

# Modification of orthogonal tRNAs: unexpected consequences for sense codon reassignment

Wil Biddle<sup>1</sup>, Margaret A. Schmitt<sup>2</sup> and John D. Fisk<sup>1,2,3,\*</sup>

<sup>1</sup>Department of Chemistry, Colorado State University, Fort Collins, CO 80523, USA, <sup>2</sup>Department of Chemical and Biological Engineering, Colorado State University, Fort Collins, CO 80523, USA and <sup>3</sup>School of Biomedical Engineering, Colorado State University, Fort Collins, CO 80523, USA

Received September 06, 2016; Revised October 05, 2016; Editorial Decision October 07, 2016; Accepted October 10, 2016

## ABSTRACT

**Breaking the degeneracy of the genetic code via sense codon reassignment has emerged as a way to incorporate multiple copies of multiple non-canonical amino acids into a protein of interest. Here, we report the modification of a normally orthogonal tRNA by a host enzyme and show that this adventitious modification has a direct impact on the activity of the orthogonal tRNA in translation. We observed nearly equal decoding of both histidine codons, CAU and CAC, by an engineered orthogonal *M. jannaschii* tRNA with an AUG anticodon: tRNA<sup>Opt</sup>. We suspected a modification of the tRNA<sup>Opt</sup><sub>AUG</sub> anticodon was responsible for the anomalous lack of codon discrimination and demonstrate that adenosine 34 of tRNA<sup>Opt</sup><sub>AUG</sub> is converted to inosine. We identified tRNA<sup>Opt</sup><sub>AUG</sub> anticodon loop variants that increase reassignment of the histidine CAU codon, decrease incorporation in response to the histidine CAC codon, and improve cell health and growth profiles. Recognizing tRNA modification as both a potential pitfall and avenue of directed alteration will be important as the field of genetic code engineering continues to infiltrate the genetic codes of diverse organisms.**

## INTRODUCTION

Genetic code engineering, expanding the set of non-canonical amino acids (ncAAs) that may be biosynthetically incorporated into proteins, is of increasing interest as a means of precisely endowing proteins with chemical functionalities that are not present in the natural set of 20 amino acids (1–3). The genetic code is often thought of as the straightforward pairing of mRNA codon with tRNA anticodon that specifies the amino acid sequence of a protein. A broader view of the genetic code includes the complete set of interactions that are involved in the conversion of in-

formation contained in the genome of an organism into the functional protein molecules that ultimately produce phenotypic behaviors.

The genetic code operates on at least three distinct levels where different sets of interactions affect the information transfer that is described in the code table. The pairing of mRNA codon with tRNA anticodon represents the primary genetic code. The idiosyncratic set of interactions between the complement of tRNA species and the suite of aminoacyl tRNA synthetase (aaRS) enzymes in an organism represents a second level of the genetic code (4,5). The combined interactions between tRNA species and aaRSs specify that the appropriate amino acid is attached to the appropriate tRNA molecule. The set of protein–tRNA interactions that determine the extent and positioning of modified nucleobases in tRNA molecules represents a third level of the genetic code. tRNA modification influences each level of the genetic code, including the interactions of mRNA codons with tRNA anticodons and the interactions of tRNAs with their aaRSs (6). In addition, modified tRNAs serve regulatory roles modulating the interactions between the process of translation and cellular stress and metabolism (7,8). The three levels of the genetic code are tightly connected: the pairing of tRNA anticodon to mRNA codon as a means of information transfer is limited by the efficiency of the joining of an amino acid to the appropriate tRNA, and the effectiveness of both processes are modulated by the modification of tRNA species.

Over 100 different modified nucleobases have been found in the various forms of RNA (9). The deamination of adenosine to inosine is the most common type of directed modification seen in RNA and is catalyzed by two classes of adenosine deaminases acting on RNA or tRNA (10,11). A to I editing of mRNA is pervasive in eukaryotes and occurs at over 100 million sites in the majority of human genes (12). The overwhelming majority of A to I changes in mRNA occur in untranslated regions, and their precise functions, presumably regulatory, have not yet been discerned. A to I editing is essential for development of the brain and is implicated in increasing the diversity of neuronal transporters and ion channels (13).

\*To whom correspondence should be addressed. Tel: +1 970 491 4115; Fax: +1 970 491 7369; Email: nickfisk@colostate.edu

Modification of tRNA is widespread and essential for the proper functioning of the translational apparatus in all kingdoms of life. In *Escherichia coli* (*E. coli*), one of the three model systems for which the full extent of modifications has been mapped, the 46 tRNA species contain an average of 7.5 modified bases, ~10% of the entire tRNA molecule (6). All *E. coli* tRNAs contain the modified bases ribothymidine and pseudouridine in the T-Psi-C loop, and 90% of *E. coli* tRNAs contain at least one dihydrouridine in the D stem loop (Figure 1). Although the most prevalent modifications occur in the T-Psi-C and D stem loops, the greatest diversity of modifications occurs in the anticodon loop. *E. coli* anticodon loops, tRNA positions 32–38, contain 21 different nucleobase modifications. In the complement of *E. coli* tRNAs, position 34, which recognizes the wobble position of a codon, may be modified to 1 of 14 non-AUGC bases to allow a single tRNA to read multiple codons. A single *E. coli* tRNA, tRNA<sup>Arg2</sup>, includes inosine modification of position 34, which allows one tRNA to decode three arginine codons: CGU, CGC and CGA. Position 37, directly 3' of the anticodon, may be modified to 1 of 7 non-AUGC bases (14–16). The large diversity of modifications that occur in the anticodon loop often have direct effects on the fidelity of translation, particularly modulating the reading of codons at the wobble position and helping to maintain the reading frame.

The introduction of an orthogonal tRNA/aminoacyl tRNA synthetase pair into a host organism enables the mRNA-directed incorporation of a non-canonical amino acid (17,18). An orthogonal aaRS does not recognize the natural complement of the host organism's tRNAs, and an orthogonal tRNA is not recognized by the host organism's suite of aminoacyl tRNA synthetase enzymes. In *in vivo* translational systems, the orthogonal tRNA interacts with the proteins that constitute the host's translational system (e.g. endogenous aaRSs, elongation factors and tRNA modifying enzymes). To date, the majority of work expanding the genetic code has focused on re-engineering existing orthogonal tRNA/aaRS pairs to activate additional non-canonical amino acids and improving levels of ncAA incorporation using these orthogonal pairs (19,20). These efforts target advancing systems with expanded genetic codes at the first and second levels of the genetic code: improving interactions between the orthogonal tRNA species and the mRNA codon, typically the amber stop, and enhancing interactions between the orthogonal tRNA and its cognate aaRS in order to increase the concentration of aminoacylated tRNA available to decode an mRNA codon.

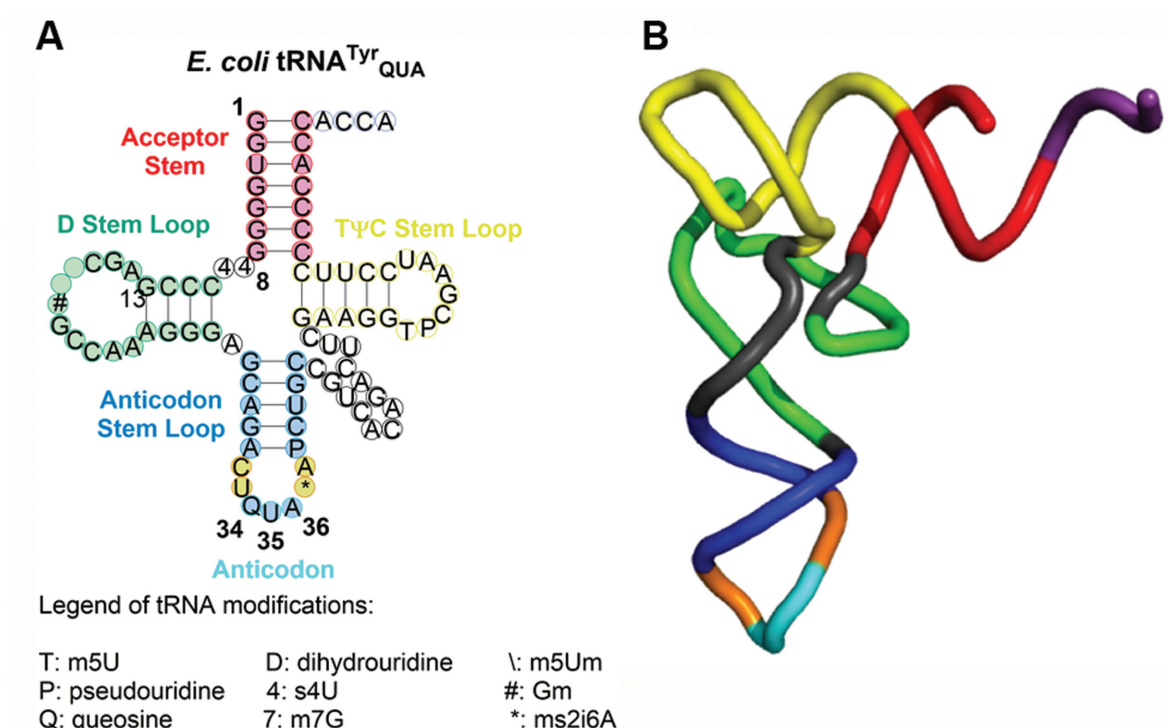
The number of codons that may be utilized for ncAA incorporation presently is limited. Most commonly, ncAAs are incorporated in response to the UAG amber stop codon. The amber stop codon is employed principally because it is the least commonly used codon in the *E. coli* genome. The introduction of an ncAA in response to a stop codon has the benefit of truncating any protein that fails to incorporate the ncAA, leading to high purity of the modified protein. The disadvantage of suppressing a stop codon is that competition with the normal termination functions leads to exponentially reduced protein yields when attempting to incorporate more than one copy of an ncAA. In general, the suppression of stop codons limits ncAA incorporation to a

single position in a given protein. The Sakamoto, Wang and Church laboratories engineered different sets of genomic changes in *E. coli* that mitigate the usual cytotoxic effect of deletion of the release factor that competes for decoding the amber stop signal (21–23). These efforts have produced cells where the meaning of the amber codon is 'free' and may be utilized to encode multiple copies of a non-canonical amino acid, generating a 21 amino acid genetic code. However, none of the methods is easily extendable to additional codons.

An alternative approach to incorporation of non-canonical amino acids, residue specific reassignment, allows for the multisite incorporation of ncAAs, but requires replacement of every occurrence of one natural amino acid with the ncAA. Residue specific reassignment operates through precisely controlling the growth medium such that the targeted natural amino acid is removed and replaced by a ncAA that is a close structural analog (24–27). The ncAA is utilized in translation in place of the removed natural amino acid, resulting in proteins with high levels of non-canonical amino acid incorporation at multiple sites without appreciable reductions in yield. The genetic code under reassignment conditions is an altered 20 amino acid code rather than an expanded code because one natural amino acid must be removed entirely.

A recently developed strategy combines aspects of the amber suppression and residue specific methods: breaking the degeneracy of the genetic code in order to reassign the meaning of individual sense codons. Sense codon reassignment has the potential to enable the simultaneous incorporation of multiple copies of multiple ncAAs into proteins (28). The genetic code is degenerate: the 61 sense codons specify 20 canonical amino acids. 18 of the 20 canonical amino acids are encoded by more than one codon. In *E. coli*, 43 tRNA species decode the 61 sense codons. The subset of 21 codons that are read through wobble interactions are potential targets for sense codon reassignment. Tirrell *et al.* reported the first example of breaking the degeneracy of the genetic code by incorporating ncAAs in response to one of two Phe codons (28). Recent reports have described the incorporation of ncAAs in response to reassigned rare arginine, isoleucine and serine codons (29–34).

We recently reported sense codon reassignment at lysine AAG, histidine CAU, asparagine AAU and phenylalanine UUU codons (35). Our evaluation utilized the orthogonal *Methanocaldococcus jannaschii* (*M. jannaschii*) tyrosyl tRNA and aaRS pair most commonly employed for the introduction of ncAAs. The *M. jannaschii* pair has been evolved to recognize and activate over 100 different ncAAs (3,19). We found that the degeneracy of the genetic code could be broken at each of the four sense selected codons with between 1% and 6% efficiency simply by introducing a variant of an engineered *M. jannaschii* tyrosyl tRNA (tRNA<sup>Opt</sup>, (36)) with an altered anticodon designed to recognize one of the selected codons through Watson–Crick base pairing. The levels of sense codon reassignment may be improved through directed evolution. The finding that many sense codons could be reassigned with the *M. jannaschii* system is significant because the anticodon is often an important identity element that allows a specific aaRS to recognize its appropriate tRNA. Changing the anticodon



**Figure 1.** (A) Sequence and secondary structure of *E. coli* tyrosyl tRNA molecule. Modified nucleobases are indicated. (B) Backbone trace of tRNA structure (pdb 1EHZ), color coded to match secondary structural elements in (A).

may affect the efficiency with which an amino acid is attached to the tRNA. In order for sense codon reassignment to be possible, the nucleotide changes in the anticodon of the orthogonal tRNA required to read the selected sense codon must not abrogate the interaction with the cognate orthogonal aminoacyl tRNA synthetase.

Breaking the degeneracy of the genetic code is likely optimal when the anticodon of the orthogonal tRNA pairs with a single sense codon on the mRNA. We utilized our GFP-based screen to evaluate the tendency of the *M. jannaschii* tyrosine-charged tRNA<sup>Opt</sup> with an altered anticodon to decode other codons via non-Watson-Crick base pairing (35). For three of the four codons, Lys AAG, Asn AAU and Phe UUU, we found that tRNA<sup>Opt</sup> with an altered anticodon discriminated between the two sense codons that specify each canonical amino acid. However, when tRNA<sup>Opt</sup> was altered to include an AUG anticodon (tRNA<sup>Opt</sup><sub>AUG</sub>) to decode one of two histidine codons, both the targeted CAU and the CAC histidine codons were decoded to approximately equal extent.

Relative to efforts to improve the behavior of orthogonal translational components through engineering mRNA codon/orthogonal tRNA anticodon interactions and orthogonal tRNA/aaRS interactions, little effort has been devoted to consideration of the third level of the genetic code: interactions between the orthogonal tRNA and host tRNA modifying enzymes. In this report, we describe experiments that identify a modification of tRNA<sup>Opt</sup><sub>AUG</sub> as the cause of the aberrant lack of discrimination between the two histidine codons. tRNA<sup>Opt</sup><sub>AUG</sub> is a substrate for the A to I deaminase TadA that is essential for maturation

of the *E. coli* tRNA<sup>Arg2</sup>. The resulting tRNA<sup>Opt</sup><sub>IUG</sub> has inosine at the first anticodon position and is capable of incorporating the amino acid it carries in response to both the CAU and CAC codons. We utilized directed evolution and our fluorescence-based screen to identify orthogonal tRNA variants that are not substrates for TadA and allow increased reassignment of the histidine CAU codon with high discrimination for the CAC codon. The unexpected modification of an orthogonal tRNA is, to our knowledge, the first described instance of a clash between genetic code engineering and tRNA modifications. These results highlight the fact that tRNA modifications should not be disregarded, as they may be both an obstacle to and target for genetic code engineering.

## MATERIALS AND METHODS

Cell strains and vector composition and construction have been described in detail previously (35); brief summaries are provided as Supplementary Material (S1–S3). pWB-Ultra plasmids used in this study differ from those reported only in the nucleotides corresponding to positions 32 through 38 of the tRNA<sup>Opt</sup> genes (Supplementary Table S1). Variants of pGFP reporters used in this study differ from those reported only in the sense codon that specifies the fluorophore position 66 (Supplementary Table S2) (37). Each histidine residue in GFP is encoded by a CAC codon in our reporter constructs; the only CAU codon in the reporter vectors occurs at position 66 to specify the critical fluorophore position. The GFP-based screen utilized to quantify sense codon reassignment has been described (35); details are provided as Supplementary Material (S4).



### Identification of inosine in tRNA anticodons

The complement of tRNA molecules was extracted from NEB 5-alpha cells harboring a pWB\_Ultra plasmid (38). Briefly, RNA was precipitated by phenol:chloroform extraction from cell lysates; a detailed protocol is provided in the Supplementary Material (S5). A total of 2.5  $\mu$ l of extracted RNA solution was used as the template for reverse transcription with primers specific for either *E. coli* tRNA<sup>Arg2</sup> (primer LW, Supplementary Table S3) or *M. jannaschii* tRNA<sup>Opt</sup> (primer LZ, Supplementary Table S3). Reverse transcription was performed using SuperScript IV reverse transcriptase (Thermo Scientific) according to the manufacturer's instructions. Duplicate reactions to which no reverse transcriptase was added were included as negative controls to ensure subsequent PCR amplification resulted from tRNA as opposed to contaminating DNA. The cDNA products from the reverse transcription reactions were amplified with Taq polymerase (New England Biolabs) using primer LX and either primer LY or primer CC (Supplementary Table S3). PCR products were purified using the GeneJET PCR Cleanup kit (Thermo Scientific) and analyzed by agarose gel electrophoresis. Amplified cDNA products were sequenced with primer LX (*M. jannaschii* tRNA<sup>Opt</sup> products) or primer LY (*E. coli* tRNA<sup>Arg2</sup>) (Genewiz LLC). Sanger sequencing .ab1 files were converted to .xml files using abi2xml.exe (freely available at [www.abi2xml.sourceforge.net](http://www.abi2xml.sourceforge.net)). Figures were created by importing data lines from the .xml file into Microsoft Excel.

### Construction of the tRNA anticodon loop library

A 64 member library of nucleotide variants at positions 32, 37 and 38 of the *M. jannaschii* tRNA<sup>Opt</sup><sub>AUG</sub> (Figure 1A) was generated via Kunkel mutagenesis using primer JQ (Supplementary Table S3). Plasmid pWB\_Ultra-Tyr-(XhoI) has an XhoI restriction site in the anticodon position of the tRNA and was used as an inactive template for library construction. Primer JQ allows A, C, G or U to be incorporated at positions 32, 37 and 38 and restores the AUG anticodon. Nearly all domains of life feature a conserved uridine at position 33 (14–16); this position was not varied. A total of 5  $\mu$ g of pWB\_Ultra-Tyr-(XhoI) ss dU DNA were used as template for library construction. Library DNA was electroporated into SS320 cells. A total of 10  $\mu$ l of the 1000  $\mu$ l transformation recovery were plated to determine transformation efficiency. The electroporation yielded  $4 \times 10^9$  unique transformants. PCR products from 8 of 13 of these colonies were not able to be digested by XhoI, suggesting a mutagenic efficiency of ~60%.

tRNA molecules are inherently highly structured, and mutation of these polynucleotides is challenging. The introduction of an XhoI restriction site into the template DNA provides a handle for improving the efficiency of mutation in addition to utilization of uridine-enriched template DNA (39). The remaining 990  $\mu$ l of the library transformation recovery media was diluted 5-fold into LB/spectinomycin 50  $\mu$ g/ml media and grown for 5 h at 37°C with shaking. Isolation of DNA from 5 ml of culture yielded 7.4  $\mu$ g of double stranded DNA, 2  $\mu$ g of which were restricted using XhoI (New England Biolabs) at 37°C for 1 h. Following PCR spin kit cleanup and elution in 35  $\mu$ l ultrapure water, one-half

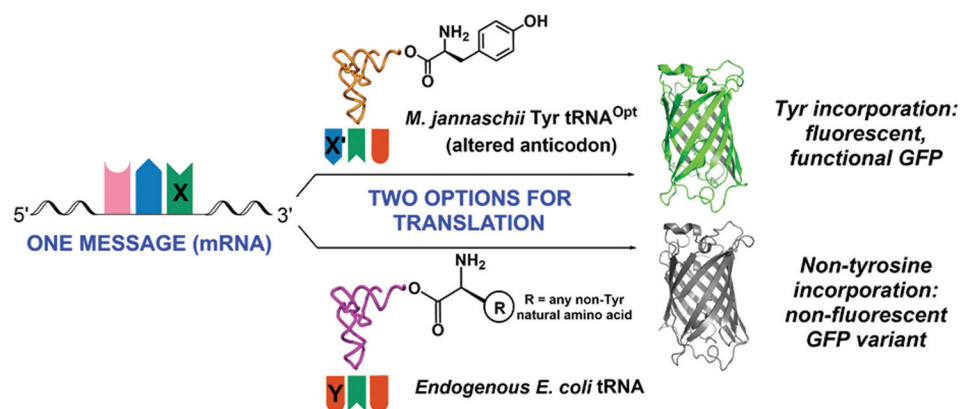
of the restricted DNA was transformed into SB3930 cells harboring the pGFP66cau reporter plasmid. The electroporation yielded  $4 \times 10^6$  unique transformants. PCR products from 13 of 13 evaluated colonies were not susceptible to XhoI digestion, suggesting that all transformants were library members.

## RESULTS AND DISCUSSION

Motivated by a general interest in reassigning the meaning of sense codons for incorporation of ncAAs into proteins *in vivo*, we recently reported a screen to evaluate the sense codon reassignment potential of the *M. jannaschii* tyrosyl tRNA/aaRS orthogonal pair (35). The screen exploits the absolute requirement for an active site tyrosine in green fluorescent protein (GFP). The screen monitors the restoration of GFP fluorescence by incorporation of tyrosine at position 66 in response to a sense codon typically assigned another meaning in the genetic code (Figure 2). When the sense codon at position 66 is read by *M. jannaschii* tyrosine-charged tRNA<sup>Opt</sup>, tyrosine is incorporated, and the resulting protein is fluorescent. When the sense codon is decoded by an endogenous *E. coli* tRNA, another canonical amino acid is incorporated, and the resulting protein is not fluorescent. The screen provides a quantitative measurement of sense codon reassignment by providing a combined measurement of the extent to which the orthogonal *M. jannaschii* tyrosyl aaRS recognizes and aminoacylates an *M. jannaschii* tRNA with an alternative anticodon and the extent to which the altered orthogonal tRNA competes against *E. coli* tRNA species to decode the codon specifying the essential tyrosine position of GFP.

Our initial evaluation assessed the levels of codon reassignment and synonymous codon discrimination at phenylalanine, histidine, asparagine and lysine codons. Unlike the high levels of codon discrimination observed for the Phe, Asn and Lys codons, tRNA<sup>Opt</sup><sub>AUG</sub> reassigned the CAU codon with an efficiency of 5% and the CAC codon with an efficiency of 4%. The lack of discrimination between histidine codons was initially puzzling. The nominal interactions determining discrimination for the synonymous Phe, Asn and His codons are identical. For sense codon reassignment of His CAU, Phe UUU and Asn AAU, we introduced an orthogonal tRNA<sup>Opt</sup> with an adenosine in the first anticodon position to specifically pair with the uridine-ending codon read via a wobble interaction by the *E. coli* tRNA. The energetic preference for Watson–Crick over wobble interactions was expected to strongly bias reassignment of one of the two degenerate codons (28,40). For histidine, phenylalanine and asparagine, *E. coli* contains a single tRNA species with a guanosine in the first anticodon position, position 34, that decodes two codons: one codon is read via a canonical Watson–Crick base pairing interaction (G–C), and the other codon is read via an energetically less favorable wobble interaction (G–U). In the case of Asn and His tRNAs, the guanosine is modified to queosine (Q), a nucleobase that pairs equally with C and U.

The supplied sense codon suppressing tRNA<sup>Opt</sup><sub>AUG</sub>, tRNA<sup>Opt</sup><sub>AAA</sub> and tRNA<sup>Opt</sup><sub>AUU</sub> each include an adenosine at position 34, a feature uncommon in *E. coli* tRNAs. Only a single *E. coli* tRNA, tRNA<sup>Arg2</sup>, is genetically-encoded



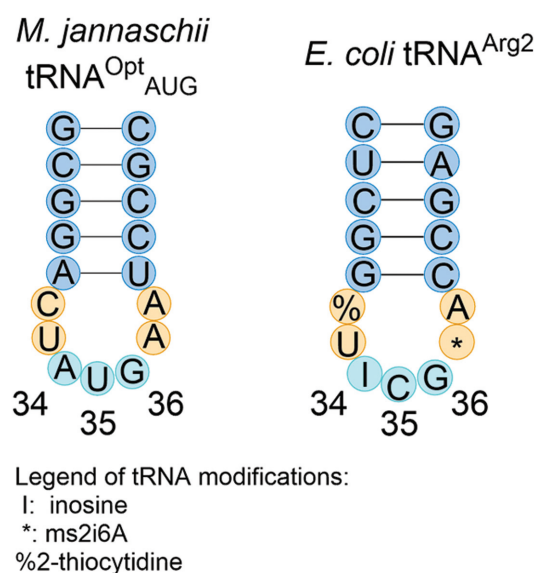
**Figure 2.** Depiction of the fluorescence-based screen utilized to evaluate sense codon reassignment.

with an adenosine at position 34. Although the anticodon of tRNA<sup>Arg2</sup> is genetically encoded as ACG, the functional form requires modification of the anticodon to ICG, where 'I' is the deaminated adenosine base, inosine. Inosine pairs with U, C and A, and the presence of inosine in an anticodon allows a single tRNA species to decode three codons, in this case CGU, CGC and CGA (41,42). The enzyme responsible for A to I deamination of position 34 of the *E. coli* tRNA<sup>Arg2</sup>, TadA, was recently identified (43-45). The set of interactions important for modification of a tRNA by TadA have not been definitively mapped, but have been localized to the anticodon stem loop. *E. coli* TadA was unable to deaminate the typically-deaminated yeast and human tRNA<sup>Ala</sup> species. Full TadA activity was maintained on truncated stem loop substrates with the Arg2 tRNA sequence, but TadA did not deaminate an adenosine equivalent to position 34 of a full tRNA in truncated stem loop substrates with a subset of nucleotide variations at positions 35 or 36 (43).

#### Evaluation of inosine modification of *M. jannaschii* tRNA<sup>Opt</sup><sub>AUG</sub>

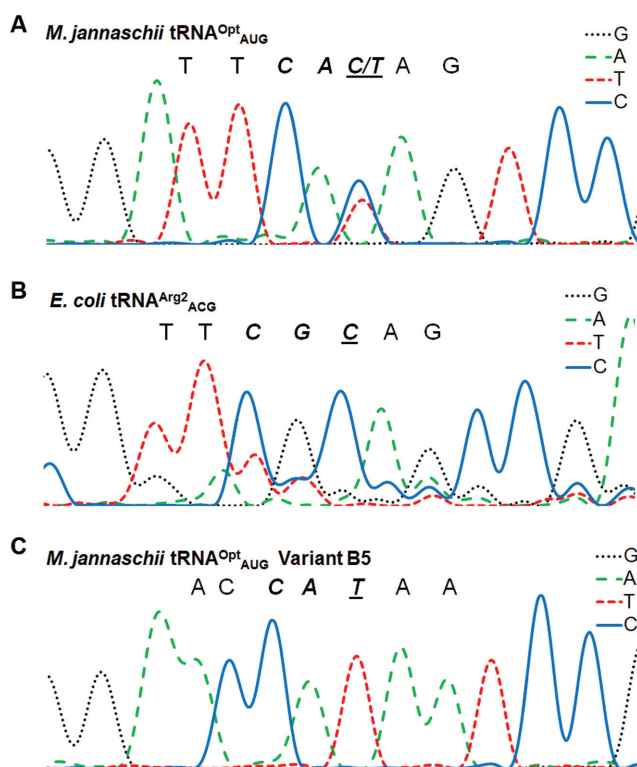
In comparing the sequences of the *E. coli* tRNA<sup>Arg2</sup> and the orthogonal *M. jannaschii* tRNA with an AUG anticodon, marked similarities and differences were apparent. The anticodon stems of the two tRNA species were largely different in sequence, sharing only one of five base pairs. The anticodon loops of precursor tRNA<sup>Arg2</sup><sub>ACG</sub> and tRNA<sup>Opt</sup><sub>AUG</sub> were remarkably similar in sequence, differing only by a single nucleotide, position 35, the middle of the anticodon (Figure 3). We hypothesized that tRNA<sup>Opt</sup><sub>AUG</sub> was a substrate for TadA, resulting in tRNA<sup>Opt</sup><sub>IUG</sub>. An IUG anticodon would theoretically be able to decode CAC, CAU and CAA codons and would explain the lack of discrimination observed for the two synonymous histidine codons. Phe and Asn codon reassigning tRNA<sup>Opt</sup><sub>AAA</sub> and tRNA<sup>Opt</sup><sub>AUU</sub> differ from the precursor tRNA<sup>Arg2</sup><sub>ACG</sub> at both positions 35 and 36. The combined differences make modification of tRNA<sup>Opt</sup><sub>AAA</sub> and tRNA<sup>Opt</sup><sub>AUU</sub> by TadA less likely.

Inosine modification of RNA may be detected by reverse transcription, amplification and sequencing of the resulting cDNA (43,46). Unmodified adenosine at position 34 of the



**Figure 3.** Comparison of anticodon stem loop sequences of *M. jannaschii* tRNA<sup>Opt</sup><sub>AUG</sub> and *E. coli* tRNA<sup>Arg2</sup><sub>IUG</sub>. Modified bases in the *E. coli* tRNA are labeled. The inosine modification of A34 confirmed in this work is the only modification noted in *M. jannaschii* tRNA<sup>Opt</sup><sub>AUG</sub>. Other modifications that have not yet been identified may exist.

tRNA is expected to trigger incorporation of only thymidine during reverse transcription. Inosine pairs most stably with cytidine and is expected to trigger significant incorporation of cytidine during reverse transcription. The complement of tRNAs was isolated from overnight cultures by phenol:chloroform extraction. Sanger sequencing of cDNA products replicated from the isolated tRNA fraction confirmed the incorporation of cytidine in response to the nucleotide at position 34 of both the *E. coli* tRNA<sup>Arg2</sup><sub>ICG</sub> and the *M. jannaschii* tRNA<sup>Opt</sup><sub>AUG</sub>. The chromatogram for the *E. coli* tRNA<sup>Arg2</sup>, which is presumably nearly quantitatively modified, shows almost exclusive incorporation of cytidine, indicating the presence of inosine (Figure 4B). The chromatogram for the original tRNA<sup>Opt</sup><sub>AUG</sub> shows an approximate 50:50 mixture of cytidine and thymidine (Figure 4A), suggesting that this tRNA is partially modified to inosine



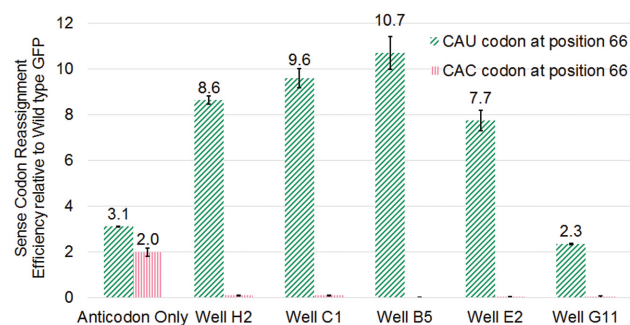
**Figure 4.** Sequence traces of reverse transcribed isolated tRNA species for (A) *M. jannaschii* tRNA<sup>Opt</sup><sub>AUG</sub>, (B) *E. coli* tRNA<sup>Arg2</sup> and (C) *M. jannaschii* tRNA<sup>Opt</sup><sub>AUG</sub> variant B5 identified from the anticodon loop library (traces of additional isolated tRNA variants are shown in Supplementary Figure S1). The sequenced stand corresponds to the reverse complement of the tRNA sequence. The portion of the chromatograms that correspond to the anticodon are italicized and bolded. The position corresponding to first anticodon position of the A to I modification is underlined. At position 34, observation of C is indicative of inosine modification; observation of T is indicative of unmodified adenosine.

at position 34. The tRNA sequencing traces suggest that the lack of discrimination observed for the *M. jannaschii* tRNA<sup>Opt</sup><sub>AUG</sub> was the result of anticodon modification to IUG, leading to increased decoding of the CAC codon.

The canonical wobble rules suggest that inosine should pair with U, C and A. The ability of the *M. jannaschii* Tyr-tRNA<sup>Opt</sup><sub>IUG</sub> to incorporate tyrosine in response to the CAA glutamine codon was examined, but no fluorescence was detected when the reporter GFP variant included a CAA codon specifying the fluorophore tyrosine residue at position 66. The lack of incorporation in response to CAA codon is likely due to the high abundance of *E. coli* Glu-tRNA<sub>UUG</sub> relative to the amount of Tyr-tRNA<sup>Opt</sup><sub>IUG</sub> in the cell. The tRNA<sup>Gln1</sup><sub>UUG</sub> which directly decodes the CAA codon is the second most abundant *E. coli* tRNA (47).

#### Directed evolution of the *M. jannaschii* tRNA<sup>Opt</sup><sub>AUG</sub> anticodon loop increases sense codon reassignment and prevents inosine modification of adenosine 34

The TadA substrate recognition elements in tRNA<sup>Arg2</sup> have not been fully mapped, but the anticodon loop was shown to be an important factor (43). We have reported that alterations to nucleotides flanking the anticodon are capa-



**Figure 5.** Codon discrimination for selected tRNA variants. Reassignment efficiencies reported as the observed optical density-corrected fluorescence divided by the optical density-corrected fluorescence of wild-type GFP for each orthogonal pair measured in response to CAU and CAC histidine codons specifying Y66 of GFP.

ble of increasing sense codon reassignment of Lys AAG codons by tRNA<sup>Opt</sup><sub>CUU</sub> (35). We hypothesized that variation within the anticodon loop of tRNA<sup>Opt</sup><sub>AUG</sub> could simultaneously increase the efficiency of sense codon reassignment at CAU codons and decrease deamination of position 34 by TadA, leading to increased discrimination between CAU and CAC codons. We constructed a focused library of 64 tRNA<sup>Opt</sup><sub>AUG</sub> variants with diversity at nucleotide positions 32, 37 and 38 in the anticodon loop. The universal U33 was maintained.

The tRNA anticodon loop library was screened using a sfGFP reporter with a CAU codon at position 66. In a screen of 85 colonies from the library, 34 samples had detectable fluorescence and 17 colonies showed fluorescence equal to or greater than the original tRNA<sup>Opt</sup><sub>AUG</sub>. The highly fluorescent colonies appeared to be clustered in groups of approximately equal fluorescence, and 6 colonies representing the apparent groups were selected for further characterization. The vector DNA specifying the orthogonal translational machinery from each of these colonies was isolated and sequenced (Table 1). Colony G10 was identified as the original tRNA<sup>Opt</sup><sub>AUG</sub> with no mutations to the anticodon loop; this variant was not characterized further. Each of the other five variants had at least one mutation in the anticodon loop, and two of these clones had multiple mutations. Alterations at each position included in the library (32, 37 and 38) result in improved reassignment of the CAU codon.

The five selected tRNA anticodon loop variants were evaluated for their ability to discriminate between the CAU and CAC codons. Each variant showed improved codon discrimination, suggesting that the mutations that improved the efficiency of sense codon reassignment of CAU also mitigated the A34I modification (Figure 5). The original tRNA<sup>Opt</sup><sub>AUG</sub> incorporated tyrosine in response to 5.1% of histidine codons, with a 3:2 preference for CAU over CAC codons. The incorporation of tyrosine in response to the CAC codon by the five tRNA<sup>Opt</sup><sub>AUG</sub> library variants was at or below the limit of detection for our in cell assay, 0.2% reassignment. The most improved systems expressing the tRNA variants selected from the library exhibit at least a 98:2 preference for CAU over CAC codons, a significant im-



**Table 1.** Sanger sequencing of reverse transcribed cDNA from isolated tRNA fractions

| Variant                             | tRNA Nucleotide Number |    |    |    |    |                |    |
|-------------------------------------|------------------------|----|----|----|----|----------------|----|
|                                     | 32                     | 33 | 34 | 35 | 36 | 37             | 38 |
| tRNA <sup>Opt</sup> <sub>IUG</sub>  | C                      | U  | I  | U  | G  | A              | A  |
| tRNA <sup>Arg2</sup> <sub>ICG</sub> | C <sup>a</sup>         | U  | I  | C  | G  | A <sup>b</sup> | A  |
| Colony B5                           | U                      | U  | A  | U  | G  | G              | U  |
| Colony C1                           | C                      | U  | A  | U  | G  | A              | C  |
| Colony E2                           | C                      | U  | A  | U  | G  | A              | U  |
| Colony G10                          | C                      | U  | I  | U  | G  | A              | A  |
| Colony G11                          | A                      | U  | A  | U  | G  | A              | C  |
| Colony H2                           | C                      | U  | A  | U  | G  | G              | A  |

<sup>a</sup>Position is modified to thiocytidine.  
<sup>b</sup>Position is modified to 2-methylthio-N6-isopentenyladenosine.

provement compared to the 3:2 preference observed with the original tRNA<sup>Opt</sup><sub>AUG</sub>. The tRNA fraction from cells expressing the tRNA<sup>Opt</sup><sub>AUG</sub> variants was isolated and sequenced. Sanger sequencing of reverse transcribed cDNA from the isolated tRNA fractions showed exclusive incorporation of thymidine, indicating that all five clones had unmodified adenosine in position 34 rather than inosine (Figure 4C and Supplementary Figure S1). Although not all possible combinations of mutations were observed, mutations at positions 32, 37 or 38 of the tRNA<sup>Opt</sup><sub>AUG</sub> appear to eliminate recognition of tRNA<sup>Opt</sup><sub>AUG</sub> as a substrate for TadaA and improve the reassignment efficiency of the targeted CAU codon.

**Prevention of inosine modification improves health of sense codon reassigning cell systems**

In general, cell growth rates for sense codon reassigning systems are slowed slightly relative to control systems in which no sense codon reassignment occurs. The maximum optical densities (carrying capacity) for cells expressing the *M. jannaschii* tRNA<sup>Opt</sup><sub>IUG</sub> average 75% of the maximum OD<sub>600</sub> of the two control systems (Supplementary Figure S2). Each tRNA<sup>Opt</sup><sub>AUG</sub> variant identified from the tRNA anticodon loop library grows to an increased relative maximal optical density, ranging from 84–90% of the two control systems. The wild-type GFP 100% and 0% ‘no reassignment’ controls each include an amber suppressing *M. jannaschii* tRNA<sup>Opt</sup><sub>CUA</sub> and aaRS which introduce tyrosine in response to amber stop codons. Inclusion of the machinery to reassign amber codons in the controls was intended to place a similar metabolic burden on all reassignment systems and attempt to minimize differences in growth profiles that could be a result of cells growing under different antibiotic conditions or with different expressed protein content.

The two synonymous histidine codons appear with approximately equal frequency throughout the *E. coli* genome; 55% of histidine is encoded by CAU and 45% of histidine is encoded by CAC. The orthogonal *M. jannaschii* tRNA<sup>Opt</sup><sub>IUG</sub> incorporated tyrosine in response to 3% of the CAU codons and 2% of the CAC codons in the GFP reporter, suggesting replacement of approximately 5% of the histidine codons throughout the *E. coli* genome. Given that the histidine side chain serves many important catalytic functions in proteins, we hypothesized that the decreased growth rates were symptomatic of cumulative negative ef-

fects of replacement of histidine with tyrosine. A more complicated explanation appears to be required, however, as each of the variants identified in the tRNA anticodon loop library exhibit both improved growth profiles and increased reassignment of histidine CAU codons throughout the *E. coli* genome. Variants H2, C1 and B5 appear to be reassigning twice as many histidine codons in the *E. coli* genome, yet exhibit more robust growth. Cell health and growth in reassignment systems is complicated and is not simply explained by the level of substitution of critical amino acids throughout the host’s genome.

**CONCLUSION**

The *in vivo* modification of tRNAs by endogenous host enzymes must be taken into account when considering the utility and orthogonality of exogenous translational machinery. Over 100 different modified RNA nucleobases have been identified in RNA, and the majority of these modifications are found in tRNA molecules. The greatest diversity of modification centers around the tRNA anticodons and has direct effects on fidelity of translation. We observed an anomalous lack of histidine codon discrimination by the *M. jannaschii* tRNA<sup>Opt</sup><sub>AUG</sub> and demonstrated that tRNA<sup>Opt</sup><sub>AUG</sub> is a substrate for the *E. coli* deaminase TadaA. *M. jannaschii* tRNA<sup>Opt</sup><sub>IUG</sub> was able to direct incorporation of the amino acid it carries in response to both the CAU and CAC codons. Sanger sequencing of cDNA products replicated from the *in vivo* tRNA transcripts confirmed that A34 is modified to inosine. We identified several *M. jannaschii* tRNA<sup>Opt</sup><sub>AUG</sub> anticodon loop variants from a focused library that simultaneously improved reassignment of the histidine CAU codon to tyrosine and decreased incorporation of tyrosine in response to the histidine CAC codon. The variations in the anticodon loop prevented recognition by TadaA, resulting in maintenance of A34 in the first position of the anticodon. Despite higher overall levels of sense codon reassignment at the His CAU codon throughout the *E. coli* genome, sense codon reassignment systems featuring non-inosine modified orthogonal tRNAs showed improved cell health and growth profiles relative to the original *M. jannaschii* tRNA<sup>Opt</sup><sub>IUG</sub>.

Variants of the *M. jannaschii* tyrosyl aaRS and tRNA are one of the most commonly utilized orthogonal pairs for incorporation of ncAAs into proteins. Although unintentional modification of orthogonal tRNA molecules has

not yet been suggested as a factor influencing the incorporation of ncAAs in response to stop codons, the extent to which orthogonal tRNA molecules are substrates for endogenous host tRNA modifying enzymes has not been systematically investigated. The alteration of orthogonal tRNA molecules to direct ncAA incorporation in response to different sense codons is a next step in the expansion of the genetic code from 21 to 22 amino acids (and beyond). Alteration of tRNA sequences will expose orthogonal translational machinery to additional interactions with endogenous host tRNA modifying machinery and may result in the unintentional generation of substrates for endogenous host enzymes. The resulting modified tRNAs may demonstrate anomalous and unexpected behavior in the process of protein translation.

Recognizing tRNA modification as both a potential pitfall and avenue of directed alteration to improve systems with expanded genetic codes will be important as the field of genetic code engineering continues to progress. The hypermodification present in tRNA species across all domains of life will undoubtedly produce other complications in translation systems, but the modifications of tRNAs will also serve as a potential handle to increase ncAA incorporation.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

National Science Foundation [1507055 to J.D.F.]. Funding for open access charge: National Science Foundation [1507055 to J.D.F.].

*Conflict of interest statement.* None declared.

## REFERENCES

- Lang, K. and Chin, J.W. (2014) Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.*, **114**, 4764–4806.
- Ngo, J.T. and Tirrell, D.A. (2011) Noncanonical amino acids in the interrogation of cellular protein synthesis. *Acc. Chem. Res.*, **44**, 677–685.
- Dumas, A., Lercher, L., Spicer, C.D. and Davis, B.G. (2015) Designing logical codon reassignment - expanding the chemistry in biology. *Chem. Sci.*, **6**, 50–69.
- de Duve, C. (1988) The second genetic code. *Nature*, **333**, 117–118.
- Giege, R., Sissler, M. and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.*, **26**, 5017–5035.
- El Yacoubi, B., Bailly, M. and de Crécy-Lagard, V. (2012) Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu. Rev. Genet.*, **46**, 69–95.
- Persson, B.C. (1993) Modification of transfer-RNA as a regulatory device. *Mol. Microbiol.*, **8**, 1011–1016.
- Manickam, N., Joshi, K., Bhatt, M.J. and Farabaugh, P.J. (2016) Effects of tRNA modification on translational accuracy depend on intrinsic codon-anticodon strength. *Nucleic Acids Res.*, **44**, 1871–1881.
- Limbach, P.A., Crain, P.F. and McCloskey, J.A. (1994) The modified nucleosides of RNA - summary. *Nucleic Acids Res.*, **22**, 2183–2196.
- Gray, M.W. (2012) Evolutionary origin of RNA editing. *Biochemistry*, **51**, 5235–5242.
- Alseth, I., Dalhus, B. and Bjoras, M. (2014) Inosine in DNA and RNA. *Curr. Opin. Genet. Dev.*, **26**, 116–123.
- Bazak, L., Haviv, A., Barak, M., Jacob-Hirsch, J., Deng, P., Zhang, R., Isaacs, F.J., Rechavi, G., Li, J.B., Eisenberg, E. *et al.* (2014) A-to-i RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.*, **24**, 365–376.
- Paul, M.S. and Bass, B.L. (1998) Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *EMBO J.*, **17**, 1120–1127.
- Dunin-Horkawicz, S., Czerwoniec, A., Gajda, M.J., Feder, M., Grosjean, H. and Bujnicki, J.M. (2006) Modomics: a database of RNA modification pathways. *Nucleic Acids Res.*, **34**, D145–D149.
- Czerwoniec, A., Dunin-Horkawicz, S., Purta, E., Kaminska, K.H., Kasprzak, J.M., Bujnicki, J.M., Grosjean, H. and Rother, K. (2009) Modomics: A database of RNA modification pathways. 2008 update. *Nucleic Acids Res.*, **37**, D118–D121.
- Machnicka, M.A., Milanowska, K., Oglou, O.O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K.M. *et al.* (2013) Modomics: A database of RNA modification pathways-2013 update. *Nucleic Acids Res.*, **41**, D262–D267.
- Furter, R. (1998) Expansion of the genetic code: Site-directed p-fluoro-phenylalanine incorporation in *Escherichia coli*. *Protein Sci.*, **7**, 419–426.
- Wang, L., Brock, A., Herberich, B. and Schultz, P.G. (2001) Expanding the genetic code of *Escherichia coli*. *Science*, **292**, 498–500.
- Liu, C.C. and Schultz, P.G. (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem.*, **79**, 413–444.
- Wan, W., Tharp, J.M. and Liu, W.R. (2014) Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta*, **1844**, 1059–1070.
- Mukai, T., Hayashi, A., Iraha, F., Sato, A., Ohtake, K., Yokoyama, S. and Sakamoto, K. (2010) Codon reassignment in the *Escherichiacoli* genetic code. *Nucleic Acids Res.*, **38**, 8188–8195.
- Johnson, D.B.F., Xu, J.F., Shen, Z.X., Takimoto, J.K., Schultz, M.D., Schmitz, R.J., Xiang, Z., Ecker, J.R., Briggs, S.P. and Wang, L. (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.*, **7**, 779–786.
- Isaacs, F.J., Carr, P.A., Wang, H.H., Lajoie, M.J., Sterling, B., Kraal, L., Tolonen, A.C., Gianoulis, T.A., Goodman, D.B., Reppas, N.B. *et al.* (2011) Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement. *Science*, **333**, 348–353.
- Dougherty, M.J., Kothakota, S., Mason, T.L., Tirrell, D.A. and Fournier, M.J. (1993) Synthesis of a genetically engineered repetitive polypeptide containing periodic selenomethionine residues. *Macromolecules*, **26**, 1779–1781.
- Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. and Huber, R. (1995) High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. *Eur. J. Biochem.*, **230**, 788–796.
- Budisa, N. (2006) *Engineering the genetic code expanding the amino acid repertoire for the design of novel proteins*. Wiley-VCH, Weinheim.
- 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.
- Kwon, I., Kirshenbaum, K. and Tirrell, D.A. (2003) Breaking the degeneracy of the genetic code. *J. Am. Chem. Soc.*, **125**, 7512–7513.
- Kwon, I. and Lim, S.I. (2014) Tailoring the substrate specificity of yeast phenylalanyl-tRNA synthetase toward a phenylalanine analog using multiple-site-specific incorporation. *ACS Synth. Biol.*, **5**, 634–643.
- Krishnakumar, R., Prat, L., Aerni, H.-R., Ling, J., Merryman, C., Glass, J.I., Rinehart, J. and Söll, D. (2013) Transfer RNA misidentification scrambles sense codon recoding. *ChemBioChem*, **14**, 1967–1972.
- Bohlke, N. and Budisa, N. (2014) Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: rare isoleucine codon AUA as a target for genetic code expansion. *FEMS Microbiol. Lett.*, **351**, 133–144.
- Zeng, Y., Wang, W. and Liu, W.S.R. (2014) Towards reassigning the rare AGG codon in *Escherichia coli*. *ChemBioChem*, **15**, 1750–1754.
- Lee, B.S., Shin, S., Jeon, J.Y., Jang, K.-S., Lee, B.Y., Choi, S. and Yoo, T.H. (2015) Incorporation of unnatural amino acids in response to the AGG codon. *ACS Chem. Biol.*, **10**, 1648–1653.
- Ho, J.M., Reynolds, N.M., Rivera, K., Connolly, M., Guo, L.-T., Ling, J., Pappin, D.J., Church, G.M. and Söll, D. (2016) Efficient



- reassignment of a frequent serine codon in wild-type *Escherichia coli*. *ACS Synth. Biol.*, **5**, 163–171.
35. Biddle, W., Schmitt, M.A. and Fisk, J.D. (2015) Evaluating sense codon reassignment with a simple fluorescence screen. *Biochemistry*, **54**, 7355–7364.
36. Young, T.S., Ahmad, I., Yin, J.A. and Schultz, P.G. (2010) An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J. Mol. Biol.*, **395**, 361–374.
37. Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. and Waldo, G.S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.*, **24**, 79–88.
38. Varshney, U., Lee, C.P. and Rajbhandary, U.L. (1991) Direct analysis of aminoacylation levels of transfer-RNAs *in vivo* - application to studying recognition of *Escherichia coli* initiator transfer-RNA mutants by glutaminyl-transfer RNA-synthetase. *J. Biol. Chem.*, **266**, 24712–24718.
39. Holland, E.G., Acca, F.E., Belanger, K.M., Bylo, M.E., Kay, B.K., Weiner, M.P. and Kiss, M.M. (2015) *In vivo* elimination of parental clones in general and site-directed mutagenesis. *J. Immunol. Methods*, **417**, 67–75.
40. Meroueh, M. and Chow, C.S. (1999) Thermodynamics of RNA hairpins containing single internal mismatches. *Nucleic Acids Res.*, **27**, 1118–1125.
41. Agris, P.F. (2004) Decoding the genome: A modified view. *Nucleic Acids Res.*, **32**, 223–238.
42. Crick, F.H.C. (1966) Codon-anticodon pairing – the wobble hypothesis. *J. Mol. Biol.*, **19**, 548–555.
43. Wolf, J., Gerber, A.P. and Keller, W. (2002) TadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. *EMBO J.*, **21**, 3841–3851.
44. Auxilien, S., Crain, P.F., Trewyn, R.W. and Grosjean, H. (1996) Mechanism, specificity and general properties of the yeast enzyme catalysing the formation of inosine 34 in the anticodon of transfer RNA. *J. Mol. Biol.*, **262**, 437–458.
45. Gerber, A.P. and Keller, W. (1999) An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science*, **286**, 1146–1149.
46. Motorin, Y., Muller, S., Behm-Ansmant, I. and Brantant, C. (2007) Identification of modified residues in RNAs by reverse transcription-based methods. *Methods Enzymol.*, **425**, 21–53.
47. Dong, H.J., Nilsson, L. and Kurland, C.G. (1996) Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.*, **260**, 649–663.