transformed into *E. coli* strain CSH108. This strain contains an episomal lacZ gene with an amber nonsense codon (TAG) in a position that has previously been shown to be tolerant of a variety of amino acid substitutions. If a suppressor tRNA is active *in vivo*, then full-length β-galactosidase should be produced resulting in a blue-colored colony upon cleavage of the galactoside analog, 5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside (X-gal). A translationally inactive suppressor or an orthogonal suppressor in the absence of its cognate aaRS will yield

**Table 1.** Oligonucleotides used to construct suppressor tRNAs

| Oligonucleotide name | Sequence (5'–3')  |
|----------------------|---|
| RH-ScWA_U30G40.1     | GCGCCTCCGTGGTACCAATTCTTTTAAAGAATTGATGGTATGAAGCGGTGGC          |
| RH-ScWA_U30G40.2     | CCTGCAACCCTTCCATTTAGAGTAGAAAAGCTCTACCATTGAGCCACCGCTTCATACCATC |
| RH-ScWA_U30G40.3     | TCTAAATGGAAGGGTTGCAGGTTCAATTCCTGTCCGTTTACCAATTTTGAACCCCGCTT   |
| RH-ScWA_U30G40.4     | GTGACGAAATGTACAGAAAAACAAAAACCCCGCCGAAGCGGGGTTCAAATTGGT        |
| RH-ScWA_C70.1        | GCGCCTCCGTGGTACCAATTCTTTTAAAGAATTGATGGTATGAAGCGGTGG           |
| RH-ScWA_C70.2        | GCAACCCCTTCGATTTAGAGTCGAAAGCTCTACCATTGAGCCACCGCTTCATACCATCAAT |
| RH-ScWA_C70.3        | GACTCTAAATCGAAGGGTTGCAGGTTCAATTCCTGTCCGTTTACCAATTTTGAACCCCG   |
| RH-ScWA_C70.4        | GTGACGAAATGTACAGAAAAACAAAAACCCCGCCGAAGCGGGGTTCAAATTGGTGAAG    |
| RH-ScWA_C70U30G40.1  | GCGCCTCCGTGGTACCAATTCTTTTAAAGAATTGATGGTATGAAGCGGTGGC          |
| RH-ScWA_C70U30G40.2  | CCTGCAACCCTTCCATTTAGAGTAGAAAAGCTCTACCATTGAGCCACCGCTTCATACCATC |
| RH-ScWA_C70U30G40.3  | TCTAAATGGAAGGGTTGCAGGTTCAATTCCTGTCCGTTTACCAATTTTGAACCCCGCTT   |
| RH-ScWA_C70U30G40.4  | GTGACGAAATGTACAGAAAAACAAAAACCCCGCCGAAGCGGGGTTCAAATTGGT        |
| RH-ScWA_AS1.1        | GCGCCTCCGTGGTACCAATTCTTTTAAAGAATTGATGGTATGAAGCGGTGGC          |
| RH-ScWA_AS1.2        | ACCTGCAACCCGCGGTTTAGAGCGCGGAGCTCTACCATTGAGCCACCGCTTCATACCAT   |
| RH-ScWA_AS1.3        | CGGCGGTTGCAGGTTCAATTCCTGTCCGTTTACCAATTTTGAACCCCGCTTCGGCGGG    |
| RH-ScWA_AS1.4        | GTGACGAAATGTACAGAAAAACAAAAACCCCGCCGAAGCGGG                    |
| RH-ScWA_AS2.1        | GCGCCTCCGTGGTACCAATTCTTTTAAAGAATTGATGGTATGAAGCGGTG            |
| RH-ScWA_AS2.2        | CAACCCCGCGCTTTAGAGGCGGAGCTCTACCATTGAGCCACCGCTTCATACCATCAATT   |
| RH-ScWA_AS2.3        | CTCTAAATGGAAGGGTTGCAGGTTCAATTCCTGTCCGTTTACCAATTTTGAACCCCGCT   |
| RH-ScWA_AS2.4        | GTGACGAAATGTACAGAAAAACAAAAACCCCGCCGAAGCGGGGTTCAAATTGGT        |
| RH-ScWA_C70AS1.1     | GCGCCTCCGTGGTACCAATTCTTTTAAAGAATTGATGGTATGAAGCGGTGGC          |
| RH-ScWA_C70AS1.2     | ACCTGCAACCCGCGGTTTAGAGCGCGGAGCTCTACCATTGAGCCACCGCTTCATACCAT   |
| RH-ScWA_C70AS1.3     | CGGCGGTTGCAGGTTCAATTCCTGTCCGTTTACCAATTTTGAACCCCGCTTCGGCGGG    |
| RH-ScWA_C70AS1.4     | GTGACGAAATGTACAGAAAAACAAAAACCCCGCCGAAGCGGG                    |
| RH-ScWA_C70AS2.1     | GCGCCTCCGTGGTACCAATTCTTTTAAAGAATTGATGGTATGAAGCGG              |
| RH-ScWA_C70AS2.2     | CCGCGCTTTAGAGGCGGAGCTCTACCATTGAGCCACCGCTTCATACCATCAATTCTTAA   |
| RH-ScWA_C70AS2.3     | CGCCTCCTAAGCGCGGGTTGCAGGTTCAATTCCTGTCCGTTTACCAATTTTGAACCCCG   |
| RH-ScWA_C70AS2.4     | GTGACGAAATGTACAGAAAAACAAAAACCCCGCCGAAGCGGGGTTCAAATTGGTGAAG    |
| RH-ScWA_AS3.1        | TAGAGCTGCCGCTCTAAAGCGGCGGGTTGC                                |
| RH-ScWA_AS3.2        | GCAACCCCGCGCTTTAGAGGCGGCGAGCTCTA                              |
| RH-ScWA_AS3.2.1      | GCGCCTCCGTGGTACCAATTCTTTTAAAGAATTGATGGTATGAAGCGGTGG           |
| RH-ScWA_AS3.2.2      | TGCAACCCGCGGTTTAGAGCGCCAGCTCTACCATTGAGCCACCGCTTCATACCATCAA    |
| RH-ScWA_AS3.2.3      | AACCGCCGGGTTGCAGGTTCAATTCCTGTCCGTTTACCAATTTTGAACCCCGCTTCGGC   |
| RH-ScWA_AS3.2.4      | GTGACGAAATGTACAGAAAAACAAAAACCCCGCCGAAGCGGGTTCA                |
| RH-ScWA_AS3.3.1      | GCGCCTCCGTGGTACCAATTCTTTTAAAGAATTGATGGTATGAAGCGGTGG           |
| RH-ScWA_AS3.3.2      | TGCAACCCCGGTTTAGAGCGGGAGCTCTACCATTGAGCCACCGCTTCATACCATCAA     |
| RH-ScWA_AS3.3.3      | TAAACCGGGGGTTGCAGGTTCAATTCCTGTCCGTTTACCAATTTTGAACCCCGCTTCG    |
| RH-ScWA_AS3.3.4      | GTGACGAAATGTACAGAAAAACAAAAACCCCGCCGAAGCGGGGTTCAAATTG          |
| RH-ScWA_AS3.4.1      | GCGCCTCCGTGGTACCAATTCTTTTAAAGAATTGATGGTATGAAGCGGTGG           |
| RH-ScWA_AS3.4.2      | TGCAACCCGCGCTTTAGAGGCGCCAGCTCTACCATTGAGCCACCGCTTCATACCATCAA   |
| RH-ScWA_AS3.4.3      | AAGCGCCGGGTTGCAGGTTCAATTCCTGTCCGTTTACCAATTTTGAACCCCGCTTCGGC   |
| RH-ScWA_AS3.4.4      | GTGACGAAATGTACAGAAAAACAAAAACCCCGCCGAAGCGGGGTTCA               |
| RH-ScWA_AS3.5.1      | GCGCCTCCGTGGTACCAATTCTTTTAAAGAATTGATGGTATGAAGCGGTGG           |
| RH-ScWA_AS3.5.2      | CAACCCCGCGCTTTAGAGGCGGGAGCTCTACCATTGAGCCACCGCTTCATACCATCAATT  |
| RH-ScWA_AS3.5.3      | CTCTAAACCGGGGGTTGCAGGTTCAATTCCTGTCCGTTTACCAATTTTGAACCCCGCT    |
| RH-ScWA_AS3.5.4      | GTGACGAAATGTACAGAAAAACAAAAACCCCGCCGAAGCGGGGTTCAAATTGGT        |
| RH-ScWA_AS3.6.1      | GCGCCTCCGTGGTACCAATTCTTTTAAAGAATTGATGGTATGAAGCGGTGG           |
| RH-ScWA_AS3.6.2      | CAACCCCGCGGTTTAGAGCGCGGAGCTCTACCATTGAGCCACCGCTTCATACCATCAATT  |
| RH-ScWA_AS3.6.3      | CTCTAAACCGCGGGTTGCAGGTTCAATTCCTGTCCGTTTACCAATTTTGAACCCCGCT    |
| RH-ScWA_AS3.6.4      | GTGACGAAATGTACAGAAAAACAAAAACCCCGCCGAAGCGGGGTTCAAATTGGT        |

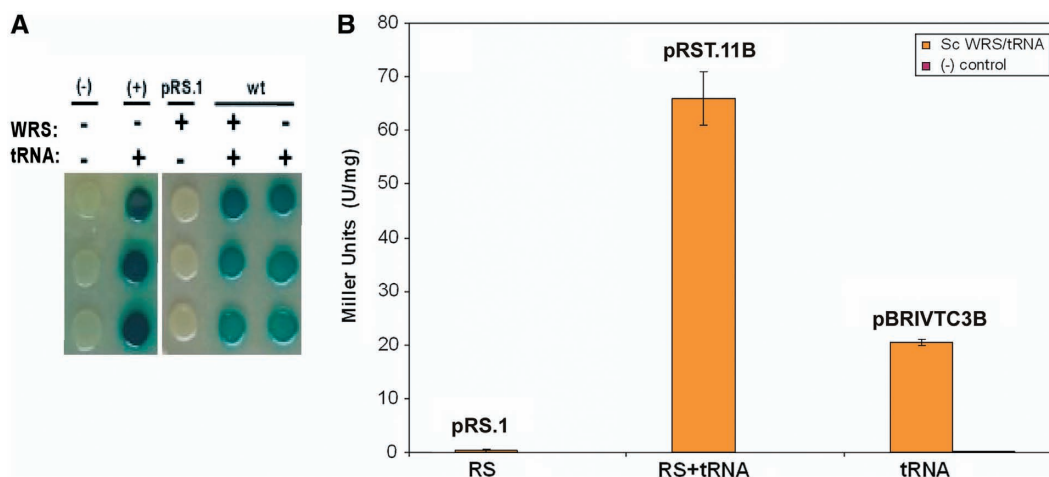
in a white colored colony due to in-frame translation termination.

As can be seen in Figure 2, when the ScWRS/tRNA<sup>Trp</sup><sub>Amb</sub> pair was expressed in CSH108, a blue colony was formed, indicating that the heterologous yeast suppressor tRNA was active in the bacterial translational system. However, a blue colony was also obtained when the Sc-tRNA<sup>Trp</sup><sub>Amb</sub> was expressed by itself in CSH108, indicating that this tRNA was not completely orthogonal to the set of *E. coli* tRNA synthetases. That said, when these strains were grown in solution, the tryptophanyl pair produced 66 Miller units of β-galactosidase per milligram of total protein (U/mg), whereas the strain carrying only Sc-tRNA<sup>Trp</sup><sub>Amb</sub>

suppressor produced 21 U/mg of enzyme. While it was likely that the ScWRS was being expressed in an active form and was able to aminoacylate its cognate suppressor tRNA in *E. coli*, the background aminoacylation activity of the Sc-tRNA<sup>Trp</sup><sub>Amb</sub> required additional modification in order to create an orthogonal suppressor.

#### Identifying background mischarging

To determine which aaRS was responsible for misacylating the yeast tryptophanyl suppressor tRNA (Sc-tRNA<sup>Trp</sup><sub>Amb</sub>), the plasmid pBRIVTC3B that expresses the Sc-tRNA<sup>Trp</sup><sub>Amb</sub> was co-transformed with pACYCSolo-DHFR\_V10Amb, a plasmid that expresses



**Figure 2.** LacZ suppression assay with Sc-tRNA<sup>Trp</sup><sub>Amb</sub>. (A) β-galactosidase suppression assay for ScWRS/tRNA<sup>Trp</sup><sub>Amb</sub> pair activity. Triplicate spots from the same initial culture are shown in column orientation. Cultures transformed with (+) or without (-) the Sc-tRNA<sup>Trp</sup><sub>Amb</sub> or ScWRS are shown in the right panel. The (-) negative control culture represents the background level of β-galactosidase activity in the absence of a suppressor tRNA. The (+) control represents β-galactosidase activity in the presence of a known amber suppressor tRNA [SupF, (Ec-tRNA<sup>Tyr</sup><sub>Amb</sub>)] which is expressed from the pBRIVTC3B plasmid similar to the experimental suppressor tRNAs. (B) Quantitation of β-galactosidase activity produced by the suppression assay shown in A. pRS.1 expresses just the ScWRS, pRST.11B expresses the ScWRS/tRNA<sup>Trp</sup><sub>Amb</sub> pair and pBRIVTC3B expresses just the Sc-tRNA<sup>Trp</sup><sub>Amb</sub> suppressor. Error bars represent 1 SD (*n* = 3).

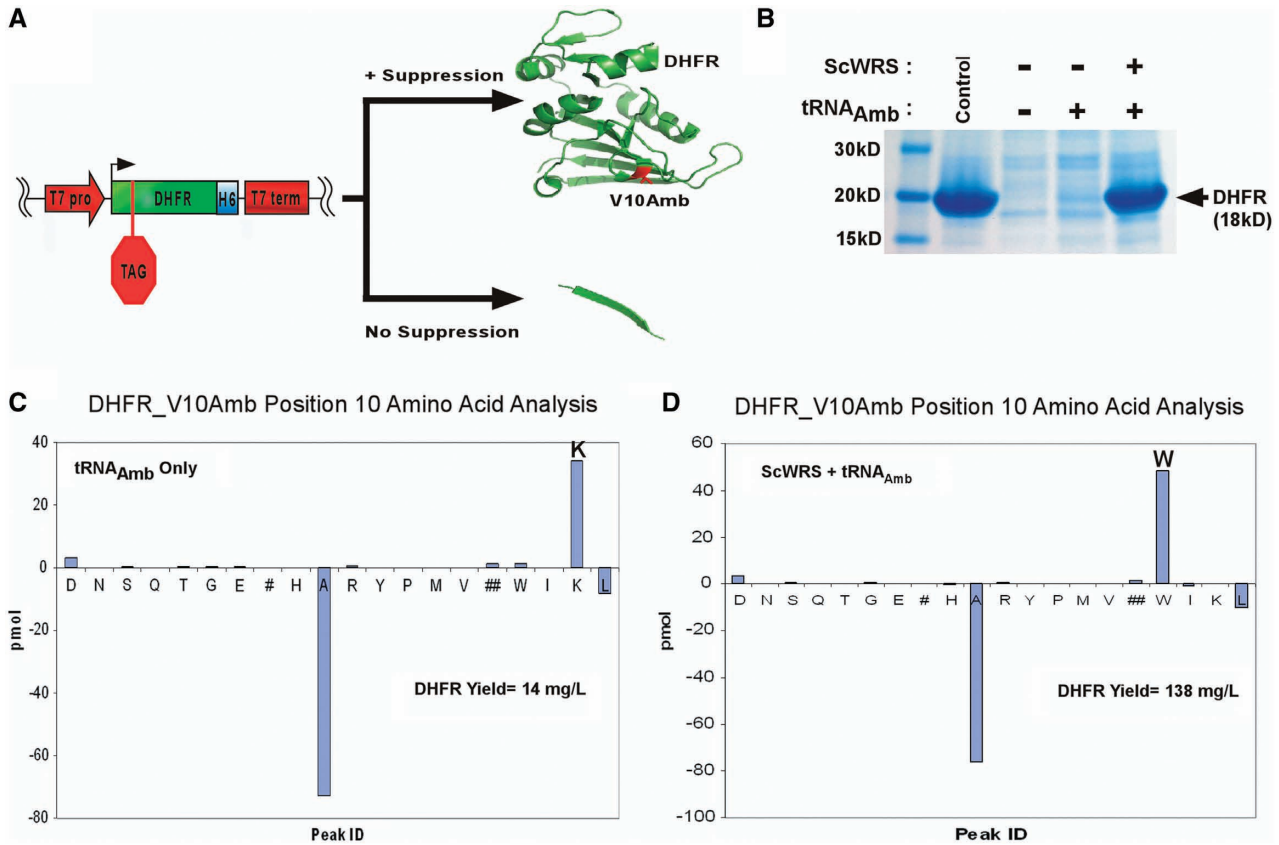
the protein DHFR. The DHFR protein sequence contains an amber nonsense codon (TAG) at amino acid position 10 (V10Amber). Therefore, only in the presence of an amber suppressor tRNA is full-length DHFR translated (Figure 3A). Expression of the pACYCSolo-DHFR\_V10Amb plasmid in the absence of the Sc-tRNA<sup>Trp</sup><sub>Amb</sub> suppressor does not yield any DHFR as determined by a Comassie stained SDS-PAGE gel (control (-) bands, Figure 3B). The DHFR expressed in the presence of the suppressor was purified via its C-terminal HisTag and submitted for N-terminal Edman peptide sequencing (37). As can be seen in Figure 3C, the Sc-tRNA<sup>Trp</sup><sub>Amb</sub> suppressor mediated the incorporation of lysine (K) at position 10 of DHFR. Indicating that the Sc-tRNA<sup>Trp</sup><sub>Amb</sub> was being misacylated by the *E. coli* lysine tRNA synthetase (EcKRS) when expressed alone in *E. coli*. In contrast, when pACYCSolo-DHFR\_V10Amb was co-transformed with plasmid, pRST.11B, which expresses both the ScWRS and the Sc-tRNA<sup>Trp</sup><sub>Amb</sub> suppressor DHFR contained primarily tryptophan at position 10 of DHFR (Figure 3D), again confirming the activity of ScWRS in *E. coli*.

As was the case with β-galactosidase assay, when the strains were grown in solution the suppressor tRNA led to tryptophan being incorporated preferentially to lysine (actually exclusively, within limits of detection for the assay). The ScWRS/Sc-tRNA<sup>Trp</sup><sub>Amb</sub> pair produced 138 milligrams of DHFR per liter of bacterial culture (mg/l), versus 14 mg/l for the Sc-tRNA<sup>Trp</sup><sub>Amb</sub> suppressor alone and 418 mg/l in the absence of the stop codon. Thus, the suppression efficiency of the ScWRS/Sc-tRNA<sup>Trp</sup><sub>Amb</sub> pair can be estimated to be about 33% versus 3% for the Sc-tRNA<sup>Trp</sup><sub>Amb</sub> alone. The greater discrimination against lysine in the DHFR expression assay relative to the β-galactosidase expression assay may be due to the fact that protein expression alone was monitored, as opposed to enzymatic activity.

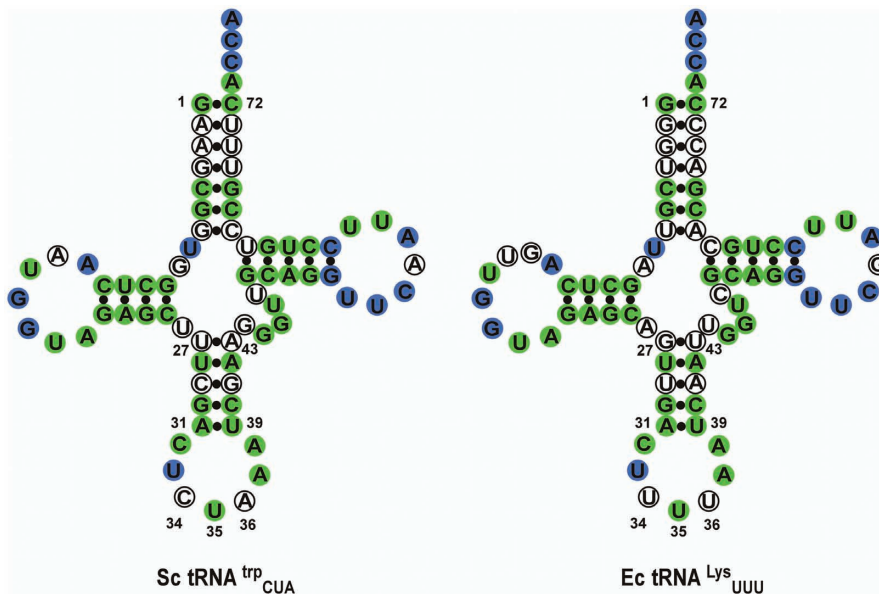
### Rational design of an orthogonal tRNA<sup>Trp</sup> amber suppressor

Since the EcKRS was determined to be responsible for causing the background aminoacylation of the Sc-tRNA<sup>Trp</sup><sub>Amb</sub>, we compared the Sc-tRNA<sup>Trp</sup><sub>Amb</sub> and the *E. coli* tRNA<sup>Lys</sup> (Ec-tRNA<sup>Lys</sup>) sequences to determine what features these tRNAs might have in common. Figure 4 shows the cloverleaf secondary structure of both tRNAs with conserved residues colored in green (identical between Ec-tRNA<sup>Lys</sup> and Sc-tRNA<sup>Trp</sup><sub>Amb</sub>) or blue (universally conserved amongst tRNAs). Surprisingly, the yeast tRNA<sup>Trp</sup><sub>Amb</sub> suppressor shares 73% sequence identity with the Ec-tRNA<sup>Lys</sup>. More importantly, Sc-tRNA<sup>Trp</sup><sub>Amb</sub> shares several key lysyl identity determinants with the Ec-tRNA<sup>Lys</sup>. In particular, the A73 discriminator base is a major identity element for both the EcKRS (38,39) and the *S. cerevisiae* WRS (40). In addition to this discriminator base, the anticodon sequence UUU for Ec-tRNA<sup>Lys</sup> (38,39) and CCA for Sc-tRNA<sup>Trp</sup><sub>Amb</sub> (23) are important identity elements for recognition by their cognate aaRSs. In the case of the yeast amber suppressor Sc-tRNA<sup>Trp</sup><sub>Amb</sub>, since the CCA anticodon sequence has been changed to CUA, the critical lysine identity element U35 (39) is inadvertently added to the anticodon sequence of the suppressor tRNA. While the U34 anticodon base in Ec-tRNA<sup>Lys</sup> is normally modified to 5-[(methylamino)-methyl]-2-thiouridine (mnm<sup>5</sup>s<sup>2</sup>U) to help read the rare AAG lysine codon (since *E. coli* lacks an isoacceptor with a CUU anticodon), the C34 in the amber suppressor is accommodated equally well (41).

The fact that the CUA amber suppressor anticodon in Sc-tRNA<sup>Trp</sup><sub>Amb</sub> is cross-reactive is not unique. In fact, the importance of the A73 discriminator and U35 anticodon bases for lysine recognition leads amber suppressor tRNAs derived from yeast tRNAs for the other



**Figure 3.** DHFR suppression assay. (A) The gene for DHFR containing a V10Amer mutation on plasmid pACYCSolo-DHFR\_V10Amb produces full-length protein when a functional nonsense suppressor is co-expressed with it. (B) Coomassie stained SDS-PAGE gel showing purification fractions from an immobilized nickel chromatography column for DHFR\_V10Amb expressed without Sc-tRNA<sup>Trp</sup><sub>Amb</sub> or ScWRS (-) with Sc-tRNA<sup>Trp</sup><sub>Amb</sub> only (tRNA<sub>amb</sub>) or with both ScWRS and Sc-tRNA<sup>Trp</sup><sub>Amb</sub> (RS +tRNA<sub>amb</sub>). M = marker, W = wash fractions and E = elution fractions. Full length DHFR appears as an 18 kDa protein band. (C) N-terminal sequencing results for DHFR position 10 from a culture containing pBRIVTC3B (Sc-tRNA<sup>Trp</sup><sub>Amb</sub> only). The bar graph represents the detected chromatographic amino acid abundances for position 10 of DHFR minus those from position 9. Positive abundance represents occupation at position 10 of the DHFR sample, whereas the negatively abundant peaks are carried over from the previous round of sequencing (position 9). (D) N-terminal sequencing results for DHFR position 10 from a culture containing pRST.11B (ScWRS and Sc-tRNA<sup>Trp</sup><sub>Amb</sub>). The yields of purified DHFR produced via suppression of the V10Amer codon are given for each sample.



**Figure 4.** Comparing Sc-tRNA<sup>Trp</sup><sub>Amb</sub> and Ec-tRNA<sup>Lys</sup><sub>UUU</sub>. Cloverleaf structures of both tRNAs showing the universally conserved residues in blue and residues conserved between the two tRNAs in green, for a total of 73% sequence identity overall.



aromatic amino acids (Phe, Tyr) to be cross-reactive to the EcKRS *in vivo* (42,43). These tRNAs also all share a common G-C base pair between positions 1 and 72 in the acceptor stem, and an anticodon loop sequence that is similar to that of Ec-tRNA<sup>Lys</sup>.

Since the anticodon loop and the 1–72 base pair sequences are required for translation activity in most tRNAs, it is unlikely that they can be altered in order to enhance discrimination and orthogonality between the Sc-tRNA<sup>Trp</sup><sub>Amb</sub> and Ec-tRNA<sup>Lys</sup><sub>UUU</sub>. Instead, it has previously been hypothesized that cross recognition may also be due to the structural plasticity of amber suppressor tRNAs to adapt to the KRS binding site (44). A number of facts support this hypothesis. While the crystal structure of the KRS-tRNA<sup>Lys</sup> complex has not been determined in its entirety, the anticodon binding domain has been studied in some detail (41,45–47). These structural and biochemical analyses of KRS reveals that it undergoes a dramatic change in conformation upon binding lysine as well as upon the formation of the Lys-AMP adenylate (41,48). Recently, it has been shown that the anticodon binding domain of KRS can enhance the binding efficiency of the lysyl-adenylate in the active site (49).

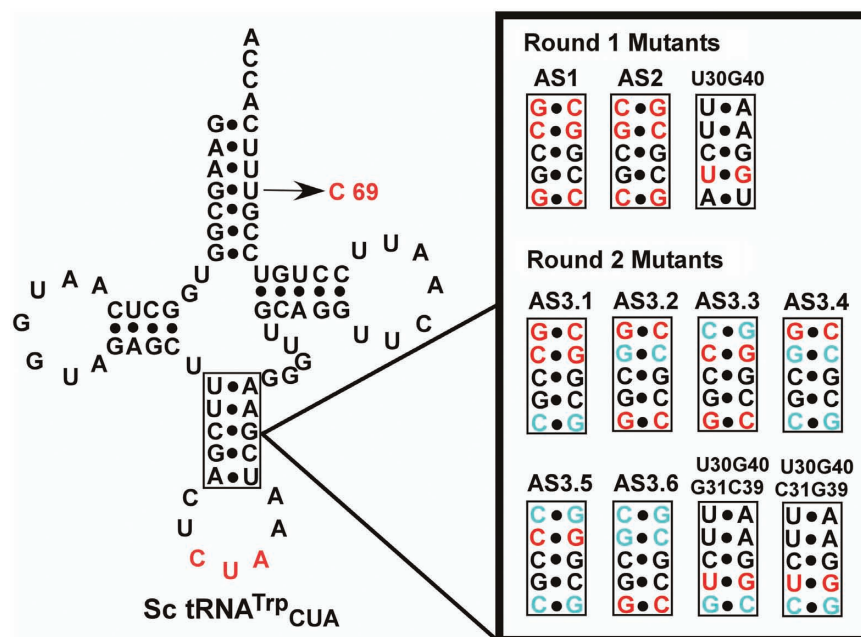
The hypothesis that the amber suppressor may be flexible enough to be charged by lysyl tRNA synthetase in turn suggests how the suppressor might be rationally engineered to avoid cross-reaction. The anticodon stem of both the *E. coli* lysine tRNA and the yeast suppressor tRNA share a relatively A-U rich (3/5 bp) stem sequence. Indeed, the Ec-tRNA<sup>Lys</sup> has only one G-C pair (G30-C40) in the anticodon stem, while most *E. coli* tRNA sequences contain three or more G-C pairs in this

stem. We therefore hypothesized that ‘stiffening’ the anticodon stem of the suppressor by making it more G-C rich should lead to a reduction in background aminoacylation by EcKRS. This hypothesis was further supported by the fact that most of the known orthogonal suppressors in *E. coli* have relatively G-C rich anticodon stems (Supplementary Figure S3) and that Fukunaga *et al.* (43) have reported that the misacylation of the Sc-tRNA<sup>Tyr</sup> amber suppressor by the EcKRS can also be reduced by G-C enrichment of the anticodon stem.

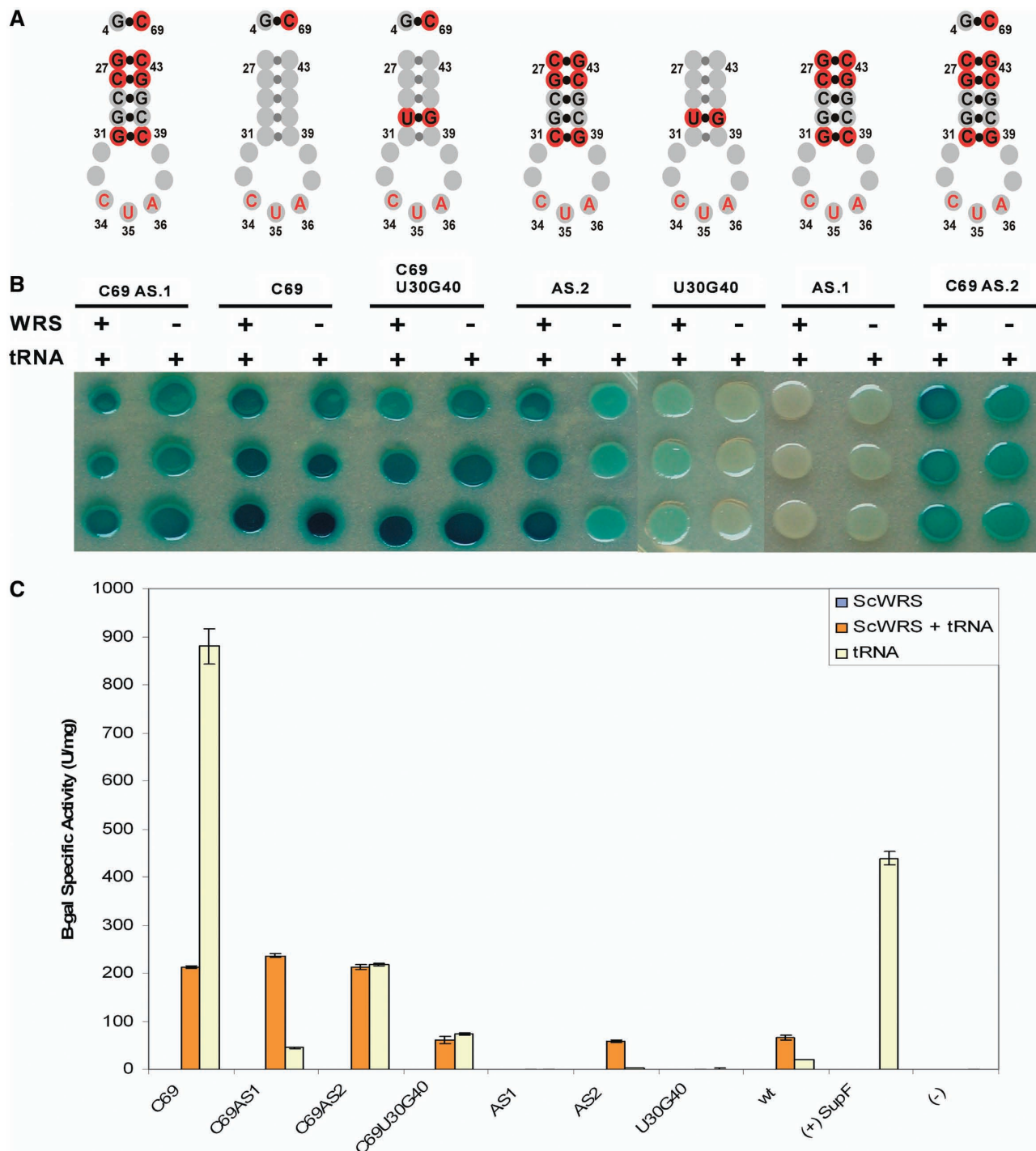
Other substitutions have also been shown to reduce background charging by EcKRS. A U30-G40 wobble pair has previously been shown to eliminate the cross reactivity of the EcKRS with the yeast tRNA<sup>Ile</sup> amber suppressor (50). This same negative discriminator was also inserted into the anticodon stem of the Sc-tRNA<sup>Phe</sup> amber suppressor which led to a reduction in the background misacylation caused by the EcKRS (7).

Based on this analysis, parallel paths were devised to eliminate the EcKRS-catalyzed background misacylation of Sc-tRNA<sup>Trp</sup>. First, we constructed G-C rich anticodon stem variants of Sc-tRNA<sup>Trp</sup> (Figure 5, Round 1 Mutants). Secondly, we introduced the negative identity determinant U30-G40 wobble pair into the anticodon stem of Sc-tRNA<sup>Trp</sup>. Finally, a U69C substitution was introduced into the acceptor stem, since replacement of a similar wobble pair (U4-G69) in yeast tRNA<sup>Tyr</sup> had previously been shown to increase the suppression activity of this tRNA. Combinations of these changes were also generated.

These suppressor variants were assayed either alone or in combination with the U69C substitution in the acceptor stem (Figure 6). The suppressor tRNA variants were



**Figure 5.** Rational Sc-tRNA<sup>Trp</sup><sub>CUA</sub> mutants. Designed anticodon stem sequences are shown. The first round of mutations from the original suppressor tRNA are shown in red. The anticodon stem mutants were also made in combination with the U69C mutation. The second round anticodon stem mutations that differ from the first set of suppressor designs are shown in cyan. Mutations preserved from the first round are shown in red. Unchanged nucleotides from the original suppressor tRNA are in black.



**Figure 6.** Suppression activity of rational suppressor tRNA mutants. (A) Structures of tRNA mutants anticodon stems. Sequence changes from Sc-tRNA<sup>Trp</sup><sub>Amb</sub> are denoted by red circles. (B)  $\beta$ -galactosidase colony spot activity assay. Each mutant tRNA was expressed with (+) or without (-) the ScWRS to test for orthogonality. Each sample was taken from the same culture and spotted in triplicate in column format. (C) Quantitation of suppression activity with  $\beta$ -galactosidase. Error bars represent 1 SD ( $n = 3$ ).

cloned into the pBRIVTC3B or pRST.11B expression vectors and transformed into *E. coli* strain CSH108 with or without the cognate ScWRS. Specific suppression of the amber nonsense mutation in the  $\beta$ -galactosidase gene by strains containing the ScWRS/Sc-tRNA<sup>Trp</sup><sub>Amb</sub> pair was compared with background suppression from the strains only containing the suppressor tRNA variants (Figure 6).

The AS.1 variant shows no suppression activity with or without the ScWRS. The U30G40 substitution not only had greatly reduced activity with ScWRS but also showed no background suppression activity. The AS.2 variant shows roughly equivalent suppression activity to its parent, but had 13-fold reduced background suppression activity (Figure 6).

Interestingly, all of the variants that contained the U69C substitution in the acceptor stem showed increased background suppression relative to the original Sc-tRNA<sup>Trp</sup><sub>Amb</sub> suppressor. N-terminal sequencing of the DHFR protein produced from the DHFR\_V10amb construct co-expressed with this suppressor revealed that the U69C substitution changes the identity of the Sc-tRNA<sup>Trp</sup> suppressor from tryptophan to histidine (data not shown). Therefore, the G4-C69 base pair introduced by this substitution appears to be a heretofore unknown identity element for the *E. coli* histidine tRNA synthetase (EcHRS) and when combined with the known G1-C72 and A73 identity elements that are also fortuitously in Sc-tRNA<sup>Trp</sup> leads to efficient aminoacylation by EcHRS.

### Optimization of an orthogonal tRNA<sup>Trp</sup> amber suppressor

Since initial results indicated that ‘stiffening’ the anticodon stem yielded improved suppressor function, each of the A-U base pairs in the original tRNA anticodon stem was systematically replaced with G-C base pairs (Figure 5, Round 2 Mutants). In addition, since the U30G40 variant eliminated background activity this substitution was also combined with G-C base pairs (Figure 5, Round 2 Mutants). These variants were again screened in the presence and absence of ScWRS (Figure 7).

The G-C replacements showed a range of activities with and without the ScWRS. Variant AS3.1 was translationally inactive, whereas variants AS3.2, AS3.3 and AS3.6 were active yet still showed significant background suppression activity. However, variants AS3.4 and AS3.5 showed a marked reduction in background suppression activity while maintaining activity in the presence of ScWRS. AS3.4 reduced background activity by 28-fold over the original Sc-tRNA<sup>Trp</sup> suppressor, while AS3.5 showed a reduction of nearly 50-fold over the Sc-tRNA<sup>Trp</sup> suppressor. All of the variants containing U30G40 were either inactive or showed residual background activity.

### Verifying the orthogonality of suppressors AS3.4 and AS3.5

To verify the orthogonality of the two promising suppressor candidates (AS3.4, AS3.5) from the LacZ screen, these suppressors were co-expressed with and without the ScWRS in the presence of the DHFR\_V10Amber plasmid. Full-length DHFR with a C-terminal histidine tag was isolated by IMAC, yielding 270 mgs/l DHFR for AS3.4 or 260 mgs/l DHFR for AS3.5 [roughly two-thirds the yield from the construct lacking the amber (TAG) stop codon (pACYCSolo-DHFR-wt)].

However, as can be seen in Figure 8A, the AS3.5 suppressor still produces a faint DHFR band on a Coomassie-stained gel. Concentration of these protein samples cells that only contained the suppressor tRNA clearly reveals that DHFR is still being expressed in the presence of the AS3.5 suppressor but not the AS3.4 suppressor (93-fold concentration; Figure 8B). N-terminal peptide sequencing of the DHFR samples isolated from

the strains expressing the AS3.4 and the AS3.5 suppressors with the ScWRS revealed that both incorporate tryptophan in DHFR in response to the amber nonsense codon at position 10 (Figure 8C and D). However, N-terminal sequencing of the DHFR band from the AS3.5 suppressor-only sample revealed that lysine was still being misincorporated at a low level by the Sc-tRNA<sup>Trp</sup>-AS3.5 suppressor (Figure 8E).

### Pathways for the acquisition of suppressor activity

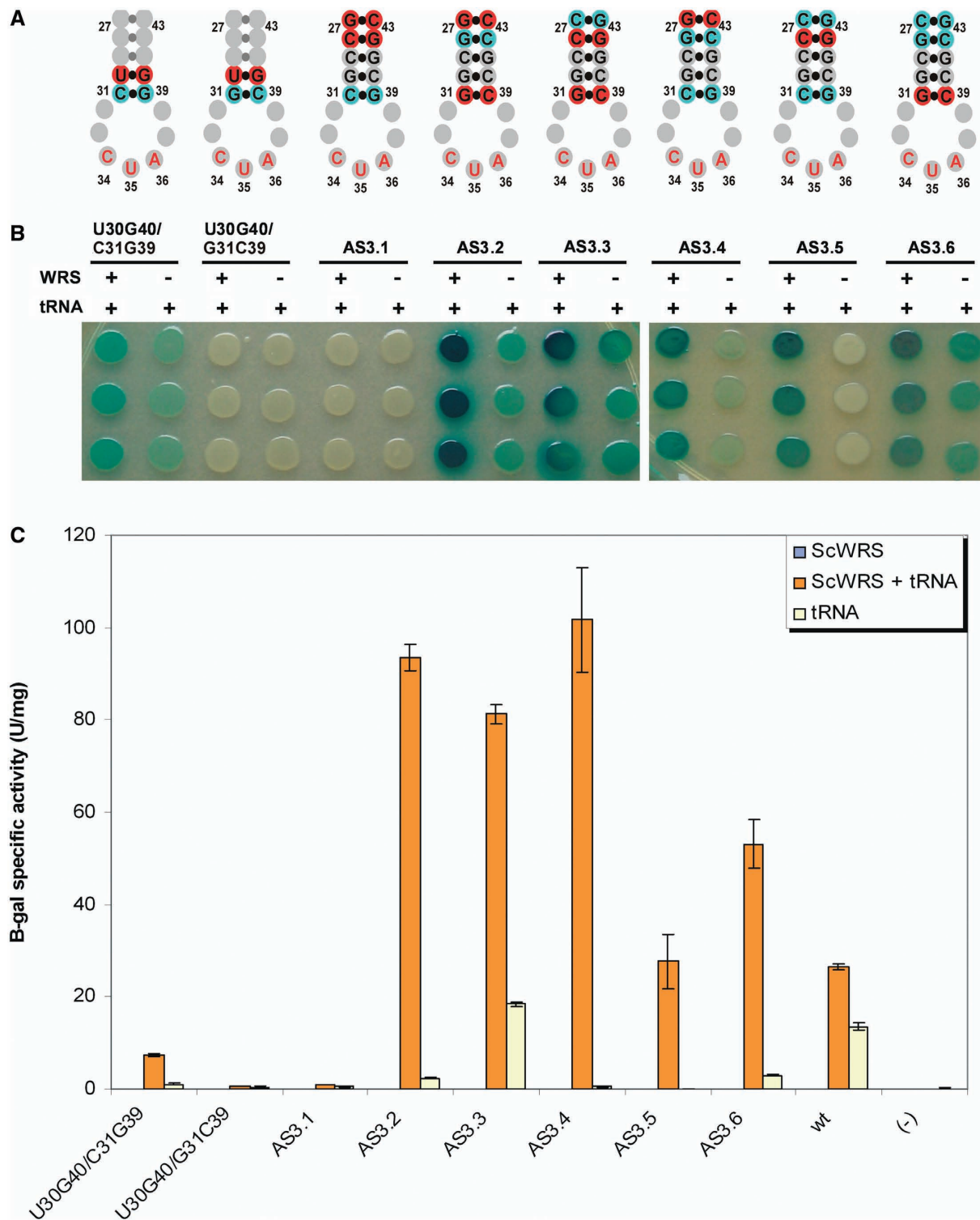
Several functional patterns emerge from these results. If the anticodon stem contains the G27-C43/C28-G42 pairs, the suppressors are non-functional (see variants AS.1 and AS3.1). If the anticodon stem contains the G31-C39 base pair, these tRNAs are functional but show varying levels of background suppression (see variants AS3.2, AS3.3 and AS3.6). If the base of the anticodon stem contains the C31-G39 base pair, the tRNA demonstrates reduced background suppression levels (see variants AS2, AS3.4 and AS3.5).

The contributions of individual base pairs towards charging and orthogonality can be better determined by comparing individual variants (Figure 9). Starting with the inactive AS1 variant, if the C28-G42 base pair is changed to G28-C42 base pair found in variant AS3.2, a significant restoration of suppression activity is restored in the presence of ScWRS ( $\Delta$  93.2 U/mg of  $\beta$ -galactosidase activity) but an increase in the background from AS1 is also seen ( $\Delta$  2.2 U/mg).

Continuing to compare substitutions, if the G31-C39 base pair at the base of the stem in variant AS3.2 is then changed to C31-G39 to yield variant AS3.4, we lose most of the background activity seen in AS3.2 ( $\Delta$  -1.9 U/mg) but gain a small amount of activity ( $\Delta$  8.1 U/mg) in the presence of ScWRS. A similar effect is seen if we go from variant AS1 to AS3.1 and then to AS3.4. If the G31-C39 base pair in AS1 is first changed to the C31-G39 base pair to yield variant AS3.1, only a modest gain in activity is seen ( $\Delta$  0.5 U/mg) but the suppressor remains largely inactive and produces a white colony in the spotting assay with or without the ScWRS. When the C28-G42 base pair in AS3.1 is finally changed to G28-C42 to yield variant AS3.4, there is a quite significant gain in suppression activity in the presence of the ScWRS ( $\Delta$  100.8 U/mg) but little in the way of background activity ( $\Delta$  0.1 U/mg). In other words, both mutational routes from AS1 to AS3.4 yield sequence intermediates whose activities can be readily rationalized.

We can also calculate the contribution of the G27-C43 base pair to suppression activity by comparing the activity of variant AS2 with variant AS3.4. In this case, the C27-G43 base pair at the top of the acceptor stem in AS2 is changed to a G27-C43 base pair found in AS3.4, and this more than doubles the activity of the suppressor with ScWRS ( $\Delta$  57.3 U/mg), while reducing the background at the same time ( $\Delta$  -0.4 U/mg).

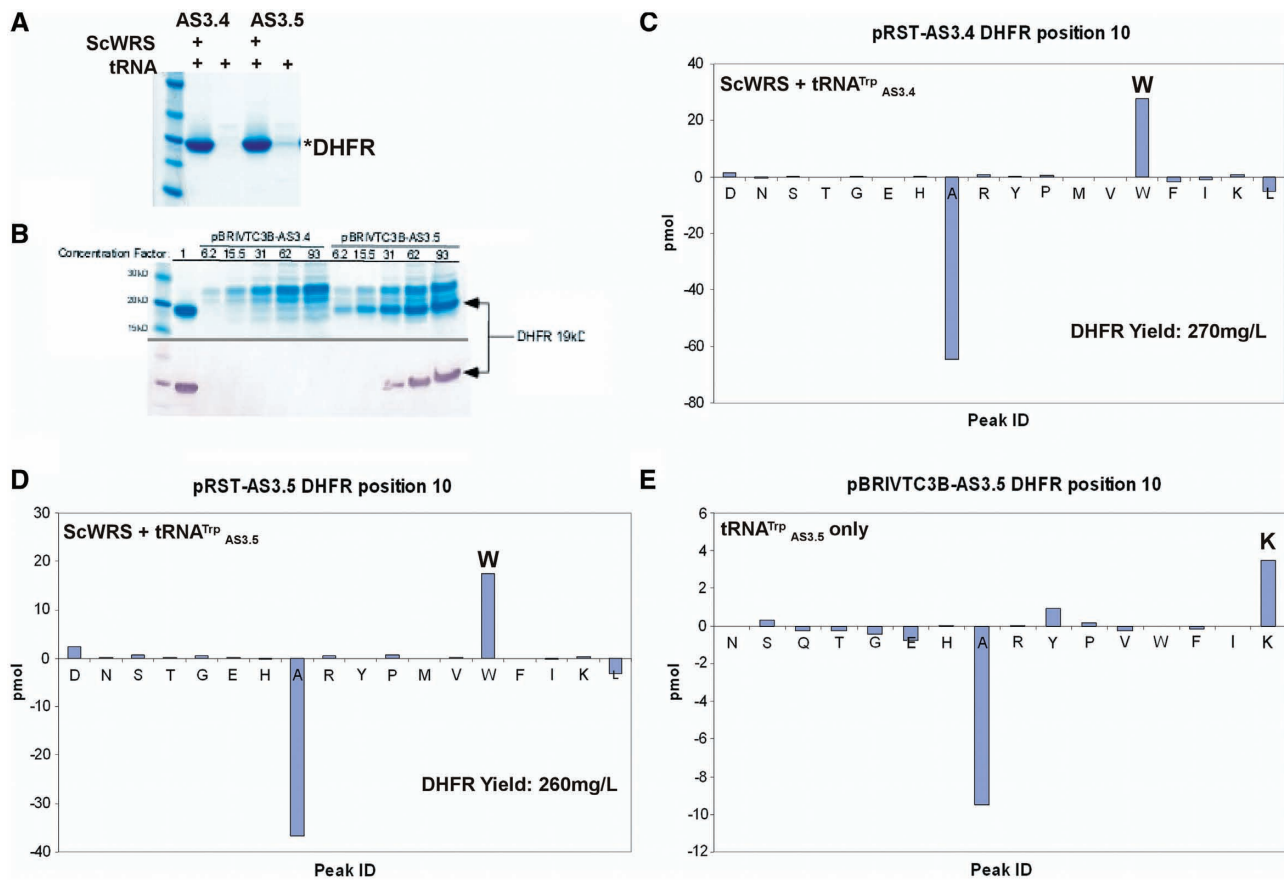
Overall from these results, it is apparent that the G28-C42 base pair is responsible for increasing the overall activity of the suppressor, but especially the



**Figure 7.** Suppression activity of optimized, rational suppressor tRNA mutants. (A) Mutant tRNA anticodon stem structures with mutations shown in red or cyan. (B)  $\beta$ -galactosidase colony spot activity assay. Each mutant tRNA was expressed with (+) or without (-) the ScWRS to test for orthogonality. Each sample was taken from the same culture and spotted in triplicate in column format. (C) Quantitation of suppression activity with  $\beta$ -galactosidase. Error bars represent 1 SD ( $n = 3$ ).

activity associated with the ScWRS. The G27-C43 base pair not only contributes to the overall activity of the suppressor in the presence of the ScWRS, but also contributes in part to reducing the background activity associated with the *E. coli* KRS. The C31-G39 base pair does

contribute greatly to the overall activity of the suppressor with the ScWRS, but is the main contributor to the reduction of background activity with the *E. coli* KRS. It is remarkable that these contributions are apparently modular, and not highly context-dependent.



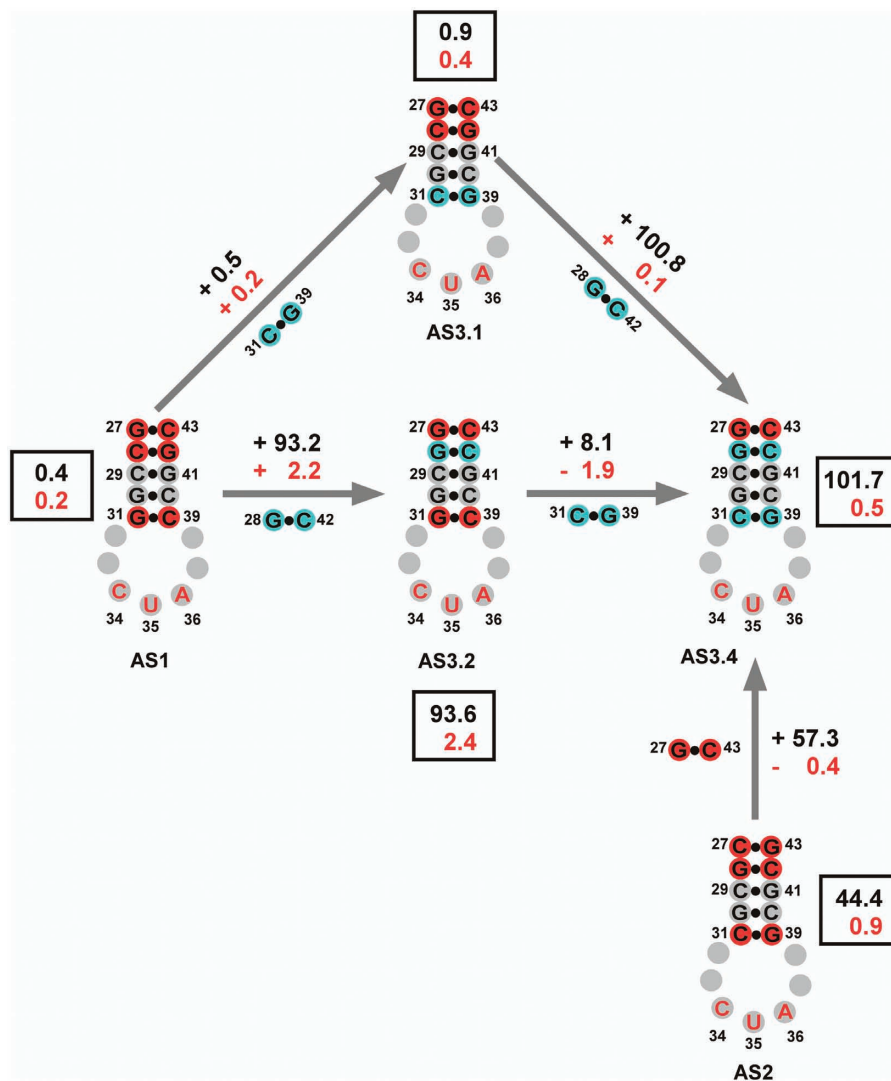
**Figure 8.** Verifying the orthogonality of AS3.4 and AS3.5. (A) Coomassie stained SDS-PAGE gel of IMAC purified DHFR expressed with and without ScWRS and with the AS3.4 or the AS3.5 suppressor reveals that low level background suppression (AS3.5 suppressor tRNA without ScWRS) still occurs with AS3.5. (B) Concentration of the suppressor tRNA-only (expressed from plasmid pBRIVTC3B) samples confirms the presence of the 18.7 kDa DHFR band in the AS3.5 fraction, but not the AS3.4 fraction. The lane next to the marker is wild-type DHFR run with no additional concentration. The elution fractions from IMAC columns for AS3.4 and AS3.5 expressed without ScWRS but with plasmid pACYCSOLO-DHFR\_V10Amb were concentrated and run on the gel at fractions representing concentration factors of 6.2-, 15.5-, 31-, 62- and 93-fold over the initially isolated protein concentration (Top panel). Western blot probing of the C-terminal HisTag on DHFR by a primary AP-conjugated anti-polyhistidine antibody confirms that DHFR is expressed at low levels in the AS3.5 sample but not in the AS3.4 sample (Bottom panel). N-terminal sequencing of the purified DHFR samples reveals that both incorporate tryptophan when the ScWRS is co-expressed with the suppressor tRNA (C and D). However, the AS3.5 suppressor is also misacylated by the *E. coli* KRS when expressed by itself (E).

## DISCUSSION

One of the requirements for developing an orthogonal aaRS-tRNA pair is that the tRNA works independently of the host aaRSs and tRNAs. This means that the tRNA must not interact with the endogenous aaRSs because this could result in the insertion of more than one amino acid at a given codon. The aaRS-tRNA recognition is governed by interactions between a given aaRS and identity elements on its cognate, the so-called ‘second genetic code’. These interactions are not only unique to each pair, but also typically exhibit species-specific differences, as well, which is one reason that heterologous pairs can often be engineered to function as orthogonal pairs (40). In the case of tryptophanyl tRNAs, the relevant identity elements are thought to be discriminator bases at position 73, either A73 in bacteria or G73 in eukaryotes (40,51).

In order to introduce multiple unnatural amino acids into proteins and whole organisms, it will be useful to have multiple orthogonal tRNA synthetase:tRNA pairs.

In particular, we have previously attempted to evolve *E. coli* to completely utilize the unnatural amino acid 4-fluorotryptophan (20). Insights from these experiments revealed that ‘top-down’ evolution of an organism’s genetic code is not a practical approach to augment the encoded amino acid content of proteins in a living cell (52). These efforts were only partially successful, owing in part to the fact that we did not attempt to pre-engineer the aaRS-tRNA constituents of the cell. Over the last decade, several groups have demonstrated the utility of using mutated or evolved aaRSs and their cognate tRNAs to expand the genetic code and add additional amino acids into proteins (2,53). These ‘bottom-up’ approaches to evolving the genetic code when combined with the cellular adaptation protocols used in the ‘top-down’ approaches may some day lead to artificial cells that have refracted genetic codes which simultaneously encode for multiple unnatural amino acids (52). In order to expand the amino acid diversity of a cell, multiple distinct aaRS-tRNA orthogonal pairs will be



**Figure 9.** Dissecting the contribution of mutations to suppressor activity. Anticodon representations of the designed suppressors are shown with the mutations made in the first and second round of designs shown in red and cyan, respectively. Grey spheres represent positions unchanged from the original Sc-tRNA<sup>Trp</sup><sub>Amb</sub> suppressor. The suppressor activities were calculated from the average values obtained from the  $\beta$ -galactosidase assay. The top numbers (black) represent the change in activity between any two variants in the presence of the ScWRS. The bottom number (red) represents the change in the background suppressor activity (tRNA only) between any two variants. The boxed values are the raw averaged suppressor activities obtained for each variant, with the top number representing the activity with the ScWRS and the bottom number the activity without the ScWRS.

required to accommodate the structural diversity multiple unnatural amino acids while maintaining the selectivity of the translation process. To date, there are relatively few orthogonal tRNA synthetase:tRNA pairs available for use in *E. coli*, therefore in order to generate an 'unColi' with an alternative genetic code and a proteome that is completely augmented with an unnatural amino acid, we have developed a new orthogonal tRNA synthetase:tRNA pair based on the *S. cerevisiae* WRS and tRNA.

We chose the WRS and its cognate tRNA as a starting point for constructing an orthogonal pair due to a number of features which make it particularly attractive for potentially adding unnatural amino acids into proteins. These features include the lack of an editing domain in the WRS and its capacious binding pocket for the largest natural

amino acid which has previously led to its use for incorporating tryptophan analogs into proteins (16,17). The expanded substrate flexibility of WRS has even supported adaptive evolution of whole bacteria and bacteria phage proteomes to tryptophan analogs (18–20). However, the engineering of the WRS (and its cognate tRNA) for the site-specific insertion of large, heterocyclic unnatural amino acids has not been sufficiently explored, especially when compared with commonly used tyrosyl tRNA synthetase (54). Zhang *et al.* (21) have used the WRS from *E. coli* and an opal (TGA) nonsense suppressor version of its cognate tRNA for the site-specific incorporation of 5-hydroxytryptophan into proteins in mammalian cells, but this orthogonal pair cannot be used in *E. coli* since it is of bacterial origin and has

overlapping recognition features which eliminates its orthogonality. Additionally, screening of synthetic suppressor tRNAs from *E. coli* has revealed that the tRNA<sup>Trp</sup> is the most efficient tRNA for incorporating large unnatural amino acids such as fluorophores into proteins in *E. coli in vitro* (55,56). Therefore, we sought to develop an orthogonal tryptophanyl pair for use in the *E. coli* genetic background.

We anticipated that the yeast tryptophanyl tRNA would be orthogonal to *E. coli* WRS based upon known differences relative to the *E. coli* tRNA<sup>Trp</sup> (22,23). The conversion of the CCA anticodon sequence to CUA to make an amber nonsense suppressor was also expected to further enhance the anti-discrimination of the *E. coli* WRS for the yeast tRNA<sup>Trp</sup><sub>Amb</sub>, as this mutation in the *E. coli* tRNA<sup>Trp</sup> changes its identity from tryptophan to glutamine (24). However, when the yeast tRNA<sup>Trp</sup><sub>Amb</sub> was initially expressed in *E. coli*, it was shown to be functional but not orthogonal to the *E. coli* translational system (Figure 2). Further analysis revealed that EcKRS was responsible for the misacylation of the original Sc-tRNA<sup>Trp</sup><sub>Amb</sub> suppressor in the absence of the ScWRS (Figure 3C). The misacylation of amber suppressor tRNAs by the EcKRS is a widely reported phenomenon, so much so that nearly a third of the amber suppressors derived from different *E. coli* tRNAs (Ile, Arg, Met (elongation), Asp, Val) all encode lysine *in vivo* (57). Other amber suppressors including those derived from the yeast tRNAs for phenylalanine (7), isoleucine (50) and tyrosine (43) have also been shown to be misacylated by the EcKRS *in vivo*.

The major identity elements of the Ec-tRNA<sup>Lys</sup> have been previously determined to be the anticodon nucleotides mnm<sup>5</sup>s<sup>2</sup>U34, U35, U36 and the discriminator nucleotide A73 (38,39). Since EcKRS can bind the mnm<sup>5</sup>s<sup>2</sup>U34 modified nucleotide or cytosine equally well this means that three of the four major identity elements are shared between the yeast tRNA<sup>Trp</sup><sub>Amb</sub> suppressor and Ec-tRNA<sup>Lys</sup> (Figure 4). Making mutations in the anticodon stems of amber suppressor tRNAs derived from the yeast phenylalanine, tyrosine and isoleucine, tRNAs has been previously reported to reduce the mischarging of the suppressor tRNAs by the EcKRS (7,43,50). Fukunaga *et al.* (43) reported that G-C enrichment of the anticodon stem of Sc-tRNA<sup>Tyr</sup><sub>Amb</sub> led to significant reduction in misacylation by EcKRS. Based on these findings, these authors hypothesized that the interaction between EcKRS and Ec-tRNA<sup>Lys</sup> requires a structural element in the anticodon stem which provides flexibility during the aaRS/tRNA recognition process (44). G-C enrichment of the anticodon stem presumably reduces its flexibility and thereby reduces its interaction with EcKRS. This hypothesis is anecdotally supported by comparison of the anti-codon stems of all of the known *E. coli* orthogonal tRNAs (Supplementary Figure S3), all of which have relatively G-C rich anticodon stems. In comparison, the tRNA<sup>Trp</sup><sub>Amb</sub> contains an A-U rich anticodon stem, much like the Ec-tRNA<sup>Lys</sup>.

Building on these findings, we made a series of anticodon stem mutations in Sc-tRNA<sup>Tyr</sup><sub>Amb</sub> aimed at disrupting the structural plasticity of this stem and thus the

aberrant interaction with EcKRS. In order to better parse out the individual effects of G-C substitutions, 16 different Sc-tRNA<sup>Tyr</sup><sub>Amb</sub> anticodon stem mutants were made and screened via a  $\beta$ -galactosidase suppression assay. Of these 16 mutants, one variant, AS3.4, demonstrated a completely orthogonal phenotype. This variant contained a highly G-C enriched anticodon stem (G28-C42, C31-G39 and G27-C43). The G28-C42 and G27-C43 base pairs at the top of the anticodon stem were found to enhance suppression activity in the presence of the ScWRS, while the C31-G39 base pair at the bottom of the anticodon stem led to the greatest reduction in background suppression by EcKRS (Figure 9). In agreement with these results, this latter mutation has also been shown to reduce the mischarging of the yeast tRNA<sup>Tyr</sup> by EcKRS (44).

Interestingly, the C31-G39 base pair in yeast tRNA<sup>Tyr</sup><sub>Amb</sub> also led to an increase in misacylation by *E. coli* glutamine tRNA synthetase. This result was similar to the results reported by Normanly *et al.* (57) two decades ago for amber suppressors derived from two isoacceptors for *E. coli* tRNA<sup>Ile</sup>. In this case, when tRNA<sup>Ile</sup><sub>1</sub> was made into an amber suppressor it was aminoacylated with glutamine, whereas the tRNA<sup>Ile</sup><sub>2</sub> amber suppressor was charged with lysine. One of the sequence differences between these two tRNA<sup>Ile</sup> isoacceptors was the 31–39 base pair which was C31-G39 in tRNA<sup>Ile</sup><sub>1</sub> and A31- $\Psi$ 39 for tRNA<sup>Ile</sup><sub>2</sub>, confirming the importance of the C31-G39 base pair for anti-discrimination by the EcKRS. These results suggest that the EcKRS could utilize a tRNA recognition and catalysis mechanism similar to the arginyl-tRNA synthetase which requires some flexibility in its cognate tRNA to undergo its induced-fit mode of catalysis (58). The reduction in the flexibility of the anticodon stem in mutant AS3.4 may have abolished recognition by the EcKRS.

Unfortunately, the previous attempts to rationally design an orthogonal suppressor tRNA based on the yeast tRNA<sup>Tyr</sup> were ultimately unsuccessful, even though the methodology was similar to that reported herein. This could be due to subtle structural differences between the engineered yeast tRNA<sup>Tyr</sup> anticodon stem relative to the engineered yeast tRNA<sup>Trp</sup> anticodon stem, as none of the mutants assayed were identical. That said, the most orthogonal tRNA<sup>Tyr</sup> suppressor that was reported had an anticodon stem sequence that was very similar to our AS3.5 tRNA<sup>Trp</sup> variant, differing at only a single base pair (G29-C41 for the tRNA<sup>Tyr</sup> variant versus C29-G41 for AS3.5).

It is especially interesting that the impact of individual G-C substitutions on both charging by the ScWRS and mischarging by the Ec-LysRS appeared to be modular and additive (Figure 9). This was not necessarily to have been expected, since many protein:RNA interactions require precise conformational fits in which any perturbation will significantly decrease affinity and activity. The apparent additivity of the interactions implies that it may be possible in the future to rationally engineer the flexibility of tRNAs to achieve novel specificities. More intriguingly, these results suggest how changes in the genetic code may have evolved. Changes in tRNA

anticodons are thought to yield quantized changes in the genetic code, with new amino acids being lost or acquired at particular codons. However, additional mutations in tRNAs that lead to ambiguous charging specificities could modulate the adoption of a new code, moving through an intermediate state in which more than one amino acid was proportionately introduced across from a given codon (59–61). The fact that changes in stem flexibility can gradually and additively lead to alterations in this ambiguity (in ‘mischarging’) implies that evolutionary routes to new codes could lead gradually through ambiguously recoded proteomes.

Previously reported refinements of orthogonal suppressor tRNAs were largely done by directed evolution from a library of mutations that tended to focus any mutations made from a proposed suppressor in the acceptor stem (4), or the loop regions of the tRNA sequence (62) to screen for mutations that enhanced the orthogonal character of a mutant tRNA. These regions of the tRNA are logical starting points for incorporating additional discrimination elements into tRNAs as these regions are commonly accessed by aaRSs. However, in light of the results presented herein, modification of the structural features of tRNAs should also be considered when designing/evolving aaRS interactions. Due to tRNA’s central role in the translation system, tRNA molecules have to interact with multiple proteins (aaRSs, EFTu) and the ribosome itself. The evolved flexibility of the tRNA molecule enables it to interact with multiple binding partners making this feature of tRNAs one of the most essential for translation (63). While interactions between tRNA identity elements and aaRSs (64), and between the tRNA acceptor stem and EFTu (65) have been extensively explored, the role of structural flexibility in fine-tuning these interactions is still largely unknown. Recent and future molecular dynamic simulations of tRNAs during translation will hopefully help to illuminate this largely overlooked property of tRNAs (63). In this regard, mutant tRNAs such as those reported herein could be useful for studying the structural dynamics of tRNA interactions between cognate and non-cognate aaRSs, much as the Hirsh suppressor tRNA<sup>Trp</sup> (66) was useful for decoding the dynamics of codon–anticodon interactions on the ribosome (67).

The Sc-tRNA<sup>Trp</sup>-AS3.4 and ScWRS orthogonal pair reported herein is to our knowledge the only reported tryptophanyl-based orthogonal pair available for use in prokaryotes and should be of great use in expanding efforts to evolve proteins that contain unnatural amino acids. In addition to potentially being used for changing the genetic code of a cell, this orthogonal suppressor system could also be used for recoding *in vitro*. For example, recombinant translation systems [similar to the PURE system (68)] have been used by Forster and colleagues (69,70) to introduce chemically acylated tRNAs containing unnatural amino acids into peptides. Similarly, Szostak and colleagues (71,72) have made peptides containing multiple unnatural amino acid analogs by taking advantage of the natural substrate flexibility of aaRSs. In either instance, the new orthogonal pair could be introduced in place of tRNA synthetase and its

cognate tRNA, or in addition to the standard complement of synthetases and tRNAs in order to more efficiently generate proteins with altered compositions.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

Funding for open access charge: National Science Foundation (grant number MCB-0943383); National Institutes of Health/National Institute of General Medical Sciences (grant number R01GM084703); The Welch Foundation (grant number F-1654).

*Conflict of interest statement.* None declared.

## REFERENCES

- Kiga,D., Sakamoto,K., Kodama,K., Kigawa,T., Matsuda,T., Yabuki,T., Shirouzu,M., Harada,Y., Nakayama,H., Takio,K. *et al.* (2002) An engineered Escherichia coli tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system. *Proc. Natl Acad. Sci. USA*, **99**, 9715–9720.
- Xie,J. and Schultz,P.G. (2006) A chemical toolkit for proteins—an expanded genetic code. *Nat. Rev. Mol. Cell Biol.*, **7**, 775–782.
- Wang,L., Brock,A., Herberich,B. and Schultz,P.G. (2001) Expanding the genetic code of Escherichia coli. *Science*, **292**, 498–500.
- Anderson,J.C., Wu,N., Santoro,S.W., Lakshman,V., King,D.S. and Schultz,P.G. (2004) An expanded genetic code with a functional quadruplet. *Proc. Natl Acad. Sci. USA*, **101**, 7566–7571.
- Pastrnak,M., Mageliery,T.M. and Schultz,P.G. (2000) A new orthogonal suppressor tRNA/Aminoacyl-tRNA synthetase pair for evolving an organism with an expanded genetic code. *Helv. Chim. Acta*, **83**, 2277–2286.
- Liu,D.R. and Schultz,P.G. (1999) Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl Acad. Sci. USA*, **96**, 4780–4785.
- Kwon,I., Wang,P. and Tirrell,D.A. (2006) Design of a bacterial host for site-specific incorporation of p-bromophenylalanine into recombinant proteins. *J. Am. Chem. Soc.*, **128**, 11778–11783.
- Santoro,S.W., Anderson,J.C., Lakshman,V. and Schultz,P.G. (2003) An archaeobacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in Escherichia coli. *Nucleic Acids Res.*, **31**, 6700–6709.
- Anderson,J.C. and Schultz,P.G. (2003) Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. *Biochemistry*, **42**, 9598–9608.
- Nguyen,D.P., Garcia Alai,M.M., Kapadnis,P.B., Neumann,H. and Chin,J.W. (2009) Genetically encoding N(epsilon)-methyl-L-lysine in recombinant histones. *J. Am. Chem. Soc.*, **131**, 14194–14195.
- Landes,C., Perona,J.J., Brunie,S., Rould,M.A., Zelwer,C., Steitz,T.A. and Risler,J.L. (1995) A structure-based multiple sequence alignment of all class I aminoacyl-tRNA synthetases. *Biochimie*, **77**, 194–203.
- Yang,X.L., Otero,F.J., Skene,R.J., McRee,D.E., Schimmel,P. and Ribas de Pouplana,L. (2003) Crystal structures that suggest late development of genetic code components for differentiating aromatic side chains. *Proc. Natl Acad. Sci. USA*, **100**, 15376–15380.
- Praetorius-Ibba,M., Stange-Thomann,N., Kitabatake,M., Ali,K., Soll,I., Carter,C.W. Jr, Ibba,M. and Soll,D. (2000) Ancient



- adaptation of the active site of tryptophanyl-tRNA synthetase for tryptophan binding. *Biochemistry*, **39**, 13136–13143.
14. Ribas de Pouplana, L., Frugier, M., Quinn, C.L. and Schimmel, P. (1996) Evidence that two present-day components needed for the genetic code appeared after nucleated cells separated from eubacteria. *Proc. Natl Acad. Sci. USA*, **93**, 166–170.
  15. Double, S., Bricogne, G., Gilmore, C. and Carter, C.W. Jr (1995) Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase. *Structure*, **3**, 17–31.
  16. Budisa, N. (2006) *Engineering the Genetic Code*. Wiley-VCH, Weinheim.
  17. Budisa, N., Rubini, M., Bae, J.H., Weyher, E., Wenger, W., Golbik, R., Huber, R. and Moroder, L. (2002) Global replacement of tryptophan with aminotryptophans generates non-invasive protein-based optical pH sensors. *Angew. Chem. Int. Ed. Engl.*, **41**, 4066–4069.
  18. Wong, J.T. (1983) Membership mutation of the genetic code: loss of fitness by tryptophan. *Proc. Natl Acad. Sci. USA*, **80**, 6303–6306.
  19. Bacher, J.M., Bull, J.J. and Ellington, A.D. (2003) Evolution of phage with chemically ambiguous proteomes. *BMC Evol. Biol.*, **3**, 24.
  20. Bacher, J.M. and Ellington, A.D. (2001) Selection and characterization of *Escherichia coli* variants capable of growth on an otherwise toxic tryptophan analogue. *J. Bacteriol.*, **183**, 5414–5425.
  21. Zhang, Z., Alfonta, L., Tian, F., Bursulaya, B., Uryu, S., King, D.S. and Schultz, P.G. (2004) Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc. Natl Acad. Sci. USA*, **101**, 8882–8887.
  22. Himeno, H., Hasegawa, T., Asahara, H., Tamura, K. and Shimizu, M. (1991) Identity determinants of *E. coli* tryptophan tRNA. *Nucleic Acids Res.*, **19**, 6379–6382.
  23. Yesland, K.D. and Johnson, J.D. (1993) Anticodon bases C34 and C35 are major, positive, identity elements in *Saccharomyces cerevisiae* tRNA(Trp). *Nucleic Acids Res.*, **21**, 5079–5084.
  24. Yaniv, M., Folk, W.R., Berg, P. and Soll, L. (1974) A single mutational modification of a tryptophan-specific transfer RNA permits aminoacylation by glutamine and translation of the codon UAG. *J. Mol. Biol.*, **86**, 245–260.
  25. Yesland, K.D., Nelson, A.W., Six Feathers, D.M. and Johnson, J.D. (1993) Identity of *Saccharomyces cerevisiae* tRNA(Trp) is not changed by an anticodon mutation that creates an amber suppressor. *J. Biol. Chem.*, **268**, 217–220.
  26. Atkin, A.L., Riazi, M.A., Greer, C.L., Roy, K.L. and Bell, J.B. (1993) The functional analysis of nonsense suppressors derived from in vitro engineered *Saccharomyces cerevisiae* tRNA(Trp) genes. *Gene*, **134**, 57–65.
  27. John, T.R., Ghosh, M. and Johnson, J.D. (1997) Identification and expression of the *Saccharomyces cerevisiae* cytoplasmic tryptophanyl-tRNA synthetase gene. *Yeast*, **13**, 37–41.
  28. de Boer, H.A., Comstock, L.J. and Vasser, M. (1983) The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc. Natl Acad. Sci. USA*, **80**, 21–25.
  29. Balbas, P., Soberon, X., Merino, E., Zurita, M., Lomeli, H., Valle, F., Flores, N. and Bolivar, F. (1986) Plasmid vector pBR322 and its special-purpose derivatives—a review. *Gene*, **50**, 3–40.
  30. Duester, G., Campen, R.K. and Holmes, W.M. (1981) Nucleotide sequence of an *Escherichia coli* tRNA (Leu 1) operon and identification of the transcription promoter signal. *Nucleic Acids Res.*, **9**, 2121–2139.
  31. Bauer, B.F., Kar, E.G., Elford, R.M. and Holmes, W.M. (1988) Sequence determinants for promoter strength in the leuV operon of *Escherichia coli*. *Gene*, **63**, 123–134.
  32. Ross, W., Salomon, J., Holmes, W.M. and Gourse, R.L. (1999) Activation of *Escherichia coli* leuV transcription by FIS. *J. Bacteriol.*, **181**, 3864–3868.
  33. Hsu, L.M., Klee, H.J., Zagorski, J. and Fournier, M.J. (1984) Structure of an *Escherichia coli* tRNA operon containing linked genes for arginine, histidine, leucine, and proline tRNAs. *J. Bacteriol.*, **158**, 934–942.
  34. Smith, M.C., Furman, T.C., Ingolia, T.D. and Pidgeon, C. (1988) Chelating peptide-immobilized metal ion affinity chromatography. A new concept in affinity chromatography for recombinant proteins. *J. Biol. Chem.*, **263**, 7211–7215.
  35. Stemmer, W.P., Cramer, A., Ha, K.D., Brennan, T.M. and Heyneker, H.L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*, **164**, 49–53.
  36. Hoover, D.M. and Lubkowski, J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.*, **30**, e43.
  37. Reim, D.F. and Speicher, D.W. (2001) N-terminal sequence analysis of proteins and peptides. *Curr. Protoc. Protein Sci.*, **Chapter 11**, Unit 11 10.
  38. McClain, W.H., Foss, K., Jenkins, R.A. and Schneider, J. (1990) Nucleotides that determine *Escherichia coli* tRNA(Arg) and tRNA(Lys) acceptor identities revealed by analyses of mutant opal and amber suppressor tRNAs. *Proc. Natl Acad. Sci. USA*, **87**, 9260–9264.
  39. Tamura, K., Himeno, H., Asahara, H., Hasegawa, T. and Shimizu, M. (1992) In vitro study of *E. coli* tRNA(Arg) and tRNA(Lys) identity elements. *Nucleic Acids Res.*, **20**, 2335–2339.
  40. Xu, F., Chen, X., Xin, L., Chen, L., Jin, Y. and Wang, D. (2001) Species-specific differences in the operational RNA code for aminoacylation of tRNA(Trp). *Nucleic Acids Res.*, **29**, 4125–4133.
  41. Cusack, S., Yaremchuk, A. and Tuskalo, M. (1996) The crystal structures of *T. thermophilus* lysyl-tRNA synthetase complexed with *E. coli* tRNA(Lys) and a *T. thermophilus* tRNA(Lys) transcript: anticodon recognition and conformational changes upon binding of a lysyl-adenylate analogue. *EMBO J.*, **15**, 6321–6334.
  42. Furter, R. (1998) Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in *Escherichia coli*. *Protein Sci.*, **7**, 419–426.
  43. Fukunaga, J., Yokogawa, T., Ohno, S. and Nishikawa, K. (2006) Misacylation of yeast amber suppressor tRNA(Tyr) by *E. coli* lysyl-tRNA synthetase and its effective repression by genetic engineering of the tRNA sequence. *J. Biochem.*, **139**, 689–696.
  44. Fukunaga, J., Ohno, S., Nishikawa, K. and Yokogawa, T. (2006) A base pair at the bottom of the anticodon stem is reciprocally preferred for discrimination of cognate tRNAs by *Escherichia coli* lysyl- and glutamyl-tRNA synthetases. *Nucleic Acids Res.*, **34**, 3181–3188.
  45. Brevet, A., Chen, J., Commans, S., Lazennec, C., Blanquet, S. and Plateau, P. (2003) Anticodon recognition in evolution: switching tRNA specificity of an aminoacyl-tRNA synthetase by site-directed peptide transplantation. *J. Biol. Chem.*, **278**, 30927–30935.
  46. Commans, S., Lazard, M., Delort, F., Blanquet, S. and Plateau, P. (1998) tRNA anticodon recognition and specification within subclass IIb aminoacyl-tRNA synthetases. *J. Mol. Biol.*, **278**, 801–813.
  47. Commans, S., Plateau, P., Blanquet, S. and Dardel, F. (1995) Solution structure of the anticodon-binding domain of *Escherichia coli* lysyl-tRNA synthetase and studies of its interaction with tRNA(Lys). *J. Mol. Biol.*, **253**, 100–113.
  48. Onesti, S., Desogus, G., Brevet, A., Chen, J., Plateau, P., Blanquet, S. and Brick, P. (2000) Structural studies of lysyl-tRNA synthetase: conformational changes induced by substrate binding. *Biochemistry*, **39**, 12853–12861.
  49. Saruwatari, Y., Wada, T., Takita, T. and Inouye, K. (2008) Substrate-induced conformational changes of the truncated catalytic domain of *Geobacillus stearothermophilus* lysyl-tRNA synthetase as examined by fluorescence. *Biochim. Biophys. Acta*, **1784**, 1633–1640.
  50. Buttcher, V., Senger, B., Schumacher, S., Reinbolt, J. and Fasiolo, F. (1994) Modulation of the suppression efficiency and amino acid identity of an artificial yeast amber isoleucine transfer RNA in *Escherichia coli* by a G-U pair in the anticodon stem. *Biochem. Biophys. Res. Commun.*, **200**, 370–377.
  51. Guo, Q., Gong, Q., Tong, K.L., Vestergaard, B., Costa, A., Desgres, J., Wong, M., Grosjean, H., Zhu, G., Wong, J.T. et al. (2002) Recognition by tryptophanyl-tRNA synthetases of discriminator base on tRNA<sup>Trp</sup> from three biological domains. *J. Biol. Chem.*, **277**, 14343–14349.

52. Bacher, J.M., Hughes, R.A., Tze-Fei Wong, J. and Ellington, A.D. (2004) Evolving new genetic codes. *Trends Ecol. Evol.*, **19**, 69–75.
53. Wang, Q., Parrish, A.R. and Wang, L. (2009) Expanding the genetic code for biological studies. *Chem. Biol.*, **16**, 323–336.
54. Xie, J. and Schultz, P.G. (2005) An expanding genetic code. *Methods*, **36**, 227–238.
55. Taira, H., Matsushita, Y., Kojima, K. and Hohsaka, T. (2006) Development of amber suppressor tRNAs appropriate for incorporation of nonnatural amino acids. *Nucleic Acids Symp. Ser.*, 233–234.
56. Taira, H., Matsushita, Y., Kojima, K., Shiraga, K. and Hohsaka, T. (2008) Comprehensive screening of amber suppressor tRNAs suitable for incorporation of non-natural amino acids in a cell-free translation system. *Biochem. Biophys. Res. Commun.*, **374**, 304–308.
57. Normanly, J., Kleina, L.G., Masson, J.M., Abelson, J. and Miller, J.H. (1990) Construction of *Escherichia coli* amber suppressor tRNA genes. III. Determination of tRNA specificity. *J. Mol. Biol.*, **213**, 719–726.
58. Guigou, L. and Mirande, M. (2005) Determinants in tRNA for activation of arginyl-tRNA synthetase: evidence that tRNA flexibility is required for the induced-fit mechanism. *Biochemistry*, **44**, 16540–16548.
59. Schultz, D.W. and Yarus, M. (1994) Transfer RNA mutation and the malleability of the genetic code. *J. Mol. Biol.*, **235**, 1377–1380.
60. Schultz, D.W. and Yarus, M. (1996) On malleability in the genetic code. *J. Mol. Evol.*, **42**, 597–601.
61. Yarus, M. and Schultz, D.W. (1997) Further comments on codon reassignment. Response. *J. Mol. Evol.*, **45**, 3–6.
62. Wang, L. and Schultz, P.G. (2001) A general approach for the generation of orthogonal tRNAs. *Chem. Biol.*, **8**, 883–890.
63. Alexander, R.W., Eargle, J. and Luthey-Schulten, Z. (2010) Experimental and computational determination of tRNA dynamics. *FEBS Lett.*, **584**, 376–386.
64. Giege, R., Sissler, M. and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.*, **26**, 5017–5035.
65. Sanderson, L.E. and Uhlenbeck, O.C. (2007) Exploring the specificity of bacterial elongation factor Tu for different tRNAs. *Biochemistry*, **46**, 6194–6200.
66. Hirsh, D. (1971) Tryptophan transfer RNA as the UGA suppressor. *J. Mol. Biol.*, **58**, 439–458.
67. Cochella, L. and Green, R. (2005) An active role for tRNA in decoding beyond codon: anticodon pairing. *Science*, **308**, 1178–1180.
68. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K. and Ueda, T. (2001) Cell-free translation reconstituted with purified components. *Nat. Biotechnol.*, **19**, 751–755.
69. Forster, A.C., Tan, Z., Nalam, M.N., Lin, H., Qu, H., Cornish, V.W. and Blacklow, S.C. (2003) Programming peptidomimetic syntheses by translating genetic codes designed de novo. *Proc. Natl Acad. Sci. USA*, **100**, 6353–6357.
70. Forster, A.C., Cornish, V.W. and Blacklow, S.C. (2004) Pure translation display. *Anal Biochem.*, **333**, 358–364.
71. Josephson, K., Hartman, M.C. and Szostak, J.W. (2005) Ribosomal synthesis of unnatural peptides. *J. Am. Chem. Soc.*, **127**, 11727–11735.
72. Hartman, M.C., Josephson, K., Lin, C.W. and Szostak, J.W. (2007) An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. *PLoS ONE*, **2**, e972.